Secondary Structure Determination of Insect Neuropeptides using High-Resolution Nuclear Magnetic Resonance and various Molecular Mechanics Calculations.

A thesis submitted in fulfilment of the requirements of the degree of MSc.

Thesis by: Katherine Sunter

Supervisor: Associate Professor G.E. Jackson

"To summarize briefly: A white rabbit is pulled out of a top hat. Because it is an extremely large rabbit, the trick takes many billions of years. All mortals are born at the very tip of the rabbit's fine hairs, where they are in a position to wonder at the impossibility of the trick. But as they grow older they work themselves even deeper into the fur. And there they stay. They become so comfortable they never risk crawling back up the fragile hairs again. Only philosophers embark on this perilous expedition to the outermost reaches of language and existence. Some of them fall off, but others cling on desperately and yell at the people nestling deep in the snug softness, stuffing themselves with delicious food and drink.

"Ladies and gentlemen," they yell, "we are floating in space!" But none of the people down there care."

Taken from "Sophies World" by Jostein Gaarder.

To my Parents

Table of Contents

Acknowledgements

I would like special thanks to go to Professor Graham Jackson for all his help, support and patience for the duration of the project. In addition, I would also like to thank Professor Gäde for supplying the peptides used in this project, Pierre Janssens for all his help in the purification process, Margie Nair for running all the NMR spectra, Graeme Broster for the preliminary work he did on Lom-AKH-I, and all my fellow colleagues in the Chemistry Department. I would also like to thank Martin Stülpner for his invaluable help and support throughout the thesis.

Abstract

Insect neuropeptides play a vital role in the hormone release processes associated with insect flight. Elucidation of the metabolic flight pathways requires a knowledge of the peptide secondary structure to allow predictions to made regarding hormone-receptor binding processes. The three insect neuropeptides under investigation were taken from the *corpora cardiaca* of the migratory locust (*Locusta migratoria*) and belonged to the Lom peptide series - Lom-AKH-I, II and III. The secondary structure of these hormones was investigated using high-resolution nuclear magnetic resonance (NMR) techniques and various molecular mechanics computations. The interproton distances and ϕ torsion angles obtained from the NMR data were used to constrain the peptides in subsequent Monte Carlo and molecular dynamics (simulated annealing) calculations. The results of these calculations indicated that Lom-AKH-I and Lom-AKH-III adopt a definite structure in d^6-DMSO solution, while Lom-AKH-II appeared to be fluctional with a minimum of three structures being required to satisfy the experimental data. Lom-AKH-I and III were shown to adopt a loose open turn structures stabilised by the presence of a single hydrogen bond. Both had an additional concave structure with hydrophobic clustering on the convex surface of the molecule. The results suggest that Lom-AKH-I and Lom-AKH-III are more reactive than Lom-AKH-II. These predictions are in agreement with literature results obtained for Lom-AKH-I and II for lipid mobilisation and phosphorylase activation. However, when measured against cAMP production, Lom-AKH-II has higher activity. Insufficient literature was available to make structure/activity comparisons for Lom-AKH-III.

Symbols and Abbreviations

AKH	Adipokinetic Hormone
Ala, A	Alanine
Asn, N	Asparagine
CD	Circular Dichroism
CC	*corpora cardiaca*
COSY	Correlated Spectroscopy
τ_c	Correlation time
cAMP	cyclic adenosine monophosphate
d^6-DMSO	deuterated Dimethylsulphoxide
Gly, G	Glycine
Homo-2DJ	Homonuclear 2DJ-Resolved
ISPA	Isolated Spin Pair Approximation
Leu, L	Leucine
MD	Molecular Dynamics
MM	Molecular Mechanics
MC	Monte Carlo
NMR	nuclear magnetic resonance
NOESY	Nuclear Overhauser and Exchange Spectroscopy
nOe	nuclear Overhauser effect
Phe, F	Phenylalanine
Pro, P	Proline
pGlu	pyroglutamic acid
ROESY	Rotating frame Overhauser and Exchange Spectroscopy
RMS	root mean square
Ser, S	Serine
SDS	Sodium Dodecyl Sulphate

Thr, T	Threonine
TOCSY	Total Correlated Spectroscopy
Trp, W	Tryptophan

Other abbreviations and symbols are defined in the chapter(s) in which they are used.

CHAPTER 1

INTRODUCTION

Insect flight is an extremely demanding activity. For this reason advanced evolution of flight processes was necessary if insects were to succeed not only at fast short distance flying, but also in covering the vast distances such as those witnessed in migration [1]. Energy rich carbohydrate and lipid reserves constitute the fuel sources required to sustain flight. During migration, the complete exhaustion of haemolymph carbohydrate reserves after approximately half an hour, coincides with the simultaneous peaking of lipid levels as they become the sole source of fuel for the remainder of the flight. These high lipid levels subsequently decline slightly and are maintained at a constant level for several hours until they too are exhausted, and the insect is forced to rest and replenish its reserves [2]. It is the subsequent frenzied feeding of these locust plagues that results in extensive crop damage and massive financial losses for many farmers.

The flight process is thus reliant on efficient oxidation of both carbohydrates and lipids. Carbohydrates, constituting the primary source of fuel for short flights, are stored as glycogen in the fat body and trehalose in the haemolymph. At the onset of flight, they are transported to the flight muscles, transformed into glucose and used for energy production. The lipids reserves, vital for longer, more demanding flights, are stored in the fat body as triacylglycerol (TG) [1]. TG needs to be converted into stereospecific 1,2 diaclyglycerol (DG) having an sn-1,2-configuration [3], before being released into the haemolymph and transported to the flight muscles. Membrane bound lipase enzymes then catalyse the hydrolysis of DG to generate fatty acids, the subsequent oxidation of which leads to energy release [1]. These mobilisation processes, particularly lipid mobilisation, are carefully controlled by the action of a series of insect neuropeptides known as adipokinetic hormones (AKHs) [1].

These hormones, synthesised and stored in the *corpus cardiacum* (CC) of a variety of studied insect species, are all structurally very similar but do exhibit a significant degree of functional diversity. Distinguishing features of these hormones include that they are octa-, nona-, or decapeptides having blocked amide (pyroglutamic acid) and carboxyl (amidated) termini and conservation of the primary sequence at positions 4 (with one exception), 8 and 9. The fact that the variant residues are the result of point mutations further demonstrates the close genetic relationship between these peptides (Figure 1.1) [1,4].

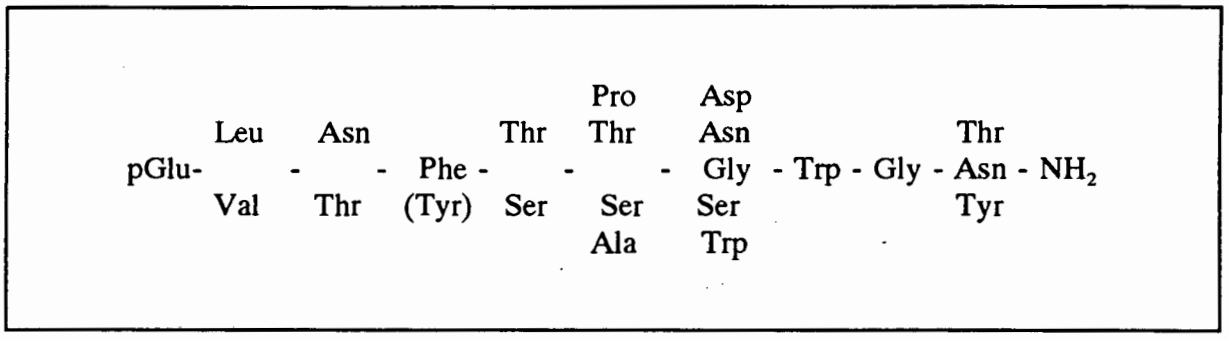


Figure 1.1: Variant and invariant residues in the AKH family.

The main focus of this thesis was on the three AKHs which have thus far been isolated from the migratory locust (*Locusta migratoria*). These three peptides were grouped together in a Lom peptide series and named Lom-AKH-I, II and III in accordance with the time of their isolation and identification. Lom-AKH-I, the first of the insect neurohormones to be completely characterised in 1976 by Stone *et. al.* [5], was also isolated from the desert locust (*Schistocerca gregaria*) [1]. Although, in 1979 a second peptide having adipokinetic activity was discovered in both *L. migratoria* and *S. gregaria*, amino acid composition determination [6,7] and sequencing [8,9] revealed two octapeptides differing by a single substitution at position five in the sequence [1]. The one pertaining to *L. migratoria* was thus named Lom-AKH-II and the other Scg-AKH-II. The third peptide, Lom-AKH-III, was only recently isolated and sequenced by Ouderjans *et al.* (1991) [10] and was found to occur exclusively in *L. migratoria.* All further discussion will therefore refer to these three Lom adipokinetic hormones [1].

Synthesis of locust AKHs takes place in the glandular lobes of the *corpora cardiaca* and their release is an indirect process controlled by neurosecretory cells residing in the lateral region

of the protocerebrum. Axons, extending from the cell bodies of these neural cells, run through the *nervi corpora cardiaca* II (NCC II) and establish synaptic contact with the glandular lobe hormone producing cells [3,11-13]. The further involvement of the NCC I in the hormone release process was demonstrated by the fact that severance of both the NCC I and II completely inhibited hormone release [12,14,15], while selective severance resulted in only partial release [3]. Although electrical stimulation of the NCC II has been shown to directly induce release of both AKH-I and II [16], NCC I stimulation only served to indirectly enhance the stimulated NCC II hormone release process [11,17]. The presence of a monophenolic amine, octopamine, in the NCC II and the glandular lobes [18] of the locust indicated the possible involvement of this molecule in the hormone release process. This has subsequently been verified in studies which showed a response identical to that observed with electrical stimulation of the NCC II in the presence of octopamine and inhibition of this response in the presence of an α-adrenergic receptor antagonist [11,19]. The fact that both NCC II stimulation and octopamine both promote cAMP formation within the glandular lobes [20] and given that cAMP plays an important part in AKH release [11,21], it has been concluded that octopamine mediates hormone release through its control of cAMP levels within the neurosecretory cells [11,20].

Although the primary physiological role of these adipokinetic hormones is lipid mobilization and transport, they have been seen to affect a number of other processes including flight muscle metabolism, protein synthesis and carbohydrate metabolism [11].

1.1 Lipid Mobilisation

In the locust, lipids are stored in the fat body in the form of triacylglycerols. The fat body constitutes the main lipid storage organ, but is also involved in carbohydrate storage in the form of glycogen [2]. Stored substances are broken down and the products used for synthesis of energy rich molecules which are then exported from the fat body to the region of interest eg. the flight muscles [2]. The role played by adipokinetic hormones in the lipid mobilisation process involves the stimulation of cyclic adenosine monophosphate (cAMP) production in the fat body [22]. This increase in cAMP production due to AKH activity has been indicated in many *in vitro* and *in vivo* studies [23-26] and leads to the conventional protein kinase chain

involving both cAMP and Ca^{2+} as secondary messengers [22]. Cyclic AMP is formed through the activation of a membrane bound adenylate cyclase enzyme which catalyses the breakdown of ATP. The activation process usually involves binding of a hormone, in this case AKH, to receptor sites on the cyclase enzyme (Figure 1.2) [2,27]. Cyclic AMP then activates a cAMP-dependent protein kinase which catalyses the phosphorylation of triacylglyceride lipase. This activated enzyme then catalyses the formation of monoacylgycerides from triacylglyceride [22]. Stereospecific sn-1,2-diacylglycerides are then generated through re-acylation of the monoacylglycerides by monoacylglyceride transferase [22]. This process of lipid mobilisation has been shown *in vitro* to be dependent on extracellular calcium [22].

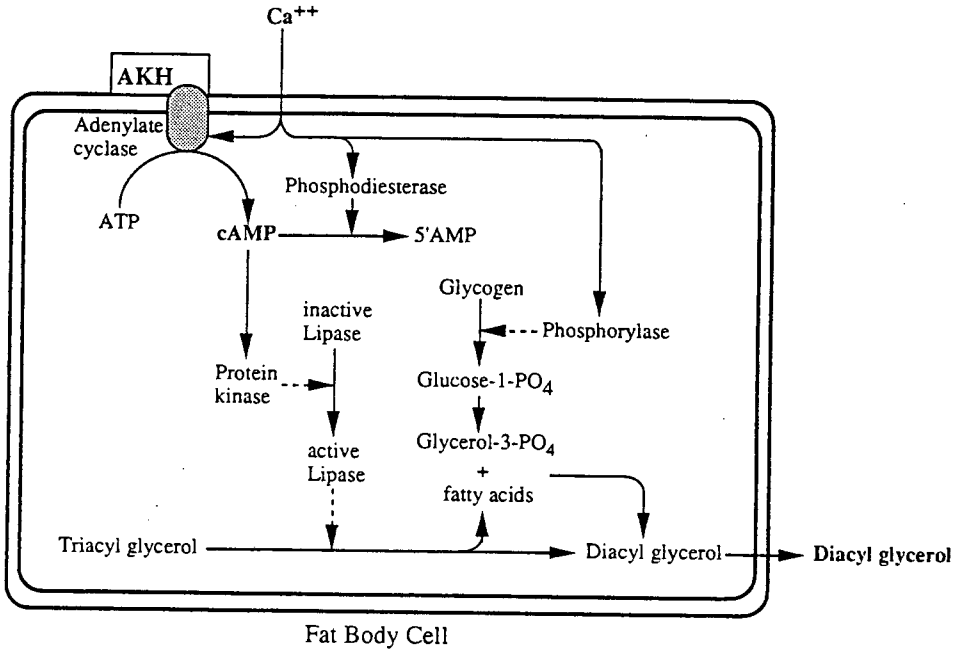


Figure 1.2: Model for the action AKH on the fat body illustrating the lipid mobilisation process. (Taken from Nijhout's "Insect Hormones" [2].)

AKH activity in terms of lipid mobilisation is not restricted to the fat body, and has been reported to stimulate the formation of DGs from lipids contained in the haemolymph [28].

1.2 Lipid Transport

Diacylglycerides, being neutral lipids are virtually insoluble in water and therefore the same applies to the haemolymph. Thus, to transport these substances to the flight muscles, use is made of lipoproteins which can readily undergo a composition change during flight [22]. In

the presence of AKH, rearrangement of haemolymph proteins and lipoproteins occurs to generate a new lipoprotein possessing a high affinity for DGs [22]. The lipoprotein of the resting haemolymph is commonly referred to as A_{yellow} (this constitutes the major form of DG loading and transport in the resting insect [11]) while the activated complex is known as A^+ lipoprotein. This A^+ lipoprotein is a high molecular weight molecule formed through the reversible association/binding of A_{yellow} with certain non-lipid containing haemolymph proteins called C_L-apoproteins [29]. This activated complex carries up to eighteen times more lipid than the corresponding inactive A_{yellow} molecule [22].

1.3 Lipid unloading

Membrane bound lipoprotein lipases present in the flight muscles catalyse the release of DG from both the activated A^+ lipoprotein and A_{yellow}. This enzyme has been shown to be four times more efficient in the release of DG from A^+ lipoprotein than A_{yellow}, and when both forms are present has shown up to a 90% decrease in activity towards the latter [22,30]. The explanation for the above lies in the inhibitory competitive binding which occurs between free C_L-apoproteins and lipoprotein lipase. In the presence of AKH the haemolymph concentration of A^+ is high i.e. the C_L-apoprotein concentration is low, inhibition of the lipoprotein lipase enzyme is minimised and DG release into the flight muscles is maximised. Conversely, a drop in AKH levels is accompanied by a decrease in A^+ lipoprotein levels and the concomitant increase in concentration of free C_L-apoprotein induces effective inhibition of the lipase enzyme which minimises the DG release into the flight muscles [22,30]. This indirect control of the unloading process by AKHs is vital for fast and efficient provision of fuel to the flight muscles during flight [21]. Also of importance, is the fact that trehalose inhibition in the presence of AKH will only occur when A^+ levels are sufficient to sustain flight, thereby preventing "stalling" within the flight muscles [22]. Thus, although initially trehalose and DGs may compete as fuel for flight, as soon AKH levels increase to the extent that significant amounts of A^+ lipoprotein are being formed, DG will become the sole source of fuel in the flight muscles [22].

1.4 Flight Muscle Metabolism

The direct action of AKHs within the flight muscle was indicated when Robinson *et al.* showed that extracts from glandular lobes of insects increased the rate of lipid oxidation within the flight muscle tissue [11,31,32]. Again to ensure continuity of fuel supply to the flight muscles octopamine levels, which carry out the initial glucose oxidation, are maintained until AKH levels are high enough to ensure sufficient oxidation of lipid to sustain flight [11,32].

1.5 Carbohydrate Regulation

The metabolism of fat body glycogen to form haemolymph trehalose is termed glycogenolysis and is catalysed by the glycogen phosphorylase enzyme. Phosphorylation of this enzyme by phosphorylase kinase generates active phosphorylase *a* from inactive phosphorylase *b* [27,33]. The activation of glycogen phosphorylase by AKH has been demonstrated in experiments involving injection of CC extracts into adult locusts [26,33,34]. Injection of synthetic AKH-I showed an increase in activity to a lesser extent than the former CC extract, and injection of storage lobe material showed glycogen phosphorylase activation but no simultaneous increase in lipid levels [34]. The same was true for "stressed" locusts (locusts were rotated in an Erlenmeyer flask and therefore constantly had to re-orientate themselves) [34]. These observations signify the existence of other glycogen phosphorylase activating substances thus suggesting AKH to be only semi-responsible for glycogen phosphorylase activation [34]. Although later studies by van Marrewijk *et al.* confirmed the activation of fat body glycogen phosphorylase by AKH-I, they were contradictory in so far as indicating dose-response relationships for CC extract and synthetic AKH-I to be virtually identical [33]. This finding, however, required an assumption on the AKH content of the CC extract which ignored factors relating to the age and sex of the locusts. An overestimation of the CC AKH content would imply the existence of other compounds also inducing glycogen phosphorylase activity [33]. In 1986 Goldsworthy *et al.* showed Lom-AKH-II to be effective in glycogen phosphorylase activation, but to a far lesser extent than AKH-I (see later - conclusion) [26,28]. Extracellular calcium has also been shown to be essential for this process to occur [35,36]. Unlike AKH, octopamine has been shown to have absolutely no effect on glycogen phosphorylase in the

locust fat body [33].

The effect of AKH on trehalosaemia has been shown to be age-dependent - although some effect was observed in larvae, none was observed in adult locusts [28,37].

1.6 Protein Synthesis

Finally synthetic AKH-I has also been shown in both *in vivo* and *in vitro* studies to inhibit fat body protein synthesis [11,25,38,39] while injection of glandular lobe extracts have demonstrated inhibition of haemolymph protein synthesis [11,39].

1.7 Effect of Age and Sex of Locusts on AKH release

Various responses to AKH release were also studied with respect to age and sex of locusts [40]. These studies indicated that lipid mobilisation in nymphs and fledglings was low relative to that observed in mature adult locusts. Lipid level studies (total lipid levels and diglyceride levels) for mature adult locusts were conducted with respect to the number of days after fledging. After injection of glandular lobe extract, the total lipid level study showed low lipid levels (a slight increase between the second and third days was observed [40]) for the first three days followed by a dramatic increase between the fourth and eighth days. These high levels of response were maintained until the locust started aging after about thirty two days at which stage a decrease in response was observed (Figure 1.3). The diglyceride levels, however, showed no increase until four days after fledging and showed a

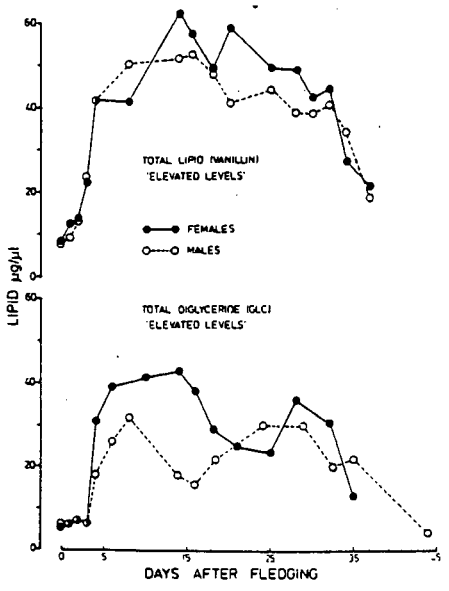


Figure 1.3: Haemolymph total lipid and diglyceride levels recorded for *Locusta* males and females after injection with glandular lobe extract ('elevated levels'). Points are representative of the average for groups of ten (lipid measurements) and five (diglyceride measurements) locusts. (Taken from Mwangi *et al.* [40].)

significant dip after approximately ten (males) or fifteen (females) days. This indicated that, in addition to DG, levels of certain other blood lipid types were also increased on injection of glandular lobe extract [40].

The reason for the low haemolymph lipid levels in nymphs, fledglings and aged locusts has been shown not to be dose dependent and requires further investigation to provide a reasonable explanation [40]. Although the total haemolymph lipid levels showed no significant difference throughout the adult life between males and females, the same was not true for diglyceride. In both males and females a dip in lipid levels was observed - in females after about 10 days and males after about 15 days (Figure 1.3). These dips appeared to correspond to the end somatic growth and the reason for their occurrence is unclear. It was proposed that they may correspond to times of optimum flight performance in both sexes [40].

1.8 The Overall Flight Process.

Extensive studies have been carried out on the actual flight process with regards to coordinating all the different hormones and substrates known to be involved in this process. The three hormones exhibiting adipokinetic activity which are released during locust flight are Lom-AKH-I, Lom-AKH-II and octopamine. The very recent isolation of Lom-AKH-III has meant its exclusion from all studies regarding its contributions to the flight process. In a study by Orchard *et al.* [41] using *Locusta migratoria*, it was shown that during the course of a sixty minute flight there occurred two separate increases in the haemolymph lipid levels - the first, occurring shortly after the onset of flight, peaked at ten minutes where it was then maintained until the second, more pronounced increase, after approximately twenty minutes of flight. Since a linear increase in utilisation of diglycerols from the onset of flight for approximately the first twenty minutes, had previously been reported [42], the plateau observed after the first ten minute peak suggested a decrease in lipid mobilisation and not lipid utilisation [41]. Hence it was predicted that during locust flight there were, in fact, two phases of lipid mobilisation [41]. Further studies on the point of release of AKH-I and AKH-II showed a marked difference between the two hormones. Lom-AKH-I was released after ten minutes of flight and showed a linear increase until approximately thirty minutes of flight, whereupon a rapid decrease in levels was observed. The activity of Lom-AKH-II, however,

only became apparent after the first fifteen minutes of flight from when it was seen to increase in concentration up until thirty minutes. This level of activity was then more or less maintained (a slight decrease may have been observed) over the next thirty minutes. These results thus suggested that only the second phase of lipid increase was attributable to AKH activity [41]. Regarding the first phase of lipid increase, Goosey and Candy showed a significant increase in octopamine levels during the first ten minutes of flight in the desert locust, *Schistocerca gregaria* [43]. Since octopamine has been shown to act directly on the fat body thereby inducing lipid release into the haemolymph it was concluded that octopamine in fact induced the initial phase of lipid increase [44]. In 1984 Orchard *et al.* [24] showed a correlation between increased haemolymph lipid levels and production of cAMP. Cyclic AMP was also produced in two phases during a sixty minute flight - the first increase peaked after approximately ten minutes of flight and the second after approximately thirty, whereafter this level was maintained for the remainder of the sixty minute flight. If the neck was ligated and AKH release to the rest of the body prevented, the second rise in cAMP was found to be absent [24]. This confirmed both that AKH release induced cAMP production in the fat body and that the second phase of lipid mobilisation was indeed the result of AKH activity [24]. Previous studies have linked octopamine to increases in cAMP levels in the locust fat body [44]. This finding, together with the fact that neck ligation did not prevent the first stage of cAMP elevation [24] and AKH titre elevations were not observed in the initial stages of flight [41,45], thus proved that the first increase in fat body cAMP and therefore the first phase of lipid mobilisation was due to octopamine activity [41].

The release of AKH-I and II has further been shown to correspond to a decrease in levels of haemolymph trehalose i.e. low levels of trehalose could constitute a further stimulus for AKH release [41,46]. This was further supported by the fact that injection of trehalose delayed, not only the flight stimulated release of AKH, but also that of octopamine [24,47]. This fact is consistent with the idea that initially carbohydrates are used as a source of energy and that a decrease in their levels, associated with longer flights, then induces the hormone controlled mobilisation and release of a new source of energy - lipids [1].

The overall model for the flight process is illustrated in Figure 1.4. The release of octopamine after ten minutes of flight induced the first elevation fat body cAMP and therefore also the

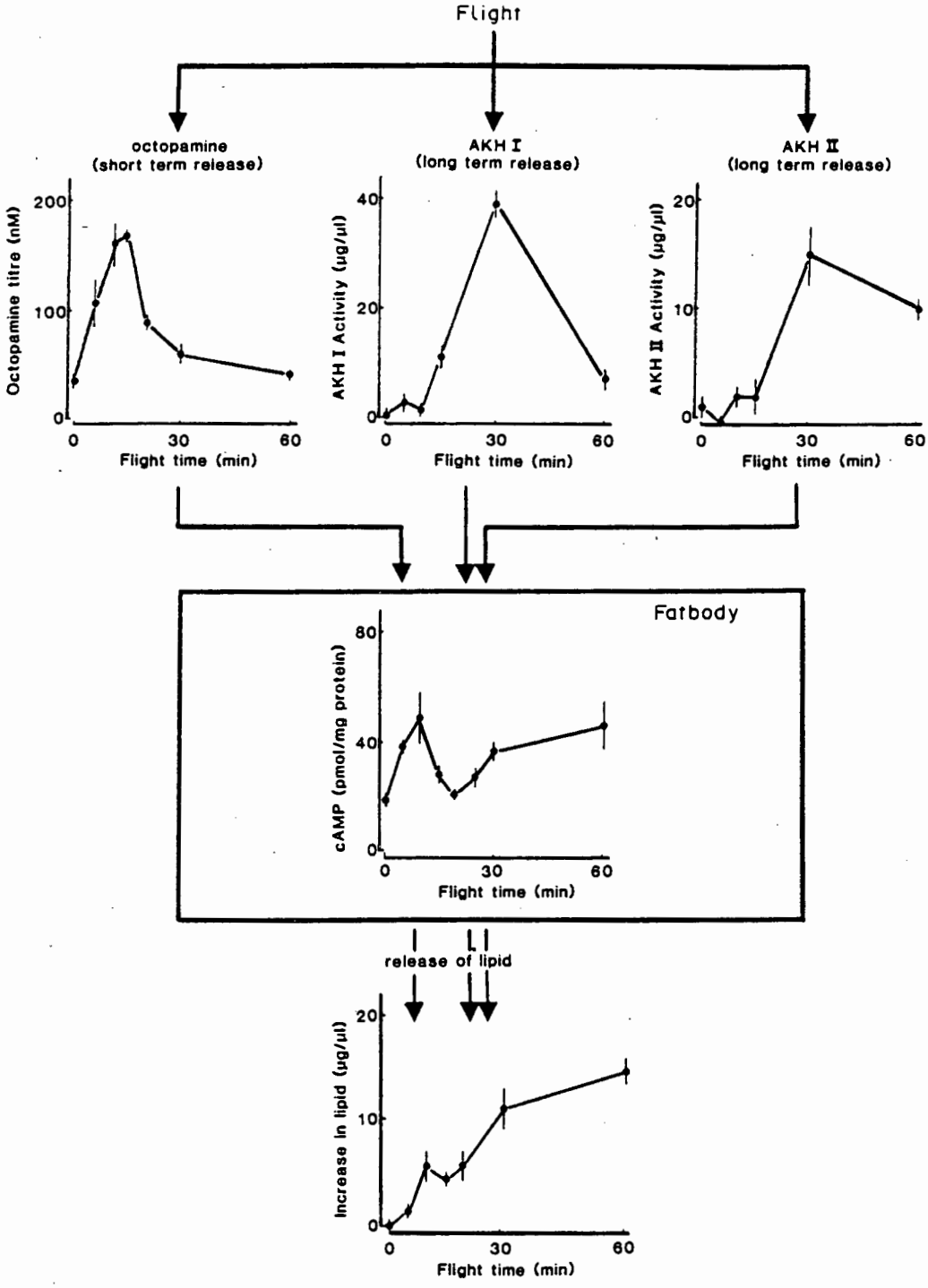


Figure 1.4: Scheme illustrating the hormonal events occurring within the overall flight process. (Taken from Orchard [11].)

first phase of haemolymph lipid increase. The release of AKH after fifteen to twenty minutes of flight induced the second cAMP increase and therefore also the second phase of haemolymph lipid increase.

The above discussion outlines the biological importance of these hormones with respect to the roles played by them in locust flight processes [11].

Since these hormones bind to a membrane receptor [2], their secondary structure may play an important role in their activity. The aim of this thesis was to determine whether Lom-AKH-I, II and III possessed a definite secondary structure, and if so, to see if there was a correlation between the secondary structure and the activity of the hormones. Achieving this would aid in further elucidation of insect flight pathways - specifically with regards to hormone-receptor binding mechanisms. Acquiring an in depth knowledge of these binding mechanisms may facilitate the synthesis of analogues, either agonists or antagonists. Both of these analogue types would be synthesised in such a way as to exhibit a much higher affinity for the receptor than the naturally occurring peptide, with the difference being that, while agonists would induce hyperstimulation resulting in rapid depletion of the substrates, antagonists would lead to inhibition of hormone induced pathways, and energy rich substrates required for flight would not be produced [48]. Because these peptides are to a large extent species specific it would be possible to generate the latter type of hormone analogues as a means of pest control i.e. use them as an effective means of locust flight inhibition. In this way the use of common neurotoxic insecticides, often poisonous to many animal species, could be reduced [5,48].

1.9 Techniques

Four methods commonly used for peptide structure determination are crystallography, circular dichroism (CD), vibrational spectroscopy (Raman, Fourier transform infrared (FTIR) and nuclear magnetic resonance (NMR) [49]. Although, crystallography constitutes a very accurate means of structure determination, it ignores the continuous dynamics of any given molecule within a living cell environment [50]. This can lead to the possible occurrence of significant differences between the crystal and solution structures. Crystallography also

requires the availability of suitable crystals. For the peptides under investigation, crystals were unobtainable and generation of corresponding crystal structures was thus not possible.

CD provides an effective means of identification of secondary structure features such as the existence of helices, β-sheets (parallel or antiparallel) and β-turns within a given protein. CD may be defined as "the difference between the extinction coefficients for left and right circularly polarized light" for a given molecule [51]. Resultant data may thus be positive or negative, with all peak wavelengths corresponding directly to those generated in a normal absorption spectrum [51]. The advantages of using this technique for protein structure determination are, first of all, that studies can be carried out in aqueous environments thus allowing better simulation of the *in vivo* cellular conditions, and secondly, they allow studies of molecular conformational changes associated with changes in conditions such as temperature, pH, salt and solvent composition. The effects of ligand binding and quaternary structure can also be observed [51]. It is the CD associated with the backbone amide groups that provides information regarding secondary structure [51]. Secondary structure determination of an unknown peptide using CD is carried out by fitting a combination of reference spectra to the spectrum of interest. The reference CD spectra are generated from peptides (often synthetic peptides synthesised under specific solvent conditions) exhibiting pure secondary structures [51]. The most successful set of reference CD spectra corresponding to the major secondary structure types viz. the α-helix, β-sheet (antiparallel), β-turn and random coil, was generated by Brahms and Brahms [52] - (Figure 1.5). Although the β-turn CD spectrum is representative of a type -I turn, Brahms and Brahms predicted that this spectrum shows remarkable similarity to those generated for all other β-turn types present in proteins. The random coil spectrum, depending on the residues involved,

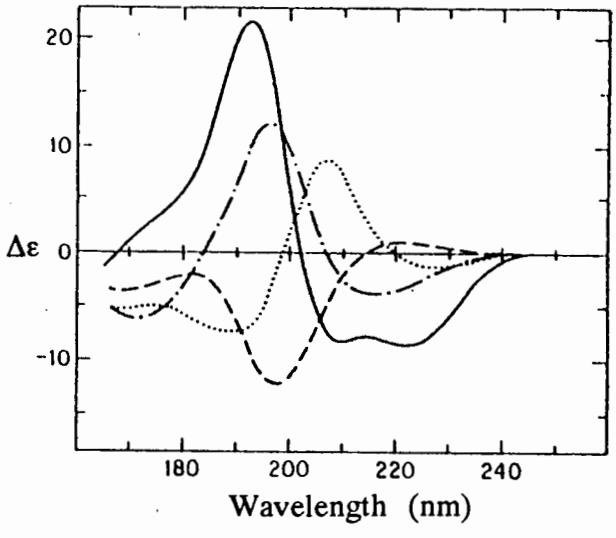


Figure 1.5: The CD spectra obtained by Brahms and Brahms [52] for pure secondary structures relating to an α-helix (solid line), β-sheet (dots and dashes), β-turn (dotted line) and random coil (dashed line). (Taken from Johnson [51].)

can deviate from the above not only in magnitude, but also in sign. The CD spectrum of any given protein can thus be assessed and interpreted in terms of these reference spectra and a reasonable prediction of the secondary structure made. Although these predictions are very accurate for α-helices, further research is required to effectively solve β-proteins. This necessitates the generation of reference spectra corresponding to additional proteins having known β-structure [51]. Other factors which may affect the CD spectrum include twists in the β-sheets and contributions from the aromatic side chains [51].

Regarding small peptides, CD spectra may have impaired quality due, not only, to the higher occurrence of rapid conformational changes often associated with these short peptides, but also, to the interference often created by aromatic side chains which may lead to masking of the true CD signals [49].

Like CD, FTIR provides an effective means of secondary structure determination of proteins in solution, along with the conformational changes experienced on alteration of temperature, pH and pressure conditions. The amide I mode of vibration is generated by the C=O stretching vibration of the amide group, in conjunction with N-H bond in-phase bending and C-N bond stretching [50]. The importance of this particular vibrational mode may be attributed to its sensitivity to hydrogen bonding and transition dipole coupling (interstrand interactions leading to large splitting of the amide I vibrational mode are a distinguishing feature in spectra generated for antiparallel β-structures) [50]. These vibrations occur over a wide range of frequencies (from approximately 1600 to 1700cm^{-1}), which is then sub-divided into regions each corresponding to a particular secondary structural type. These frequency ranges were again determined through the use of reference peptides having pure secondary structure. Because larger proteins often exhibit a combination of all different structural types there can be a large amount overlap within any given FTIR spectrum generated for an unknown [50]. The two ways of solving this overlap problem include band narrowing and pattern recognition [50]. Band narrowing is based on Fourier deconvolution and is sometimes referred to as resolution enhancement. This technique attempts to decrease band width in order to separate out the peaks as much as possible and increase the spectral resolution. The limiting factor in this case is the signal:noise ratio [50]. The pattern recognition technique is based on the construction of a calibration matrix from a series of proteins of known secondary

structure [50]. The most recent method for generation of a calibration matrix was reported by Lee *et al.* [53]. Making use of factor analysis and multiple linear regression, the best results for matrix construction were acquired from the sole use of the amide I vibrational mode in spectra generated for proteins of known structure [50]. One limitation of FTIR is the inability to carry out a quantitative assessment of protein secondary structure. One reason for this is the possible distortion of results by interferences created by absorption of certain animo acid side-chains in the region of interest [50]. Even with its limitations, however, FTIR has been proven to be an effective means for determination of overall protein secondary structures and is particularly useful for observation of subtle backbone conformational changes associated with ligand binding as well as changes in environmental conditions [50].

Finally, Raman spectroscopy has also been shown to be a useful tool in protein secondary structure determination, being applicable to both crystals and solutions [54]. As above with FTIR, it is the amide I bands between 1630 and 1700 cm^{-1}, that are used for determining such information. Initially secondary structure determinations were carried out using spectra from proteins whose secondary structure had been previously determined by X-ray crystallography. These spectra were then used to compute sets of reference intensities with each set being representative of a spectrum corresponding to one or more conformational types [54]. The secondary structure of a given protein was then predicted by fitting the spectrum of that protein to a linear combination of the above reference spectra [54]. The major problem with this method lay with the fact that X-ray crystallography only gave very crude estimates with regards to secondary structure i.e. secondary structure characteristics were categorized according to helix, β-sheets, parallel and antiparallel, and reverse turns. This form of analysis ignores the fact that a large degree of diversity exists within each of these categories. Raman spectroscopy is however very sensitive to small changes in backbone conformation ($1cm^{-1}$ changes are detectable) and thus has the ability to provide more detail with regards to different possible conformation types existing within each of the above four categories. Averaging of the spectra within each class therefore leads to a loss of this additional information in the structure determination procedure [54]. A new method was proposed in 1983 by R.W. Williams in which "the amide I Raman band of a protein is analysed directly as a linear combination of amide I bands of proteins whose secondary structures are known" [54]. This method allows estimations of secondary structures, independent of crystallographic

information relating to the corresponding proteins, to be made, thus generating more accurate information regarding backbone conformations [54].

The last three methods discussed all constitute "low-resolution" spectroscopic methods and serve to provide a more global perspective of the overall secondary structure of a given protein [51], rather than the highly specific information gained from NMR and crystallographic methods.

In this thesis secondary structure determination of Lom-AKH-I, II and III was carried out using NMR restrained molecular dynamics (MD).

REFERENCES

1. Gäde, G. *Zool. Jb. Physiol.* **1992**, *96*, 211-225.

2. Nijhout, H.F. *Insect Hormones*, Princeton University Press, Princeton, New Jersey, 1994, 3-49.

3. Beenakkers, A.M.T. *Endocrinology of Insects*, Alan R. Liss, Inc., New York, 1983, 441-450.

4. Gäde, G. *J. Insect Physiol.* **1990**, *36(1)*, 1-12.

5. Stone, J.V.; Mordue, W.; Batley, K.E.; Morris, H.R. *Nature* **1976**, *263*, 207-211.

6. Carlsen, J.; Herman, W.S.; Christensen, M.; Josefsson, L. *Insect Biochem.* **1979**, *9*, 497-501.

7. Gäde, G. *J. Insect Physiol.* **1984**, *30*, 729-736.

8. Siegert, K.; Morgan, P.J.; Mordue, W. *Biol. Chem. Hoppe-Seyler* **1985**, *366*, 723-727.

9. Gäde, G. Schaffer, M.H.; Cook, J.C.; Rinehart, K.L. *Biochem. Biophys. Res. Commun.* **1986**, *134*, 723-730.

10. Oudejans, R.C.H.M.; Kooiman, F.P.; Heerme, W.; Versluis, C.; Slotboom, A.J.; Beenakkers, A.M.T. *Eur. J. Biochem.* **1991**, *195*, 351-359.

11. Orchard, I. *J. Insect Physiol.* **1987**, *33(7)*, 451-463.

12. Rademakers, L.H.P.M. *Cell Tiss. Res.* **1977**, *184*, 213-224.

13. Rademakers, L.H.P.M. *Cell Tiss. Res.* **1977**, *184*, 381-395.

14. Goldsworthy, G.J.; Johnson, R.A.; Mordue, W. *J. comp. Physiol.* **1972**, *79*, 85-96.

15. Goldsworthy, G.J.; Mordue, W.; Guthkelch, J. *Gen. comp. Endocr.* **1972**, *18*, 545-551.

16. Orchard, I.; Lange, A.B. *J. Insect Physiol.* **1983**, *29(5)*, 425-429.

17. Orchard, I.; Loughton, B.G. *Comp. Biochem. Physiol.* **1981**, *68A*, 25-30.

18. Orchard, I.; Martin, R.J.; Sloley, B.D.; Downer, R.G.H. *Can. J. Zool.* **1986**, *64*, 271-274.

19. Orchard, I.; Loughton, B.G. *J. Neurobiol.* **1981**, *12*, 143-153.

20. Orchard, I.; Loughton, B.G.; Gole, G.W.B.; Downer R.G.H. *Brain Res.* **1983**, *258*, 152-155.

21. Pannabecker, T.; Orchard, I. *Molec. cell Endocr.* **1986**, *48*, 153-158.

22. Goldsworthy, G.J.; Mordue, W. *Biol. Bull.* **1989**, *177*, 218-224.

23. Gäde, G.; Holwerda, H.A. *Insect Biochem.* **1976**, *6*, 535-541.

24. Orchard, I.; Lange, A.B. *J. Insect Physiol.* **1984**, *30(12)*, 901-904.

25. Asher, C.; Moshitzky, P.; Ramachandran, J.; Applebaum, S.W. *Gen. comp. Endocr.* **1984**, *55*, 167-173.

26. Goldsworthy, G.J.; Mallison, K.; Wheeler, C.H. *J. Insect Physiol.* **1986**, *32(1)*, 95-101.

27. Mathews, C.K.; van Holde, K.E. *Biochemistry*, The Benjamin/Cummings Publishing Company, Inc., Redwood City, California, 1990.

28. Raabe, M. *Recent Developments in Insect Neuropeptides*, Plenum Press, New York, 1989.

29. Wheeler, C.H.; Goldsworthy, G.J. *J. Insect Physiol.* **1983**, *29(4)*, 349-354.

30. Goldsworthy, G.J.; Wheeler, C.H. *Insect flight* (Danthanarayana, W.; ed.), Springer-Verlag, Berlin, 1986, 49-59.

31. Robinson, N.L.; Goldsworthy, G.J. *J. Insect Physiol.* **1977**, *23*, 9-16.

32. Goldsworthy, G.J. *Adv. Insect Physiol.* **1983**, *17*, 149-204.

33. Van Marrewijk, W.J.A.; Van den Broek, A.T.M.; Beenakkers, A.M.T. *Gen. comp. Endocr.* **1983**, *50*, 226-234.

34. Gäde, G. *J. Insect Physiol.* **1981**, *27(3)*, 155-161.

35. McClure, J.B.; Steele, J.E. *Insect Biochem.* **1981**, *11*, 605-613.

36. Van Marrewijk, W.J.A.; Van den Broek, A.T.M.; Beenakkers, A.M.T. *Insect Biochem.* **1991**, *21*, 375-380.

37. Van Marrewijk, W.J.A.; Van den Broek, A.T.M.; Van Der Horst, D.J.; Beenakkers, A.M.T. *Insect Biochem.* **1984**, *14*, 151-157.

38. Carlisle, J.; Loughton, B.G. *Nature* **1979**, *282*, 420-421.

39. Carlisle, J.; Loughton, B.G. *J. Insect Physiol.* **1986**, *32*, 573-578.

40. Mwangi, R.W.; Goldsworthy, G.J. *Physiol. Ent.* **1977**, *2*, 37-42.

41. Orchard, I.; Lange, A.B. *J. Insect Physiol.* **1983**, *29(8)*, 639-642.

42. Van Der Horst, D.J.; Houben, N.M.D.; Beenakkers, A.M.T. *J. Insect Physiol.* **1983**, *26*, 441-448.

43. Goosey, M.W.; Candy, D.J. *Insect Biochem.* **1980**, *10*, 393-397.

44. Orchard, I.; Carlisle, J.A.; Loughton, B.G.; Gole, G.W.B.; Downer R.G.H. *Gen. comp. Endocr.* **1982**, *48*, 7-13.

45. Cheeseman, P.; Goldsworthy, G.J. *Gen. comp. Endocr.* **1979**, *37*, 35-43.

46. Van Der Horst, D.J.; Van Doorn, J.M.; Beenakkers, A.M.T. *Insect Biochem.* **1978**, *8*, 413-416.

47. Cheeseman, P.; Jutsum, A.R.; Goldsworthy, G.J. *Physiol. Ent.* **1976**, *1*, 115-121.

48. Gäde, G. *Proc. Conf. Insect Chem. Ecol. Tabor* **1990**, 407-427. Academia Prague and SPB Acad. Publ. The Hague, 1991.

49. Dyson, H.J.; Wright, P.E. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 519-538.

50. Surewicz, W.K.; Mantsch, H.H.; Chapman, D. *Biochemistry* **1993**, *32(2)*, 389-394.

51. Johnson, W.C. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 145-166.

52. Brahms, S.; Brahms, J. *J. Mol. Biol.* **1980**, *138*, 149-178.

53. Lee, D.C; Harris, P.I.; Chapman, D.; Mitchell, R.C. *Biochemistry* **1990**, *29*, 9185-9193.

54. Williams, R.W. *J. Mol. Biol.* **1983**, *166*, 581-603.

CHAPTER 2

NMR OF PROTEINS

Over the past thirty years, nuclear magnetic resonance (NMR) has proved to be an invaluable tool in peptide and protein structure determination [1-10]. The major advances in this field of research occurred with the development of two dimensional techniques which led to vast improvements in dispersion and thus made progress towards solving problems related to overlapping peak assignment - a common feature of peptide spectra. The reliability of NMR as a method for structure determination has been demonstrated by the similarity between the crystal and solution structures obtained for certain proteins - pancreatic trypsin inhibitor (BPTI) [11], α-amylase inhibitor Tendamistat [12,13] and barley serine proteinase inhibitor-2 [14]. This agreement allows valid comparisons to be carried out when the crystalline and non-crystalline structures are different. It also means that if suitable crystals cannot be produced (as is the case with many proteins), a high degree of confidence can be placed in any three dimensional protein structure generated using NMR. In addition, because the peptides are in solution, not only can physiological conditions be simulated more closely, but also various studies can be carried out with regards to the effects of changing conditions such as pH, temperature, and ionic strength [15]. NMR experiments serve to provide information - dihedral angle and distance constraints - which may then be used in subsequent restrained molecular mechanics calculations for generation of low-energy peptide or protein conformers. Certain secondary structure characteristics including helices and β-sheet conformations, can, however, be identified from the NMR data alone (see later), thus allowing general structural predictions to be made before implementing the more complex computational procedures [15].

All two dimensional pulse sequences possess the same general features viz. a preparation, an evolution and a detection period [16]. The preparation period constitutes a time to allow the

system to return to thermal equilibrium or a steady state and then "prepare" it for the evolution period. This is done by applying a radio frequency pulse (usually 90°), or series of pulses [16], which force the magnetisation into the transverse x-y plane, thereby inducing phase coherence (single quantum coherence) between the α and β energy states (Figure 2.1) [17]. During the evolution period, t_1, the coherence is allowed to evolve. Not only do the magnetisation vectors acquire differing phases in accordance with their chemical shift differences, but they also split due to J coupling [18]. A second pulse, sometimes referred to as a mixing pulse, then causes transfer of the single quantum coherence of a given proton to all other coherences available within a given spin system, i.e. coherence transfer is the result of redistribution of magnetisation amongst the coupled nuclei of a given spin system, and is responsible for generation of crosspeaks in the resultant spectrum. The extent of coherence transfer is dependent on the length of the evolution period, t_1, the coupling constant, J, and the duration of the second pulse [17]. This second pulse may then be followed either by a complex series of pulses and delays, constituting a mixing time, or, in the simplest case (that of the COSY experiment), directly by the detection period, t_2, when recording of the FID occurs [16].

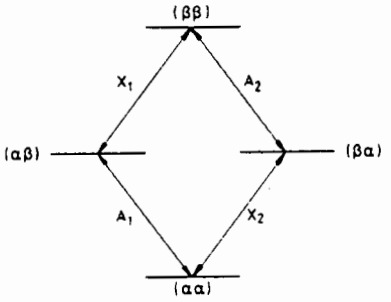


Figure 2.1: The energy levels of a homonuclear AX system. (Taken from Derome [17])

The basic theory behind 2-D NMR lies in the incrementation of the evolution time period, t_1, in the relevant pulse sequences [18]. By incrementing t_1 and recording the corresponding FID's, it is possible to generate an $s(t_1,t_2)$ data matrix, which will yield the 2-D frequency spectrum $S(\nu_1,\nu_2)$, in which ν_1 and ν_2 are correlated [17]. Extensive variations in the above simplified sequence give rise to the many unique pulse sequences available today, many of which generate spectra differing in informational content.

Although it is not discussed here, it should be noted that the use of 3D NMR [19] for large protein structure determination is becoming popular as a means of solving the extensive peak overlap often witnessed in the 2D spectra of these proteins. For small peptides like those of the AKH family, however, 2D spectra show sufficient resolution to solve the overlap in the 1D proton spectra, and the use of 3D NMR is therefore unnecessary for the current problem.

Using NMR as a means of solving peptide structures (secondary or tertiary) requires accurate identification and assignment of the relevant spin systems. This may be effectively carried out using COSY (**CO**rrelation Spectroscop**Y**) [20,21], TOCSY (**T**otal **CO**rrelation Spectroscop**Y**) [22,23], NOESY (**N**uclear **O**verhauser and **E**xchange Spectroscop**Y**) [24-27] and/or ROESY (**R**otating frame **N**uclear **O**verhauser and **E**xchange Spectroscop**Y**) [28-31] experiments in conjunction with a series of other experiments depending on the problem at hand. COSY and TOCSY rely on coherent transfer through scalar coupling while NOESY and ROESY rely on incoherent transfer (the result of transition saturation) which requires through space or dipolar coupling [16,17].

2.1 Through-bond correlation

A variety of 2D-experiments have been developed to allow identification of *J*-coupled protons, the most well known of which include the COSY and TOCSY experiments. The unique spin systems of many of the amino acid residues generate characteristic correlation patterns which can be readily identified in these spectra. In the COSY experiment, crosspeak generation is limited to ^{2}J and ^{3}J-coupling, thus allowing sequential intraresidue assignments to be made [15]. The TOCSY is, as its name implies, designed to give complete correlation between all the protons of a given spin system, i.e. additional coupling generates crosspeaks between protons which are not directly coupled but share common coupling partners [32]. These two experiments used in conjunction with each other are essential for initial peak assignments of the protein ^{1}H spectrum. Limiting features of the COSY spectrum include firstly that the diagonal and crosspeaks in the spectrum are out of phase by 90° in both dimensions and therefore require extensive weighting to improve crosspeak lineshapes [33]. Secondly, the components of a multiplet are in antiphase, which means that if coupling is less than the peak linewidth, self-cancellation can result in reduced peak intensities [33,34]. In an absolute value COSY spectrum crosspeaks tend to adopt a star shaped appearance with the "tails" resulting from the dispersive mode components of the peak. In isolation these peaks are effective in producing the relevant coupling information. However, as soon as there is overlap, the problem of distinguishing peaks from each other is complicated by the fact that overlap between two peaks can generate a ridge with higher intensity than either contributing peak. This can be misleading, particularly if the threshold is set at an intensity higher than

that of the original peaks. The dispersive "tails" also prevent accurate spectral interpretations in the region of the diagonal to be carried out [34].

An alternative to the absolute value COSY is the phase sensitive COSY [35,36] produced by collecting two FIDs differing in phase by 90° for each t_1 period, and combining each pair to generate a single complex signal [16]. This experiment allows the crosspeaks to be purely absorptive, thus giving rise to improved resolution and sensitivity. However, if the crosspeaks are purely absorptive, then the diagonal will be purely dispersive, sometimes to the extent of obscuring crosspeaks located near the diagonal [15,16]. This experiment, by increasing resolution of multiplet components, generates information regarding the coupling patterns as well as the coupling constants, and has gained significant popularity in recent years [15]. By introducing multiple quantum filters, it is possible to simplify the spectrum and improve this phase sensitive COSY experiment [17].

Double quantum filter (DQF)-COSY [37,38] experiments simplify the spectrum, particularly the diagonal region, through the elimination of singlet peaks from the spectrum [17]. Other advantages are that the intensities of the diagonal peaks are reduced to more or less the same as those for the crosspeaks [37], the spectrum shows significantly improved peak lineshapes [33], and the diagonal peaks and crosspeaks can now both be purely absorptive in the spectrum (although the diagonal peaks of systems with more than two spins do still contain some dispersive contributions, this occurs to far lesser extent than before) [17]. Triple quantum filter (TQF)-COSY [39] experiments exclude AB and AX systems from the spectrum thus leading to even greater spectral simplification [40]. The major problem experienced with all of these multiple quantum methods is the significant decrease in peak intensity in these spectra [40,37].

Certain amino acids such as Arg, Lys, Leu and Pro have long side chains which can prove to be problematic. Either they generate regions of extensive overlap (mostly in the region close to the diagonal) which often cannot be resolved using only the COSY experiment, or they generate such low intensity complex multiplets that they are absent from the spectrum [32].

These problems may be solved by carrying out a TOCSY rotating frame (spin lock) experiment in which the usual mixing pulse is now replaced by a mixing period. This mixing period, referred to as isotropic mixing [37], is made up of a series of pulses applied in such a way as to momentarily eliminate the effects of chemical shift, thereby allowing rapid distribution of coherences between scalar coupled nuclei throughout a given spin system [33,37]. The advantages of the TOCSY experiment are that all the crosspeaks are in-phase, and both the crosspeaks and the diagonal can be displayed in absorption mode. The former allows for superior resolution of low intensity multiplets, particularly ones in which small couplings previously led to self-cancellation due to their antiphase nature in the COSY spectrum, while the latter generates peaks with improved lineshape, and therefore also leads to increased sensitivity [33]. If a sufficiently long mixing period is used, it is possible to generate a spectrum in which all the coupling within any given spin system is observed [32]. The longer the mixing period, the greater the extent of longitudinal relaxation, and the lower the intensity of the crosspeaks. In practice, a compromise has to made between signal to noise ratios and the number of coherence transfers.

Further problems are, however, also experienced in differentiating the Asn, Asp, Cys, His, Phe, Ser, Trp and Tyr residues from each other. All of these amino acids have AMXY spin systems (homonuclear couplings to the rest of the side chain are negligible), and are therefore not clearly distinguishable from each other in the COSY and TOCSY spectra [32]. It is possible to distinguish the Ser residue from the others because of the small Hβ-Hβ' coupling [32], as well as through the presence of HO-Hβ and HO-Hβ' crosspeaks (the presence of these peaks is dependent on the rate of exchange of the mobile hydroxyl proton with the solvent). The β-protons of amino acids containing aromatic systems i.e. His, Phe, Trp and Tyr, tend to have slightly further downfield chemical shifts due to the extensive delocalisation within the ring system which leads to greater deshielding of the adjacent β-protons. This information is only of real use when dealing with small peptides that contain a very limited number of these AMXY spin system amino acid types. In order to conclusively distinguish these residues from each other, further experiments revealing dipolar coupling information, are required.

Another possibility is the use of heteronuclear experiments such as heteronuclear correlated

spectroscopy (HETCOR) [34] and heteronuclear multiple quantum correlation (HMQC) [41,42] experiments. These spectra serve to correlate shifts between protons and directly bound heteronuclei [32]. Such experiments have proved to be very useful for solving overlap problems in crowded homonuclear spectra due to the chemical shift dispersion associated with heteronuclei. These experiments require either uniform ^{13}C and ^{15}N labelling of the peptides or extremely high solution concentrations [33].

2.2 Through-Space Correlation

The above methods were useful for identification of the spin systems of individual amino acid residues. The next step in the process is to identify the sequence of the residues within the peptide or protein through the generation of spectra showing through-space correlation between protons, i.e. dipolar coupling resulting from build-up in nuclear Overhauser effects (nOe) [32]. Short and medium range nOes occurring between protons within five residues of each other are important for sequential interresidue assignments leading to both primary sequence [43] and regular secondary structure determination. Thereafter, the longer range interactions become important for tertiary structure predictions to be made. The most common experiments used for measuring these dipolar interactions are the NOESY and ROESY experiments.

NOe originates from the change in intensity experienced by a particular NMR signal on irradiation of a different one, and is given by the following expression:

$$\eta_i(s) = \frac{(I-I_0)}{I_0} \qquad (2.1)$$

where I_0 is the equilibrium intensity i.e. the intensity without irradiation, I is the measured intensity after irradiation, s is the nucleus being irradiated and i is the nucleus experiencing the intensity change [17]. Depending on whether the resultant intensity is greater or less than the initial intensity, the nOe measured may be positive or negative respectively. In order to understand the origin of this effect it is necessary to consider the different pathways available for relaxation. Consider an AX system in which no scalar coupling is observed. This system

will have four spin states viz. αα, ßß, αß, and ßα. Because chemical shifts are small relative to Larmor frequencies, the difference in energy between the *s* and *i* transitions can be considered to be negligible, and the energies of the αß and ßα levels to be degenerate (Figure 2.2) [17]. Perturbation of the system leads to a change in the population of the energy levels. Relaxation will then occur in order to return the system back to equilibrium. The perturbation may be carried out through saturation of, say, the *s* transitions resulting in a zero population difference between these energy levels. Saturating *s* here does not affect the intensity of the *i* transition, since only single quantum transitions are present as relaxation pathways [17]. Figure 2.3 shows the initial direction of cross relaxation directly after saturation of *s*. Both W_1^s and W_1^i are irrelevant - the former due to the fact that the population difference is being controlled in the saturation process, and the latter because the population difference remained unchanged even after irradiation. Of importance are W_0 and W_2, both of which show changes in population difference after irradiation. The population difference of W_0 before irradiation was zero, and after was δ, and for W_2 before was 2δ, and after was δ (Figure 2.3). To re-establish equilibrium, population transfer will either occur from ßα to αß or from ßß to αα. The former results in a decrease in population at the bottom of the one transition and an increase at the top of the other. This leads to an overall decrease in the intensity of the *i* transition, assuming of course that $W_0 > W_1^i$ (the counteractive relaxation process trying to maintain the equilibrium situation), and thus results in a negative nOe. The latter, in turn, results in a decrease in population at the top of one transition and an increase at the bottom of the other. This leads to an overall increase in intensity of the *i* transition, assuming $W_2 > W_1^i$, and therefore generates a positive nOe [17].

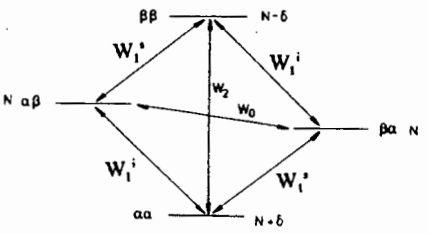


Figure 2.2: Energy levels and populations of a homonuclear AX system. Also indicated are the various relaxation pathways within the system. (Taken from Derome [17])

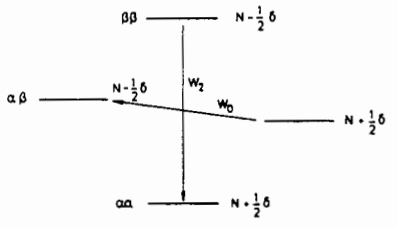


Figure 2.3: The initial direction of cross relaxations directly after saturation of the s transition. (Taken from Derome [17])

W_0 and W_2 are cross relaxation processes resulting from

dipolar or through space coupling. In order to relate dipolar coupling to the longitudinal relaxation time, T_1, it is necessary to consider the motion of the molecule. Correlation time, τ_c, is a measure of the lifetime of any particular conformation of the molecule, assuming that the change from one conformation to another is instantaneous. Large molecules will undergo slow motion and so have high values of τ_c, while smaller molecules will, in general, undergo rapid molecular motion and so have small τ_c. τ_c is chosen not as an average time, but as an outer limit, i.e. it is chosen as a minimum value such that conformational lifetimes shorter than τ_c seldom occur. From a plot of spectral density, $J(\omega)$, versus frequency for different values of τ_c, it is possible to show that an inverse relationship exists between τ_c and T_1 (Figure 2.4) - as τ_c decreases, the number of molecules with specific frequencies corresponding to those of the fluctuating magnetic field decreases, and therefore the relaxation time increases. The transition frequency of the i relaxation pathway (W_1^i) together with the frequencies of the cross relaxation pathways W_0 and W_2 are central to the concept of nOe. The energy of the W_2 process will be the greatest and that of the W_0 process the least i.e. $W_0 < W_1^i < W_2$. Looking at Figure 2.5 it may readily be seen that in going from a smaller to a larger molecule, W_2 will move over the edge and diminish, thereby causing the W_0 relaxation process to dominate in this case. This explains the observed change from positive to negative nOe in going from a small to a large molecule. (For small molecules W_2 is the dominant relaxation pathway and nOe generated is **positive**, while for large molecules W_0 is dominant and a **negative** nOe is now generated.) At some particular molecular size $W_0 \approx W_2$, and the corresponding nOes of these molecules approach zero [17], a problem which will be discussed later in this chapter.

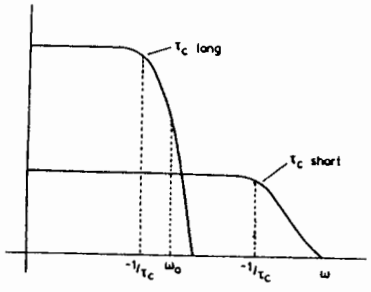


Figure 2.4: Spectral density functions for two different values of τ_c showing the inverse relationship between τ_c and T_1. (Taken from Derome [17])

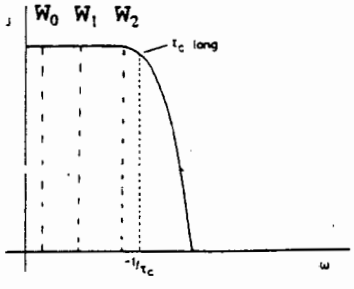


Figure 2.5: The effect of going from a small molecule to a large molecule will cause W_2 to move over the edge thereby inducing dominance of the W_0 process. (Modified from Derome [17])

2.3 The Isolated Spin Pair Approximation (ISPA)

The NOESY pulse sequence consists of three 90° pulses - the incremented t_1 evolution period lies between the first two, with the second being followed by the mixing period, τ_m, which is kept constant throughout the experiment [43]. The mixing period is the time during which cross relaxation (magnetisation transfer between proximal nuclei) inducing build-up of nOe occurs [16]. The nOe intensity may be expressed as follows:

$$NOE \propto \frac{1}{\langle r^6 \rangle} . f(\tau_c) \qquad (2.2)$$

i.e. as a function of internuclear distance, r, and the correlation time, τ_c [44]. Short mixing times limit the amount of nOe build up to shorter internuclear distances - 3 to 3.5Å, while longer mixing times generate nOes for interproton distances of up to 5Å [43]. The ISPA (**I**solated **S**pin **P**air **A**pproximation) is based on the approximation that cross peak intensity is indirectly proportional to the sixth power of the distance separating two protons:

$$r_{ij} = r_{ref} \left(\frac{a_{ref}}{a_{ij}} \right)^{1/6} \qquad (2.3)$$

where a^{ij} is representative of peak intensity. Using the ISPA it should therefore be possible to determine unknown internuclear distances making use of a reference peak and its corresponding interproton distance, for example the peak and distance corresponding to geminal or aromatic ring protons [45,46]. There are, however, significant problems associated with the direct use of this equation.

The first problem is related to the fact that τ_c values are not necessarily constant for different locations within the molecule or protein [44], and reported values are averages of all the internal motions within the molecule. In the case of large molecules the condition of $\tau_c\omega >> 1$ is true and the rate of nOe build-up increases linearly with τ_c. However for smaller molecules in the case $\tau_c\omega < 1$, the nOe build-up is opposite in sign to the diagonal. The problem now lies with the fact that in macromolecules there can exist pockets of fast local motion, i.e. pockets

having lower τ_c values, which lead to reductions in nOe build-up rates in such molecules [43,47].

The second problem relates to spin diffusion effects which result in diminished nOe intensities due to the competitive relaxation pathways present in any multispin system. Spin diffusion effects become particularly problematic at longer mixing times. For this reason the ISPA is only appropriate for shorter mixing times. One major problem, however, associated with short τ_m values is low signal to noise ratios resulting from insufficient nOe build-up. At these low ratios there is also the added possibility of peak intensities being affected by spectral artifacts and zero-quantum coherence. It is in fact more accurate not to consider the absolute nOe generated for a given τ_m, but rather to relate the rate of nOe build-up to internuclear distances through generation of nOe versus τ_m plots where the slope represents the build-up rate. This is only possible where mixing times used are short enough such that deviations from linearity due to spin diffusion effects do not occur [48].

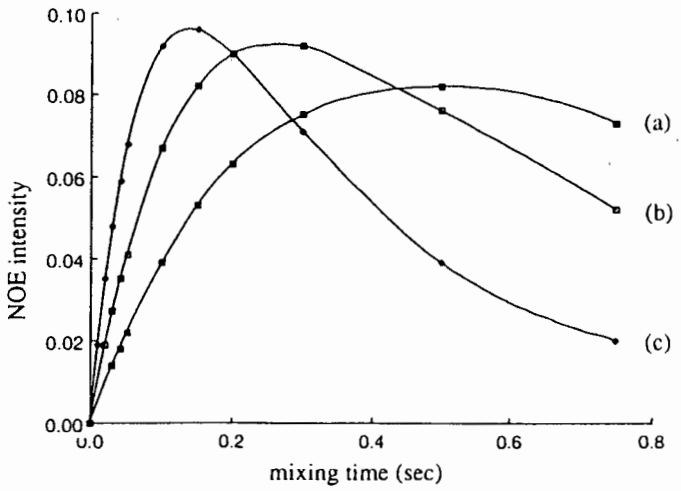


Figure 2.6: Effect of increasing τ_m and τ_c ((a) 2ns, (b) 4ns and (c)8ns) values on the nOe intensity of the crosspeak generated for the δH-εH interaction in the aromatic ring of Tyr68 of dihydrofolate reductase. (Taken from Roberts [46])

The effect of mixing and correlation time on the nOe intensity is shown in Figure 2.6. From the figure it can seen that the linear relationship between nOe intensity and τ_m, necessary for accurate application of the ISPA, only exists at short mixing times. As soon as longer mixing times are used the nOe intensities are seen to decrease - the result of indirect nOe/spin diffusion effects. Also illustrated is the relationship between τ_c and nOe intensity. As τ_c increases the region of linearity applicable to the ISPA decreases as does the τ_m value at which maximum nOe intensity is observed [46].

The problem therefore is that the short mixing times required for the ISPA lead to signal to noise ratios that are insufficient for accurate peak measurements. In general the ISPA tends to overestimate short distances and underestimate long distances, with the error increasing with increasing τ_m and τ_c values [46].

In order to deal with the problem of limited τ_m values, use is made of a relaxation rate matrix approach which generates theoretical NOESY data sets which may be compared to the experimental data sets (see later - chapter 3) [46].

2.4 NOESY versus ROESY

Whereas nOe is the result of longitudinal magnetisation, rOe results from transverse magnetisation. The ROESY experiment is based on generation of homonuclear nOes under spin locked conditions. The advantage of doing this is that spin locked nOes are always positive, and therefore the molecular correlation time ceases to be of importance, i.e. the problem of small or zero nOes in the region of $\tau_c\omega=1$ is no longer applicable. Instead, a steady increase in spin-locked nOe with increasing τ_c is observed [29,49,50]. ROESY experiments are therefore advantageous for generation of dipolar coupling information for medium sized (MW=1000) molecules, where NOESY results show near zero nOe measurements, or even a combination of positive and negative nOe which is useless in terms of subsequent distance determination calculations [32]. In the ROESY experiment spin diffusion is less important than in the NOESY experiment due to indirect nOe effects being opposite in sign to the direct relaxation pathways [32]. This therefore leads to a decrease in nOe intensities for longer distances meaning that these distances are now overestimated and not underestimated as before in the NOESY experiment. Thus when using the ROESY experiment, only the lower bound distance constraint need be adjusted when taking spin diffusion effects into consideration. The rOes are also influenced to a far lesser extent by intramolecular inconsistencies in τ_c than the corresponding nOes [32].

The disadvantage of the ROESY experiment lies in occurrence of spurious crosspeaks in the spectrum resulting from coherent transfer between scalar coupled nuclei of which there are two types - the Hartmann-Hahn(TOCSY)-type and the COSY-type. Although most of the time

these artifacts are readily identifiable due to the fact that they are opposite in sign/phase to the "real" ROESY peaks, two-step ROESY-TOCSY or TOCSY-ROESY transfer can also occur where the spurious peaks are now in the same phase [32]. A second disadvantage is that the ROESY peaks are less intense than in the corresponding NOESY, and that they often have variable intensities, which need to be corrected before distance calculations can be performed [32,51].

Thus the decision regarding the use of ROESY or NOESY experiments to gain dipolar coupling information should be based on the specific nature of the problem to be solved.

2.5 Measurement of nOe Intensity

Measurement of NOESY peak intensities in 2D spectra relies on calculation of peak volume and is thus dependent on factors such as lineshape, the window functions applied during processing, and peak overlap. Peak volume, V, determination can be carried out in a variety of ways. A very coarse method involves a simple estimation of the volume from the peak height or amplitude of the maximum [46]. This may be done directly or by counting the number of contours [46]. The obvious disadvantage of this method is the assumption that lineshapes and linewidths are identical for all crosspeaks. This leads to an overestimation of narrow peak and underestimations of broad peak intensities [46]. In cases where the digital resolution is high the actual lineshape becomes an important factor in the volume calculations. It is then necessary to approximate a geometry factor, k, which is estimated to be proportional to the contour area at half height i.e.

$$V = kI_{max} \tag{2.4}$$

where I_{max} is the maximum amplitude [46].

An alternative means of volume calculation is through numerical integration methods, one of which is equivalent to amplitude summation over a specified region around the maximum [46]. The problem with this approach is its sensitivity to selection of the integration region as well as to spectral artifacts and peak overlap [46].

Assuming that crosspeaks have identical lineshapes but different linewidths, peak volume becomes the product of the amplitude and linewidths in both dimensions, ω_1 and ω_2, as follows [46]:

$$V = k\Delta_1\Delta_2 I \tag{2.5}$$

where I again is the amplitude and Δ_1 and Δ_2 are the linewidths corresponding to ω_1 and ω_2 respectively. The advantage of this method is its decreased sensitivity to peak overlap relative to the previous method. Errors associated with this algorithm arise from the fact that lineshapes are affected in multiplet peaks. This problem can, however, be overcome to a certain extent by introducing line broadening [46].

Another method which is effective, again for reasonable estimation of overlapping peak volumes, involves fitting the rows and columns of the 2D NOESY spectrum to a set of reference lines. The reference lineshape of a given resonance is defined in both dimensions by using that part of a row or column which corresponds to a well resolved unambiguously assigned cross-peak of that resonance [46]. The limitation of this method lies with the generation of sufficient reference lines, which is a computationally demanding, and therefore time consuming, process [46].

Finally there is a method in which the 2D peak is defined in terms of the row and column in the spectrum which correspond to a specified amplitude, A_{ij} (if not the maximum amplitude then another point of superior resolution in the cross-peak) [46]. Generation of the 1D integrals of the specified row and column then allows calculation of the peak volume as follows [46]:

$$V = I_1 I_2 / A_{ij} \tag{2.6}$$

where I_1 and I_2 correspond to 1D row and column integrals (peak areas), respectively. In the event of overlap, integral determination can be carried out using curve fitting based on theoretical lineshapes [46]. This method, however, has the disadvantage of being slow due to the amount of manual interaction required [46].

A general equation which may be used for calculation of the error in the peak volume is given below:

$$\Delta V = N + \varepsilon V \tag{2.7}$$

where N is the spectral noise level and ε, the error, associated with the method used for volume calculation [46]. In all volume calculations, it is important to estimate the error associated with the measurement.

2.6 NOe and Secondary Structure Determination

The dipolar coupling between nuclei which can be measured via the NOESY pulse sequence provides a good basis for prediction of the secondary structure of peptides and proteins. The interactions of relevance include both those between backbone protons and those between the backbone and β-sidechain protons. Any two protons generating a crosspeak in the NOESY spectrum must lie within a 4.5 - 5Å distance of each other. Such interactions can be divided into three different categories - short-range or sequential interactions which occur between adjacent residues, medium range interactions which are generated by dipolar coupling between protons separated by not more than five residues, and finally long range interactions which occur between protons separated by six residues or more. To describe the distance between, say, the α-proton of one residue and the amide proton of the next, the following notation is generally used: $d_{\alpha N}(i,i+1)$ - Figure 2.7, where i refers to one residue and (i+1) to the adjacent residue. A residue two away is referred to as (i+2) etc [15].

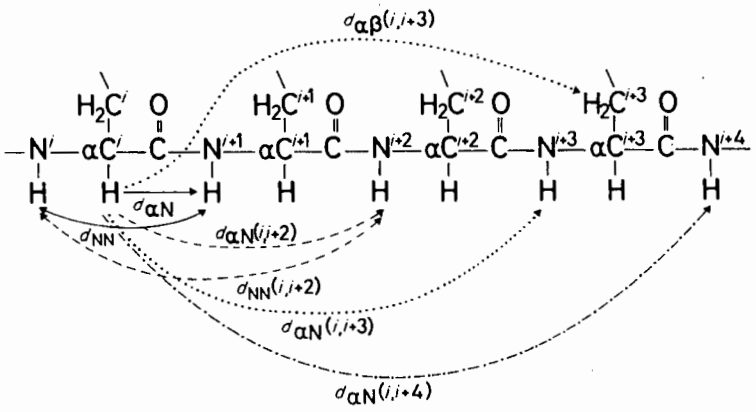


Figure 2.7: Short and medium range connectivities useful for secondary structure determination. (Taken from Wüthrich [15])

Predictions regarding regular secondary structure are based on the occurrence of a variety of

short ^1H-^1H distances generating nOe signals in the NOESY spectrum (Table 2.1).

Table 2.1: The short sequential and medium range interproton distances expected for different regular secondary structure types. (Taken from Wüthrich [15])

Distance	α-helix	3_{10}-helix	β-sheet (antiparallel)	β-sheet (parallel)	Type-I turn[a]	Type-II turn[a]
$d_{\alpha N}(i,i+1)$	3.5	3.4	2.2	2.2	3.4 3.2	2.2 3.2
$d_{\alpha N}(i,i+2)$	4.4	3.8			3.6	3.3
$d_{\alpha N}(i,i+3)$	3.4	3.3			3.1-4.2	3.8-4.7
$d_{\alpha N}(i,i+4)$	4.2					
$d_{NN}(i,i+1)$	2.8	2.6	4.3	4.2	2.6 2.4	4.5 2.4
$d_{NN}(i,i+2)$	4.2	4.1			3.8	4.3

[a]The first of the two numbers applies to the distance between residues two and three, the second to that between residues three and four (see figure 2.8). The range indicated for $d_{\alpha N}(i,i+3)$ corresponds to the distances adopted if ψ_1 is varied between -180 and 180°.

The criteria for helix formation include relatively strong $d_{NN}(i,i+1)$ and $d_{\alpha N}(i,i+1)$ nOe connectivities (the former being stronger than the latter). The α-helix type is then characterised by strong $d_{\alpha N}(i,i+3)$ and $d_{\alpha N}(i,i+4)$ connectivities and is readily distinguishable from the 3_{10} helix type which shows strong NH-αH interactions between the i and (i+2) and i and (i+3) residues [15,52].

β-sheet conformations are indicated by strong sequential $d_{\alpha N}(i,i+1)$ nOes along with several interstrand interactions which are essential for confirmation of a β-sheet structure. These include $d_{\alpha N}(i,j)$, $d_{NN}(i,j)$ and $d_{\alpha\alpha}(i,j)$ connectivities (where j is applicable to a second strand) [52]. Parallel and anti-parallel β-structures are distinguishable on the grounds of different strengths of the nOe intensities of the NH-NH and αH-αH interactions. Both, in particular the latter, generate far stronger interactions in antiparallel than in parallel β-sheets (Figure 2.8) [15].

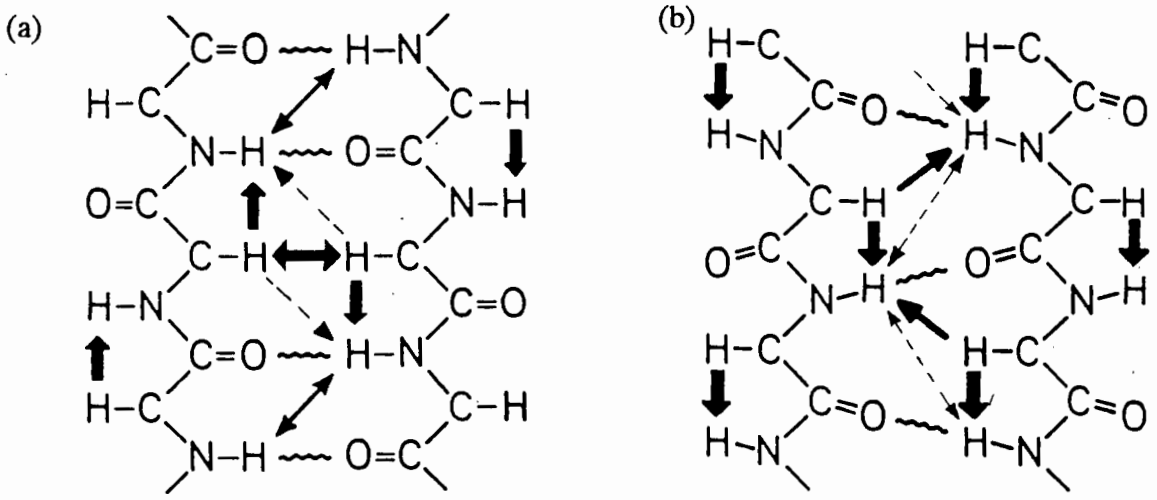


Figure 2.8: Short and medium range backbone interproton distances in β-sheets. Wavy lines indicate hydrogen bonds and the thick vertical lines $d_{\alpha N}(i,i+1)$. (a) Antiparallel β-sheets - interstrand interactions are indicated by thick horizontal arrows, $d_{\alpha\alpha}(i,j)$, thin solid arrows, $d_{NN}(i,j)$ and broken arrows, $d_{\alpha N}(i,j)$. (b) Parallel β-sheets - interstrand interactions are indicated by thick horizontal arrows, $d_{\alpha N}(i,j)$, and broken arrows, $d_{NN}(i,j)$. (Taken from Wüthrich [15])

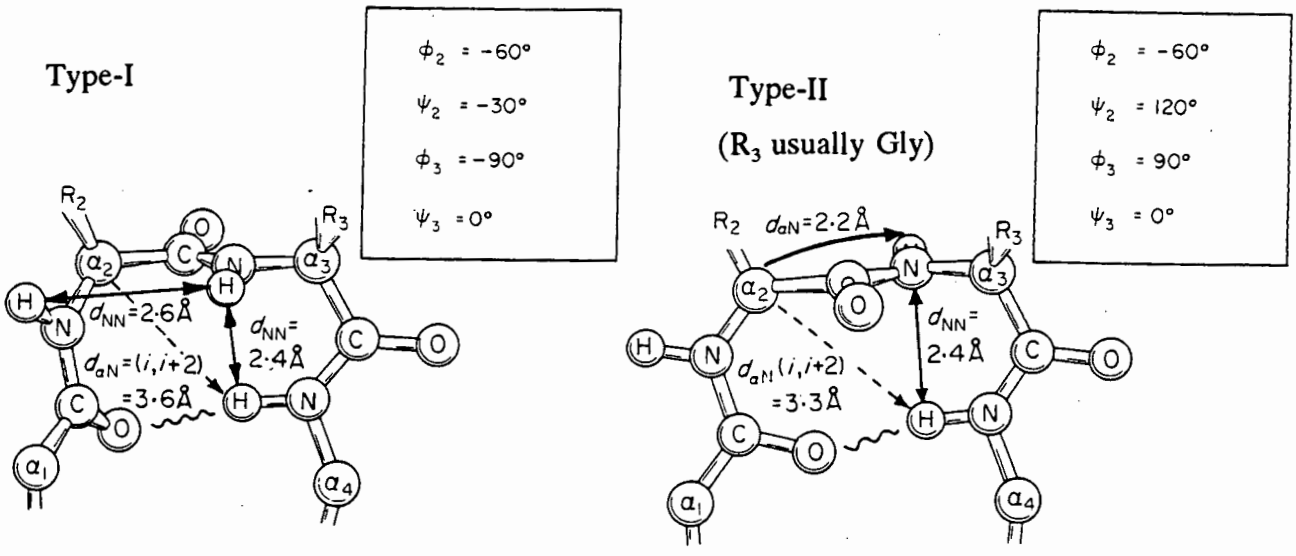


Figure 2.9: Short and medium range interproton distances expected for type-I and type-II tight turns. The wavy lines represent hydrogen bonds. (Taken from Wüthrich [15])

Criteria suitable for the formation of tight turns are more difficult to generate due to the large number of known variations. In general a turn region will contain four residues. Most turns will show $d_{NN}(i,i+1)$ and $d_{\alpha N}(i,i+1)$ nOe connectivities between residues three and four and $d_{\alpha N}(i,i+2)$ nOe connectivities between residues two and four. Although $d_{\alpha N}(i,i+1)$ nOe connectivities between residues two and three are observed in both turn types, the interaction is significantly stronger in a type-II turn. A type-I turn will show an additional NH-NH interaction between residues two and three (Figure 2.9) [15,52].

The possibility also exists, particularly for small peptides, that, although a definite structure may be generated, it does not exhibit the characteristics of any regular secondary structure. Such peptides are then referred to as being unstructured or as having a random coil configuration [52].

2.7 Measurement of J Couplings

Three bond scalar coupling can be related to torsion angles through the Karplus equation, the general form of which is given by:

$$J(\theta) = A\cos^2\theta - B\cos\theta + C \qquad (2.8)$$

where A, B and C are dependent on the specific dihedral angle being determined; for example, A, B and C are 6.4, 1.4 and 1.9 respectively for the HN-C_αH torsion angle. The angle θ, therefore refers to the angle between the hydrogens, where values of 0° and 180° will refer to *cis* and *trans* arrangements respectively. The equation $\theta = |\phi - 60°|$ allows for the subsequent calculation of the ϕ torsion angle [43]. Determination of $^3J_{\alpha N}$ coupling constants can very rarely be obtained directly from the proton spectrum. The possibility of using a phase sensitive COSY experiment is limited not only by the complex

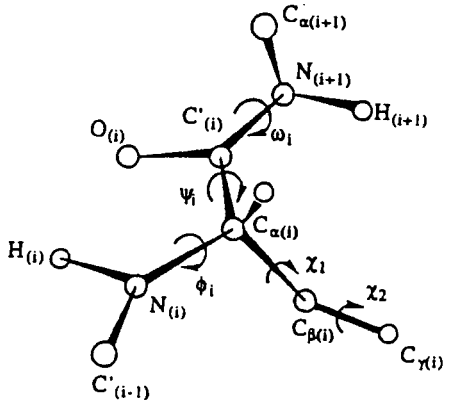


Figure 10: Standard nomenclature for atoms and torsion angles along the peptide chain. (Taken from Roberts [46])

multiplets generating extensively overlapped crosspeaks, but also by the significant overestimation of coupling constants in cases where the linewidth exceeds the *J* coupling [43]. The development exclusive correlation spectroscopy (E-COSY) [53-55] and PE-COSY [56] experiments, where coherence transfer is restricted to transitions directly connected in the energy level diagram, has resulted in significant simplification of the COSY multiplets thus facilitating coupling constant determination in multiple spin systems [53]. E-COSY experiments are superior to DQF-COSY experiments due to the former exhibiting increased resolution of multiplets in the region of the diagonal [53]. Another advantage of E-COSY or PE-COSY experiments is their use in calculation of $^3J_{\alpha\beta}$ coupling constants which are then used for determination of χ_1 dihedral angles which generate information regarding the orientation of given amino acid side chains with respect to the backbone. Where only one β-proton is present, calculations involving the Karplus equation are simple, however, in the presence of two β-protons there exist three different rotamers each corresponding to a different value of χ_1, but each containing a characteristic set of coupling constants [57].

Another possibility for the determination of $^3J_{\alpha N}$ coupling constants is using a homonuclear 2DJ-Resolved (Homo-2DJ) experiment which separates the chemical shifts and coupling constants through introduction of a 180°$_x$ pulse in the middle of t_1. The F_1 dimension contains the multiplicity and coupling constants, and the F_2 dimension the chemical shifts [18].
The importance of $^3J_{\alpha N}$ coupling constant determination resides with making predictions regarding possible secondary structures adopted by the relevant peptide or protein in solution. Relatively small coupling constants ranging from ~3-5Hz indicate helical (α or 3_{10} helices) or turn formations, while larger coupling constants ranging from ~9-12Hz indicate a β-structure (parallel or anti-parallel β-sheets). Coupling constants lying between these extreme values indicate that the peptide is either not constrained in a regular secondary structure conformation, or that a certain amount of conformational averaging is taking place [57].

2.8 Solvent Suppression

Many NMR experiments involving peptides and proteins are carried out in an aqueous environment - 90%H_2O:10%D_2O [37]. The major problem with this approach is the generation of a dominant spectral peak resulting from the high proton concentration of the solvent. The

high intensity of the solvent peak leads to a problem of dynamic range. The first aim of solvent suppression is therefore to decrease the size of the NMR signal entering the receiver coil to below that of the analogue-to-digital converter (ADC) thus ensuring that all solute peaks are detectable. The second use is to decrease the diffuseness of the solvent peak mainly in the F1 dimension of 2D spectra [37].

There are two main methods of solvent suppression - the first and most effective is presaturation, and the second includes solvent non-excitation methods. The limitation of the former method lies in its only being suitable for solvent suppression if chemical exchange between the solvent and sample protons either does not occur or occurs at a rate slower than their longitudinal relaxation rate [33] (the rate of exchange can, to a certain extent, be controlled by the pH and temperature conditions of the experiment [37]).

Presaturation involves the delivery of a soft pulse at the exact frequency of the solvent peak. This weak radio frequency field serves to saturate the population difference across the solvent transition and is switched off just before the final pulse, for the data acquisition period. This is the case with 1D experiments. With 2D experiments, although ideally the narrow frequency pulse should be administered only for the duration of the recovery period, this can lead to a significant build-up of solvent signal during t_1 and therefore it is sometimes necessary to leave it on until the start of t_2. This can however have an adverse effect on the spectral symmetry due to Bloch-Siegert shifts (frequency of a signal is shifted from its usual value due to irradiation of a nearby peak) in the F1 dimension [17]. There are two possible methods of peak irradiation - one using the decoupler and the other, the transmitter. In both cases there is always the overall formation of a residual solvent peak due to significant build-up of solvent magnetisation during the evolution period of the experiment. Using the transmitter has the advantage that the phase of the small residual peak is known, and therefore, through signal averaging, can be decreased substantially. When using the decoupler, however, the phase relative to the transmitter is unknown and so signal averaging is less effective. Thus, although using the transmitter for solvent suppression is more effective, this method is more demanding on the instrument in that there has to be a continuous, rapid of transmitter power. Additional disadvantages of this method are that its implementation requires either that the transmitter saturation frequency is central in the spectrum (this forces an extended spectral

width if all peaks are to be included in a resultant 2D spectrum and thus leads to recording of large amounts of extraneous data and decreased spectral resolution) or that there is additional rapid switching of the transmitter frequency. The choice of method is therefore governed by the experimental requirements.

Solvent non-excitation methods are useful if chemical exchange between the solute protons of interest and solvent protons, occurs. These methods involve maximising the transverse magnetisation of transitions involving the sample protons while at the same time minimising those associated with the solvent [33]. This can be achieved by replacing the non-selective hard pulse by a series of non-selective pulses and delays to generate non-uniform excitation - a disadvantage in that this prevents quantitative comparisons from being between different parts of the spectrum [17]. The simplest of these experiments and one on which many of the more complex solvent suppression methods are based, is the "jump and return" (JR) sequence [58]. This sequence involves two 90° pulses of opposite phase separated by the time, τ [33]. By setting the transmitter frequency to exactly that of the solvent peak, after the first pulse, the associated transverse magnetisation remains static in the rotating frame while all other signals are allowed to precess normally. With the second pulse, the sample proton transverse magnetisation will remain unaffected, while that of the solvent will return to the z axis [17]. Many new pulse sequences have been developed to cope more specifically with 2D solvent suppression problems, for example the Sklenar-Bax sequence [37,59]. The use of solvent non-excitation methods in 2D experiments is, however, far from trivial, therefore making presaturation a preferable method of solvent suppression wherever possible.

The above solvent suppression techniques have, to a large extent, been replaced on modern spectrometers by shaped pulse and pulse field gradient techniques [37]. However, these techniques were not available on the instrument used and therefore will not be discussed further.

Effective solvent suppression methods are a necessity if sufficient spectral resolution for proteins and peptides is to be obtained. Since 1984 there has been increased interest in and development of new techniques for solvent suppression. A review and comparison of many of these methods was carried out by W.E. Hull [37].

REFERENCES

1. Wüthrich, K. *Acc. Chem. Res.* **1989**, *22*, 36-44.

2. Wüthrich, K.; Billeter, M.; Braun, W. *J. Mol. Biol.* **1984**, *180*, 715-740.

3. Gampe, Jr, R.T.; Olejniczak, E.T.; Rockway, T.R.; Fesik, S. W. *Transactions ACA*, **1988**, *24*, 155-165.

4. Tonge, A.P.; Murray-Rust, P; Gibbons, W.A.; McLachlan, L.K. *J. Comp. Chem.* **1988**, *9(5)*, 522-538.

5. Hinds, M.G.; Welsh, J.H.; Brennand, D.M.; Fisher, J; Glennie, M.J.; Richards, N.G.J.; Turner, D.L.; Robinson, J.A. *J. Med. Chem.* **1991**, *34*, 1777-1789.

6. Yu, C.; Bhaskaran, R.; Chuan, L.C.; Yang, C.C. *Biochemistry* **1993**, *32*, 2131-2136.

7. Fry, D.C.; Madison, V.S.; Bolin, D.R.; Greeley, D.N.; Toome, V.; Wegrzynski, B.B. *Biochemistry* **1989**, *28*, 2399-2409.

8. Barden, J.A.; Kemp, B.E. *Biochemistry* **1993**, *32*, 7126-7132.

9. Breeze, A.L.; Harvey, T.S.; Bazzo, R.; Campbell, I.D. *Biochemistry* **1991**, *30*, 575-582.

10. Wüthrich, K. *NMR in Biological Research: Peptides and Proteins*, North-Holland Publishing Company, Amsterdam, 1976.

11. Wagner, G.; Braun, W.; Havel, T.F; Schaumann, Th.; Gó, N.; Wüthrich, K. *J. Mol. Biol.* **1987**, *196*, 611-639.

12. Kline, A.D.; Braun, W.; Wüthrich, K. *J. Mol. Biol.* **1986**, *189*, 377-382.

13. Pflugrath, J.W.; Wiegand, G.; Huber, R.; Vértesy, L. *J. Mol. Biol.* **1986**, *189*, 383-386.

14. Clore, G.M.; Gronenborn, A.M.; James, M.N.G.; Kjer, M.; McPhalen, C.A.; Poulsen, F. M. *Protein Eng.* **1987**, *1*, 313-318.

15. Wüthrich, K. *NMR of Proteins and Nuceic Acids*, Wiley: New York, 1986.

16. Wemmer, D.E.; Reid, B.R. *Ann. Rev. Phys. Chem.* **1985**, *36*, 105-37.

17. Derome, A.E. *Modern NMR Techniques for Chemistry Research*, Pergamon Press, Oxford, 1987.

18. Benn, R.; Günther, H. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 519-538.

19. Griesinger, C.; Schwalbe, H.; Schleucher, J.; Sattler, M. *Two Dimensional NMR Spectroscopy, Applications for Chemists and Biochemists* (Croasmun, W.R.; Carlson, R.M.; eds.), VCH Publishers, Inc., 1994, 457-580.

20. Piantini, U.; Sørensen, O.W.; Ernst, R.R. *J. Am. Chem. Soc.* **1982**, *104*, 6800.

21. Rance, M.; Sørensen, O.W.; Bodenhausen, G.; Wagner, G.; Ernst, R.R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1983**, *117*, 458.

22. Braunschweiler, L.; Ernst, R.R. *J. Magn. Reson.* **1985**, *53*, 521.

23. Bax, A.; Davis, D.G. *J. Magn. Reson.* **1985**, *65*, 355.

24. Neuhaus, D.; Williamson, M. *The Nuclear Overhauser Effect in Structural and Conformational Analysis*, VCH Publishers, New York, 1989

25. Jeener, J; Meier, B.H.; Bachmann, P.; Ernst, R.R. *J. Chem. Phys.* **1979**, *71*, 4546.

26. Kumar, A.; Wagner, G.; Wüthrich, K.; Ernst, R.R. *J. Chem. Phys.* **1981**, *103*, 3654.

27. Macura, S.; Ernst, R.R. *Mol. Phys.* **1980**, *41*, 95.

28. Bothner-By, A.A.; Stevens, R.L.; Lee, J.; Warren, C.D.; Jeanloz, R.W. *J. Am. Chem. Soc.* **1984**, *106*, 811.

29. Bax, A.; Davis, D.G. *J. Magn. Reson.* **1985**, *63*, 207-213.

30. Kessler, H.; Griesinger, C.; Kerssebaum, R.; Wagner, K.; Ernst, R.R. *J. Am. Chem. Soc.* **1987**, *106*, 811.

31. Bull, Et. E. *J. Magn. Reson.* **1987**, *72*, 397.

32. Kessler, H.; Seip, S. *Two Dimensional NMR Spectroscopy, Applications for Chemists and Biochemists* (Croasmun, W.R.; Carlson, R.M.; eds.), VCH Publishers, Inc., 1994, 619-654.

33. Frenkiel, T.A. *NMR of Macromolecules, a Practical Approach* (Roberts, G.C.K.; ed.), Oxford University Press, Oxford, 1993, 35-70.

34. Gray, G.A. *Two Dimensional NMR Spectroscopy, Applications for Chemists and Biochemists* (Croasmun, W.R.; Carlson, R.M.; eds.), VCH Publishers, Inc., 1994, 1-65

35. Marion, D.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1983**, *113*, 967-974.

36. States, D.J.; Haberkorn, R.A.; Ruben, D.J. *J. Magn. Reson.* **1982**, *48*, 286-292.

37. Hull, W.E. *Two Dimensional NMR Spectroscopy, Applications for Chemists and Biochemists* (Croasmun, W.R.; Carlson, R.M.; eds.), VCH Publishers, Inc., 1994, 67-456.

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

39. Homans, S.W.; Dwek, R.A.; Boyd, J.; Mahmoudian, M.; Richards, W.G.; and Rademacher, T.W. *Biochemistry*, **1986**, *25*, 6342.

40. Homans, S.W. *NMR of Macromolecules, a Practical Approach* (Roberts, G.C.K.; ed.), Oxford University Press, Oxford, 1993, 289-313.

41. Bax, A.; Subramanian, S.J. *J. Magn. Reson.* **1986**, *69*, 565.

42. Müller, L. *J. Am. Chem. Soc.* **1979**, *101*, 4481.

43. Bax, A. *Annu. Rev. Biochem.* **1989**, *58*, 223-256.

44. Wüthrich, K. *Science* **1989**, *243*, 45-50.

45. Bruch, M.D.; Noggle, J.H.; Gierasch, L.M. *J. Am. Chem. Soc.* **1985**, *107*, 1400-1407.

46. Barsukov, I.L.; Lian, L.Y. *NMR of Macromolecules, a Practical Approach* (Roberts, G.C.K.; ed.), Oxford University Press, Oxford, 1993, 315-357.

47. Landy, S.B.; Rao, B.D.N. *J. Magn. Reson.* **1989**, *83*, 29-43.

48. Baleja, J.D.; Moult, J.; Sykes, B.D. *J. Magn. Reson.* **1990**, *87*, 375-384.

49. Bax, A.; Davis, D.G. J. *Magn. Reson.* **1985**, *65*, 533-535.

50. Simonis, U.; Dallas, J.L.; Walker, F.A. *Inorg. Chem.* **1992**, *31*, 5349-5350.

51. Griesinger, C.; Ernst, R.R. *J. Magn. Reson.* **1987**, *75*, 261.

52. Dyson, H.J.; Wright, P.E. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 519-538

53. Griesinger, C.; Sørensen, O.W.; Ernst, R.R. *J. Am. Chem. Soc.* **1985**, *107*, 6394.

54. Griesinger, C.; Sørensen, O.W.; Ernst, R.R. *J. Magn. Reson.* **1987**, *75*, 474.

55. Griesinger, C.; Sørensen, O.W.; Ernst, R.R. *J. Chem. Phys.* **1986**, *85*, 6837.

56. Müller, L. *J. Magn. Reson.* **1987**, *72*, 191.

57. Dyson, H.J.; Wright, P.E. *Two Dimensional NMR Spectroscopy, Applications for*

Chemists and Biochemists (Croasmun, W.R.; Carlson, R.M.; eds.), VCH Publishers, Inc., 1994, 655-698.

58. Plateau, P.; Guéron, M. *J. Am. Chem. Soc.* **1982**, *104*, 7310.

59. Sklenar, V.; Bax, A. *J. Magn. Reson.* **1987**, *75*, 378.

CHAPTER 3

COMPUTATIONAL METHODS

Molecular mechanics calculations, sometimes referred to as force field calculations, constitute an effective means of molecular conformation determination [1]. A molecule may be represented as a group of atoms held together by simple harmonic forces [1]. It is possible to generate a potential energy function which models the molecular forces to produce a feasible potential energy surface for all the internal degrees of freedom within a given molecule [2]. The manner in which this potential energy function is defined is determined by the force field used. The standard energy equation forming the basis for all molecular mechanics calculations is as follows:

$$E_{tot}=E_{str}+E_{bend}+E_{tor}+E_{vdw}+E_{ele}+E_{oop}+E_{hb}+.... \quad (3.1)$$

where the various terms refer to bond stretching (E_{str}), angle bending (E_{bend}), torsion angles (E_{tor}), van der Waals energy (E_{vdW}), Coulombic interaction energy (E_{ele}), out of plane bending (E_{oop}) and hydrogen bond energy (E_{hb}), respectively [3]. There are many other energy terms which may be added to this basic equation, including those associated with solvent effects, off-diagonal or cross terms, various input constraints etc. [3]. E_{tot} is a measure of the extent to which any given molecular conformation deviates from the "ideal" (zero energy) conformation and thus on its own is meaningless. It represents the sum of the energy differences between the real and ideal/equilibrium values associated with the various defined energy parameters [2,3]. The possibility exists that the hypothetical state may not even be accessible to certain molecules [3]. In this case, the three-dimensional coordinates of atoms in a molecule are manipulated and modified in an attempt to generate a structure with minimal steric energy - the assumption made, is that, of a family of conformers, the one having the lowest steric energy is the most favourable [1]. The energy terms used and the

manner in which the actual atoms are defined constitute the distinguishing features of any given force field [1].

3.1 Molecular Mechanics vs Quantum Mechanics.

The major difference between the use of Quantum Mechanics (Molecular Orbital Theory) and Molecular Mechanics (MM) is that, while Quantum Mechanics is based on the electronic configuration of the molecular system [2], MM approximates the molecular system to spheres connected by springs which are under the influence of classical Newtonian forces [3], i.e. that are joined together by bonds defined by the potential energy function associated with any given force field [2].

The Quantum Mechanical approach requires solution of the Schrödinger equation:

$$H\Psi(Q',q) = E\Psi(Q',q) \qquad \textbf{(3.2)}$$

where H is the Hamiltonian operator, Ψ is the wavefunction which, as indicated, is a function of the coordinates of the nucleus, Q', and the electrons, q, and E is the energy of the system [4]. To solve this equation as it stands is virtually impossible, and for this reason it was necessary to make certain approximations. This was done by Born and Oppenheimer who, based on the significant mass difference between the electrons and nuclei (the motion of the electrons is considerably greater than that of the nuclei), made the approximation that nuclei were stationary and therefore had no associated kinetic energy [4,5]. This approximation allows the Schrödinger equation to be split into two new functions - one describing the motion of the electrons and the other the motion of the nuclei [4]. The equation representing the electron motion is now given by:

$$H\psi(Q',q) = E(Q')\psi(Q',q) \qquad \textbf{(3.3)}$$

where ψ is the wavefunction for the electrons, and the energy, E, is now a function only of the coordinates of the nuclei [4]. It is this energy that describes the potential energy surface of a given molecule [4].

The second equation describing the motion of the nuclei on this surface is:

$$H\Phi(Q')=E\Phi(Q') \tag{3.4}$$

where Φ is the wavefunction of the nuclei [4]. This equation is useful for calculating the structure or time evolution of a molecule [4]. This requires solving the first equation which is not a trivial task. The difficulty lies in mapping out the potential energy, surface and this is what makes Quantum Mechanics approach computationally unfeasible [2].

The assumption that nuclei are significantly heavier than electrons means that quantum effects can be considered negligible [4]. Hence it is possible to replace equation (3.4) by Newton's equation of classical mechanics given in equation (3.5) [4].

$$-\frac{dE}{dQ'}=m\frac{d^2Q'}{dt^2} \tag{3.5}$$

where m is the nuclear mass. Molecular mechanics calculations now solve this equation using an empirical fit to the potential energy surface (see force fields below) [4]. They ignore time evolution factors and concentrate on the generation of various structures and their corresponding energies [4].

3.2 Molecular Mechanics

Extensive research has been conducted in the application of MM calculations to peptide and protein secondary and tertiary structure determination. The previous chapter illustrated the use of one and two dimensional NMR in the gathering of important information regarding interproton distances and various backbone and sidechain torsion angles. This data is used to constrain the molecule in subsequent computational calculations implemented for structure determination.

3.2.1 Advantages and Disadvantages of Molecular Mechanics

The primary advantage of the use of molecular mechanics is the speed at which the calculations can be carried out. The fact that the potential energy surface, as defined by the appropriate force field, constitutes the starting point for MM calculations makes them orders of magnitude faster than those based on quantum mechanical methods which rely on accurate construction of the potential energy surface [2]. The fact that the computer code has been very well developed allows MM to be applied to very large molecules (thousands of atoms), while molecular orbital calculations are restricted to smaller molecules in the region of 100-150 atoms. MM has proved to be the only means of realistically searching conformational space and generating accurate geometries and associated relative energies which allow subsequent identification of low energy minima [3].

However, there are a few problems with MM that need to be considered when using this method for structure determination. This method deals only with atomic nuclei and thus problems arise if electronic effects such as orbital interactions, bond breaking and formation etc., contribute to the properties of interest [2]. There are also limitations in so far as the choice of class of molecule is concerned - this choice being dependent on the availability of a suitable parameter set [3]. Other drawbacks associated with the implementation of MM include the fact that coulombic forces are poorly dealt with, no kinetic information can be gained from the results, and application of this method is restricted to equilibrium chemistry i.e. this method is unable to deal with excited states [3]. A further consideration is that although all parameter information should theoretically be based on information gained from experimental results, this is not always possible, and in these cases it is necessary for the relevant parameters to be derived using Molecular Orbital Theory calculations [3]. Finally, it should also be noted that the conformational search will not necessarily generate a global minimum, and there always exists the possibility that certain low energy structures are in fact only trapped in relatively deep local minima or saddle points [2]. Location of a "global minimum" is, therefore purely speculative and results can only be assessed through a comparison of the relative energies of various conformers generated from runs in which the same force field is used with the same parameter set.

3.2.2. Generation of a Starting Structure

3.2.2.1 Distance Geometry

Structure information should first be introduced into a program which optimises the straight chain molecule in accordance with the constraints. There are two program types which carry out calculations such as these: one based on metric matrix distance geometry and the other on distance geometry in torsional space [6]. The former operates to convert the initial distance constraints to cartesian co-ordinates via a metric matrix [6] and includes programs such as DGEOM [7], DG-II [8], DISGEO [9], DSPACE [10], EMBED [11] and XPLOR/dg [12]; while the latter depends on a variable target function method in which bond lengths and bond angles are fixed and only torsion angles are treated as independent variables [6]. The DISMAN [13,14] and DIANA [15,16] programs adopt this approach.

Regarding the metric matrix approach, upper and lower bound distance constraints input into the programs are used in the generation of a distance matrix [6]. The upper bounds, if not specified, are assigned lengths greater than the overall protein size, while the lower bounds are restricted to the sum of the van der Waals radii of the two relevant atoms [6]. The structure possibilities are narrowed down by a process referred to as bounds smoothing. This process involves decreasing the upper bounds and increasing the lower bounds to minimise geometrical impossibilities [6]. This is done using a triangle inequality which is based on the following reasoning: if three atoms i, j and k are considered, then the furthest possible distance between, say, i and k will occur when i, j and k are in a straight line with j lying between the two. If the corresponding upper bounds are u_{ij}, u_{ik}, and u_{jk}, then the following is true [6]:

$$u_{ik} \leq u_{ij} + u_{jk} \qquad \textbf{(3.6)}$$

Thus by decreasing u_{ik} until the equation is satisfied, new upper bound values for each atom pair can be determined. It is possible to carry out lower bound smoothing in a similar manner. If the same three atoms are considered, then the smallest possible distance between, say, i and k will occur if i, j and k are in a straight line with i and k lying adjacent to each other. The lower bounds are given by l_{ij}, l_{ik} and l_{jk}, and l_{ik} is then increased in such a way as to satisfy

the following [6]:

$$l_{ik} \geq l_{ij} - l_{jk} \tag{3.7}$$

An inability to satisfy the triangle inequality indicates problems with the input distance constraints [6].

On completion of bounds smoothing, a set of trial distances obeying the upper and lower bound constraints is generated. The manner in which the trial distance sets are chosen is one of the distinguishing features of the various programs. Those in which distances are chosen from a random distribution face the problem of limited sampling of conformational space [6,17]. This problem can be overcome by narrowing the number of inconsistent distances using a set of self-consistent distance matrices [6]. An alternative method involves choosing the trial distances to fit a particular distribution function [6].

The next step involves an embedding process in which distances are converted into co-ordinates via the generation of a metric matrix. Once the three dimensional structure has been generated it is necessary to check the chirality before optimisation. During optimisation initially just the chirality, and then the chirality and distances are optimised through minimising a target function to below a specified threshold value [6]. This metric matrix approach generates structures which are likely to possess high potential energies and therefore need to be further refined using molecular mechanics calculations performed by programs such as AMBER [18], MacroModel [19,20], CHARMM [21], DISCOVER [22], GROMOS [23] and XPLOR [24].

Distance geometry carried out in torsional space makes use of a variable target function and provides the basis of the DISMAN and DIANA programs. The use of torsional space greatly reduces the number of variables when compared to the previous method, thereby also reducing the amount of computer memory required to operate these programs [6]. This approach is based on fixing the bond distances and angles so that the torsion angles constitute the only independent variables in the system. Although variations of the standard target function do exist, it commonly takes the form [25]:

$$T=\sum_{i<j}[\theta(D_{ij}-U_{ij})(D_{ij}-U_{ij})+\theta(D_{ij}-L_{ij})(D_{ij}-L_{ij})] \quad \textbf{(3.8)}$$

where D_{ij} corresponds to the distance between two atoms i and j, and U_{ij} and L_{ij} to the associated upper and lower bounds, respectively. $\theta(x)$ is zero if $x \leq 0$ and one if $x > 0$. Function optimisation is attained when $T = 0$.

Structure determination is carried out through minimisation of the target function starting from a random structure and a set of relevant distance constraints. Optimisation of this function involves first derivative determination which facilitates rapid correction of distance violations through alteration of the relevant torsion angles [6,13]. To reduce the multiple minimum problem, a variable optimisation process is implemented. This involves initially optimising short-distance intraresidue constraints of a given residue i and then gradually extending this optimisation to include the longer range constraints associated with neighbouring residues, first the (i±1) and then the (i±2) residues, until eventually the whole molecule is successfully minimised [6,26]. As before with the metric matrix method, the resultant structures obtained are in good agreement with the distance constraints, but also probably have a high potential energy and thus need to be submitted to one of the above mentioned molecular mechanics programs. The primary function of these distance geometry programs is thus to provide suitable starting structures for refinement through either restrained molecular dynamics, simulated annealing or Monte Carlo calculations (see later).

3.2.3 Conformational Space Searching

3.2.3.1 Force Fields.

Force fields are derived as a means of expressing the potential energy surface associated with the movement of atoms within a given molecule [2]. All force fields used in this study are based on a combination of internal co-ordinates, to express the bonding features, and interatomic distances, to describe van der Waals and other electrostatic interactions, all of which contribute to the target function [4]. Thus it is the manner in which atom types and the overall potential energy function are defined that distinguishes one given force field from

another. The most well developed and commonly used force fields include those formulated for small organic molecules containing minimal heteroatoms and those developed for peptide, polynucleotide and protein macromolecules. The former includes the MM2, MMP2 and MM3 forcefields and are based on pioneering work of Allinger [27-29], while the latter includes the AMBER [18], DISCOVER [22], CHARMM [21] and GROMOS [23] force fields [4,6].

3.2.3.2 Minimisations

The simple minimisation steps applied to sample structures minimise the potential energy by reducing the nett force on each atom to zero, thereby locating stable structures on the potential energy surface [4]. The minimisation functions assume an approximately harmonic potential energy surface. This approximation is based on the fact that the surface becomes essentially harmonic in the region close to a minimum [4]. Any given minimisation procedure requires generation of a suitable objective function which describes the energy of a system as a function of the co-ordinates for a given conformation [4]. Thereafter, this objective function is iteratively minimised through subtle conformational changes. The minimisation procedure, relying on determination of the distance to and direction of the minimum, is based on calculation of the first derivatives of the objective function. The resultant gradient vector only points downhill, and not necessarily towards the minimum [4]. Therefore the initial gradient vector is iteratively searched (in one dimension) for a local energy minimum which occurs at a point tangential to an energy contour. To further reduce the energy of the system the new gradient vector must then be perpendicular to the previous one [4]. In this manner the potential energy surface is explored until the first derivatives are equal or nearly equal to zero, at which stage convergence is said to occur [4]. Common minimisation techniques used include Steepest Descent, Conjugate Gradient and Newton-Raphson methods [4]. The first of these three is a very robust method usually applied to initial structures generated from molecular mechanics, while the latter two are subsequently used to refine the structures further [4].

3.2.3.3 Monte Carlo Conformation Space Search

Monte Carlo conformational space searches can operate in either an internal (torsion angles)

or external (cartesian) coordinate system.

(a) Internal Coordinate Searches

A non-linear molecule having N atoms will have 3N-6 vibrational degrees of freedom [30]. The fact that the number of three dimensional structures expands geometrically with the number of degrees of freedom, makes the task of locating all possible minima for a large molecule that much more difficult. Since it is essentially only the torsion angles that contribute to major conformational changes, it is possible to differentiate between the more energetically constrained bond lengths and angles and the significantly less constrained torsion angles [31]. This decreases the number of degrees of freedom to include only the torsion angles and therefore restricts the search to torsional space. Thus an R^{3N-6} problem becomes an $R^{N'}$ problem (where R is the sampling density along each torsional co-ordinate and N' is the number of torsions) [31,32].

The two possible methods for implementing a conformational space search are (i) a deterministic or tree search or (ii) a stochastic or random search. The former constitutes a means of covering all conformational space. Its efficiency is, however, severely limited by the size of the molecule due to the increase in area of conformational space associated with greater flexibility [32]. The tree search is based on the systematic sampling of torsion space - unconstrained torsion angles are specified and each of these angles is then systematically rotated through a given angle - usually 30°. Therefore there are 12 sampling points for each torsion and an overall of N^{12} conformations to be generated where N is the number of variable torsions [3]. Although this is an efficient means of generating starting structures and complete coverage of conformational space is guaranteed, it becomes impractical for larger molecules as the number of variable torsions increases [3,30,31]. Introduction of constraints simplifies the search since any constraint violation leads to pruning off of entire branches of the "tree". However, even with constraints, this type of search is only feasible for molecules having six or less variable torsions [32]. It is possible to increase its feasibility, for instance, by increasing the size of the angle through which the torsion is rotated, but this reduces the coverage of conformational space to the extent that certain minima may be missed. It is also possible to ignore some "non-relevant" angles such as for example, those connected to methyl

groups [3].

Because the tree search is not only time consuming, but also imposes restrictions on the number of variable torsions, a random search method constitutes the method of choice for Monte Carlo searches. These stochastic methods of searching conformational space are based on starting structure generation through random variation of the appropriate torsion angles [32]. Although, the area of conformational space that is searched is reduced, the search may still be considered adequate if a run is allowed to proceed for sufficient time [31]. There are two methods employed - the random walk method and the superior use-directed method. A single Monte Carlo step involves: generation of a starting structure, applying the relevant random variations to the chosen co-ordinates, minimising the resultant structure, comparing the minimised structure to previous structures and finally saving it if it is unique or discarding it if it is a duplicate [32]. In the former method, starting structures are generated through energetic considerations, i.e. any structure generated having an energy greater (within a specified energy window) than that of the immediate global minimum is discarded. If the energy is less, then the structure is maintained and used as the starting structure for the next Monte Carlo step. The disadvantage of this method is that certain structures may be located more often than others and are thus more frequently used as starting structures. This leads to a bias in the choice of starting structure [31]. To solve this problem, a use-directed method can be employed in which structures are considered in terms of the global minimum - any structure having an energy within the specified energy window of the global minimum is admissible. All stored structures are then assessed and the least used structure chosen as the starting structure, thus ensuring a more effective means of conformational space coverage [31]. One of the major problems is the question of search convergence. Convergence is achieved if a number of non-identical searches produce the same conformational results [32]. Another criterion for search convergence is a high minimum duplication rate, i.e. all saved structures have been repeatedly found a significant number of times [32].

(b) External Co-ordinate Searches

An external (cartesian) co-ordinate search method based on a random translation algorithm was proposed by Saunders *et al.* [31]. This involved generation of new starting conformations

through random translation of atoms [31], and as would be expected, produced structures having high steric energy. These high energy interactions were rapidly reduced in an ensuing optimisation step carried out using a simple two-body force field. The product structure was then further minimised to a local minimum in accordance with the usual procedure followed in MM calculations [31].

(c) Summary

Of the above three methods supplied for conformational space searching, the use-directed internal co-ordinate search method has been shown to be the most effective for maximal coverage of conformational space [32]. The efficacy of a search is to a large extent dependent on the way in which starting structures are generated. It was found that low energy starting structures generally produced low energy conformers. For this reason the use-directed method, which operates by choosing least used of previously stored low-energy structures as starting structures, was particularly effective and generated high A_E[1] and low SD_{100}[2] values [31]. The number of variable torsion angles also serves to increase the efficacy of a given search; the greater the number of torsions, the greater is the sampling of conformational space [31]. With regards to internal versus external co-ordinate searches, the former has been shown to be more effective in terms of generating a greater number of well-defined low energy conformers per CPU time, even though starting structures are more easily generated in the latter [31].

3.2.3.4 Molecular Dynamics

Molecular Dynamics (MD) is based on a study of molecular motion and involves solution of Newton's equations of motion [4]. Such calculations require an initial set of atomic co-

---

[1] A_E = Acceptance rate - the percentage of Monte Carlo steps generating structures within the specified energetic window. This is a measure of how selectively a search generates structures within the energetic range [31].

[2] SD_{100} = the standard deviation of the duplication rate per 100 acceptances. This is a measure of the uniformity of a given search of low-energy conformational space [31].

ordinates and velocities, the former being generated either through distance geometry calculations, NMR or X-ray crystallographic structure data or through model-building, while the latter are randomly assigned from the Maxwell-Bolzmann distribution function for a specified temperature [33]. After an initial optimisation needed to reduce high energy steric and other internal interactions, MD calculations are performed to solve Newton's second law of motion [33]:

$$F_i = m_i a_i$$

$$\Rightarrow -\frac{dE}{dr_i} = m_i\left(\frac{d^2 r_i}{dt^2}\right) \tag{3.9}$$

where, for any given atom i, F_i denotes force, r_i, the position, a_i, the acceleration and m_i, the mass. The energy, defined by a given potential energy function (dependent on the force field), and the mass of the system are known, thus making it possible to determine the acceleration [3,4,33]. Knowing the initial accelerations and velocities, it is possible to determine the velocity, v_i, and as well as the position of each atom at time $(t+\Delta t)$ [4,33]. This is best done using a Taylor series, such as the one given below [33].

$$r_i(t+\Delta t) = r_i(t) + v_i(t)\Delta t + \frac{1}{2}a_i(t)(\Delta t)^2 + \ldots \tag{3.10}$$

In order to prevent the solution from diverging, the size of the timestep, Δt, must be smaller than the highest frequency vibration (typically 10^{-14}s for a C-H bond stretching vibration), and therefore is usually in the region of 10^{-15}s [4]. Dynamic simulations are temperature dependent, with the temperature of the system being measured as a function of average kinetic energy as follows [33]:

$$\frac{1}{2}\sum_{i=1}^{N} m_i\langle v_i^2\rangle = \frac{3}{2}Nk_BT \tag{3.11}$$

where N is the number of atoms in the system, k_B, the Boltzmann constant, and $\langle v_i^2\rangle$ is the average velocity squared of atom i.

The MD trajectory is time dependent and therefore allows certain conformational properties, such as folding, to be monitored on a time-scale [4]. The internal motions of proteins have been shown to occur on the picosecond timescale, and therefore instantaneous conformations may differ significantly from the average structures obtained using MD [33]. Since molecular motion generally occurs on a 10^{-3} to 10^{-9}s timescale [3], the limitation of MD lies in the extensive CPU time needed to simulate these events [33]. It only becomes feasible if an adequate set of constraints is provided [33]. These take the form of distance and torsion angle constraints, which may be obtained from NMR data. Introduction of these constraints into the calculation occurs through the addition of an energy constraint term (E_{const}) to the potential energy function [33] i.e.

$$E_{const}=E_{nOe}+E_{tor} \tag{3.12}$$

where E_{nOe} denotes the energy contribution from the distance constraints, E_{tor} the energy contribution from torsional constraints. The most commonly used distance constraint function is that based on a simple harmonic function as follows [4,33]:

$$E_{nOe}=k_{nOe}(r-r_{nOe})^2 \tag{3.13}$$

where r and r_{nOe} are the theoretical and experimental interproton distances respectively and k_{nOe} is a specified force constant. This equation may be modified to accommodate a square well potential [4]. The torsion angle restraint function can be applied as a simple harmonic function analogous to equation (3.13), or it may take on the form of a periodic function:

$$E_{tor}=k_{tor}(1+\cos(n\phi-\phi_0)) \tag{3.14}$$

where k_{tor}, is the force constant, n is an integer denoting periodicity which equals 1 for torsion angle forcing, ϕ is the input torsion constraint and ϕ_0 the phase angle [4].

The advantage of using MD as opposed to Monte Carlo methodology is that the former allows a time evolution study of processes such as diffusion and molecular folding to be carried out [4]. It also has the advantage over normal mode dynamics of not being restricted to harmonic motion, thus allowing crossing over of energy barriers and greater sampling of conformational space [4,33]. This is important because at temperatures above 300K, molecular motion is

essentially anharmonic due to high side chain mobility [33]. The higher the temperature at which MD is carried out, the greater the chance of a molecule being able to cross high energy barriers. Therefore one advantage of high temperature dynamics is that the possibility of molecules being trapped in local minima is significantly reduced. In addition, the lifetime of any given conformer is shortened, thus extending the coverage of conformational space [4]. However, high temperature dynamics also tend to yield a greater proportion of high-energy conformers after minimisation of the sampled structures [31].

The main limitation associated with using MD is that the extent of conformational space coverage is severely restricted by the time required for these calculations. Thus, although it is possible, in theory, to completely cover conformational space, it is unfeasible in practice. High temperature dynamics does allow more efficient crossing of energy barriers, but problems such as cis-trans conversions can become problematic at these elevated temperatures [4].

3.2.3.5 Simulated Annealing

Simulated annealing is a relatively new optimisation technique derived as an alternative means of global minimum location [34]. The concept behind simulated annealing is analogous to the annealing processes witnessed in crystals and metals [34]. In order for the ground state to be attained, it is necessary to very slowly decrease the temperature to ensure that the equilibrium is maintained at all times during the procedure i.e that there is sufficient time for the atoms to align themselves in such a way as to minimise the energy [35]. Thus the liquid goes from a high temperature disordered system where atoms have significant thermal motion to a low temperature highly ordered crystalline state [35]. If the cooling procedure occurs too rapidly the system equilibrium is destabilised and defects occur, resulting in the formation of higher energy amorphous crystal or polycrystalline structures which correspond to local minima [36]. This process of slowly lowering the temperature in order to slowly reduce atom mobility is referred to as annealing. The original algorithm, developed in 1953, to simulate the annealing process as a means of solving optimisation problems, was the Metropolis algorithm [35-37]. The principle behind this algorithm lies in the system being in thermal equilibrium corresponding to a given temperature, T [35]. This equilibrium allows conformational

distribution over a number of different energy states i.e. the possibility that higher energy levels are populated is not excluded [35]. The extent of population is given by the Boltzmann probability distribution [35,36]:

$$Prob(E) \sim \exp(-E/k_B T). \qquad \textbf{(1.15)}$$

Thus one of the distinguishing features of the Metropolis algorithm is the fact that uphill steps to overcome local energy barriers are possible [35,36]. Previously, searches were conducted merely through generation of a series of starting structures and subsequently minimising them as quickly as possible through steepest descent methods where energy only decreases monotonically. The local minima were subsequently compared and the lowest energy minimum accepted as a possible global minimum. The current method, however, allows transitions to higher energy levels to occur. Such transitions are more probable at higher temperatures than at lower temperatures when atom mobility is significantly reduced [38]. The Metropolis algorithm models the probability of a transition to a higher energy level occurring [34-36,39,40]. The objective function (E_1), describing a given potential energy surface, is evaluated for an initial co-ordinate set where r_i, $i=1,....,N$ corresponds to the particle/atomic positions [34]. A random step, Δr_i, $i=1,....,N$, is then taken and the new associated energy, E_2, calculated [34]. The magnitude of the step is dependent both on the properties of the objective function as well as on the extent of accuracy and resolution required [39]. The resultant energy change between the two system configurations is then assessed in terms of the Boltzmann distribution function where the probability of the configuration change occurring is given by $p = exp(-(E_2-E_1)/kT)$. If $E_2 \leq E_1$, then p is assigned the value of one and the new configuration is always accepted as the new starting configuration i.e downhill moves leading to lower objective function values will always be accepted. If, however, $E_2 > E_1$ then the probability is calculated and compared to a random value chosen on the interval [0,1]. If this value is less than p, the new configuration is accepted and an uphill step is taken. Otherwise the new configuration is discarded and the old configuration used again as a starting structure [35,36,39]. The initial starting temperature should be high in order to allow for extensive sampling of conformational space. Once no further improvements have been observed, or a specified search termination criterion has been met, the temperature is reduced [35]. The process is continued periodically, slowly reducing the temperature.

Use of this algorithm requires (i) generation of a suitable objective function, (ii) definition of reasonable starting structures, (iii) deciding on the manner in which the random steps should be taken, (iv) definition of the selection criteria (acceptance or rejection of structures) and (v) how the temperature of the system is controlled and gradually lowered [35,36].

Simulated annealing has been successfully applied to protein side chain conformation analysis using a molecular mechanics potential energy function [41,42]. Alternatively, a more simplistic application can be carried out in conjunction with MD. The system is initially equilibrated at high temperatures to ensure sufficiently large coverage of conformational space. The dynamics of the system is then modelled at various decreasing temperatures until at a very low final temperature the system ideally adopts the global minimum conformation. Successful annealing is, however, dependent on a slow decrease in temperature. Because, in practice, the amount of time required to perform such a run is unrealistic, the size of the temperature step is largely controlled by the time available for the run [43].

3.2.4 Assessment of Theoretical versus Experimental results.

Once the sample structures have been generated it is necessary to assess each of these structures in terms of their agreement with the experimental data. In the case of NMR data, this can be done effectively using a relaxation matrix approach which serves to calculate a theoretical NOESY data set which can then be compared to the original experimental NOESY results [44]. Programs which readily perform these calculations include MORASS [45,46], MARDIGRAS [47] and IRMA [48].

The **M**ultispin **O**verhauser **R**elaxation **A**nalysi**S** and **S**imulations (MORASS) program considers multispin systems in its calculations thus taking spin diffusion effects into account as well as the usual cross relaxation resulting from dipolar coupling between two nuclei [45]. A relaxation rate matrix relating peak intensity and mixing times is generated by a set of coupled differential equations, more specifically Bloch simultaneous equations, and may be calculated from the following [45]:

$$\frac{d}{dt}V(\tau_m)=-\Gamma V(\tau_m) \qquad \textbf{(3.16)}$$

where $V(\tau_m)$ is the intensity matrix and Γ is the symmetrical relaxation rate matrix containing σ_{ij} off-diagonal elements - the result of cross relaxation between nuclear spins i and j - and ρ_i diagonal elements [45] (equation 3.17).

$$\Gamma = \begin{bmatrix} \rho_1 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \cdots \\ \sigma_{21} & \rho_2 & \sigma_{23} & \sigma_{24} & \cdots \\ \sigma_{31} & \sigma_{32} & \rho_3 & \sigma_{34} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \tag{3.17}$$

Solving and rearranging equation 3.16 generates the following:

$$-\Gamma = \frac{1}{\tau_m} \ln[V(V_0)^{-1}] \tag{3.18}$$

where V_0 is the diagonal matrix corresponding to τ_m=0. Solution of the relaxation rate matrix can be reduced to an eigenvalue problem through substitution of the logarithm by eigenvalues and eigenvectors of $V(V_0)^{-1}$ [45]. Thus, generation of the relaxation rate matrix requires no knowledge of relaxation mechanisms or motional correlation times [45]. In order to simulate theoretical NOESY peak intensities for a given structure it is necessary to determine the σ_{ij} and ρ_i values corresponding to the relevant interproton distances in that structure. The equation set given below (equation 3.19) illustrates how this can be achieved with additional correlation time, τ_c, information, where ω and γ are the frequency and gyromagnetic ratio for ^{1}H nuclei, respectively, d_{ij} is the interproton distance, $\hbar$ is the Planck's constant divided by 2π and $J(\omega)$ is the spectral density.

$$\sigma_{ij} = \frac{\gamma^4 \hbar^2}{10\langle d_{ij}^3 \rangle^2} [6J(2\omega) - J(0)] \tag{3.19}$$

$$\rho_i = \sum_{j \neq i}^{n} \frac{\gamma^4 \hbar^2}{10\langle d_{ij}^3 \rangle^2} [J(0) + 3J(\omega) + 6J(2\omega)] \tag{3.20}$$

$$J(\omega)=\left(\frac{\tau_c}{1+\omega^2\tau_c^2}\right) \tag{3.21}$$

MORASS then creates a hybrid intensity matrix by merging the experimental (generated from experimental NMR NOESY results) and theoretical intensity matrices. From this new hybrid intensity matrix, it is possible to generate a hybrid relaxation rate matrix which can subsequently used to calculate the corresponding interatomic distances using equation 3.19 [49]. These distances can be used as constraints in subsequent molecular mechanics calculations to further improve the agreement between the experimental and theoretical data. This process is repeated until the system converges to below a specified threshold value. The approach described here in which the structure is iteratively modified until convergence criteria are met, is also used in the IRMA algorithm [44].

The MARDIGRAS algorithm, on the other hand, relies on modification of the relaxation rate matrix generated from the hybrid intensity matrix. The modified relaxation matrix is converted back into an intensity matrix, which is then used, in conjunction with the experimental data, to calculate a new corresponding hybrid intensity matrix. This process is repeated until convergence criteria are met. Thereafter the final relaxation rate matrix is used to generate distances which may be applied as constraints in subsequent MM calculations for accurate structure generation [44].

3.2.4.1 Determination of an *R*-Factor

Quantification of the agreement between the experimental and theoretical data is problematic and has lead to much discussion in the literature. However, one generally used method of comparison is the *R*-factor. This is based, to a large extent, on the well known crystallographic *R*-factor [50] -

$$R_1=\frac{\sum_{ij}^{N}|V_{ij}^{theo}-V_{ij}^{exp}|}{\sum_{ij}^{N}|V_{ij}^{exp}|} \tag{3.22}$$

$$R_2 = \left[\frac{\sum_{ij}^{N} (V_{ij}^{theo} - V_{ij}^{exp})^2}{\sum_{ij}^{N} (V_{ij}^{exp})^2} \right]^{1/2} \qquad (3.23)$$

The problem with this approach is the overemphasis of short interproton distances i.e these close interactions tend to dominate the equation [44,50,51]. For this reason alternative forms of equations 3.22 and 3.23, generating a weighted *R*-factor, have been derived which take into account the relationship between peak intensity and internuclear distances - $V \propto r^{-6}$ [51]

$$R_1^{1/6} = \frac{\sum_{ij}^{N} |(V_{ij}^{theor})^{1/6} - (V_{ij}^{exp})^{1/6}|}{\sum_{ij}^{N} (V_{ij}^{exp})^{1/6}} \qquad (3.24)$$

$$R_2^{1/6} = \left[\frac{\sum_{ij}^{N} [(V_{ij}^{theor})^{1/6} - (V_{ij}^{exp})^{1/6}]^2}{\sum_{ij}^{N} (V_{ij}^{exp})^{1/6}} \right]^{1/2} \qquad (3.25)$$

Although this sixth root scaling allows the longer-range interactions to be taken into account, the short range interactions are still slightly weighted. This is because larger peaks are less affected by signal to noise ratios than are the smaller ones [51]. A third means of calculating the *R*-factor if a number of mixing times are being considered is as follows:

$$R = \left[\frac{\sum_{ij}^{N} W(\tau_m)[V_{ij}^{theor}(\tau_m) - V_{ij}^{exp}(\tau_m)]^2}{\sum_{ij}^{N} W(\tau_m)[V_{ij}^{exp}(\tau_m)]^2} \right]^{1/2} \qquad (3.26)$$

where $W(\tau_m)$ is set to 1, or else to τ_m if the lower peak intensities at shorter mixing times are to be taken into account [44]. Equation 26 may also be expressed in terms of an $R^{1/6}$-factor with, of course, the exclusion of the weighting factor [44].

REFERENCES

1. Cox, P. *Journal of Chemical Education* **1982**, *59(4)*, 275-277.

2. Boyd, D.B.; Lipkowitz, K.B. *Journal of Chemical Education* **1982**, *59(4)*, 269-274.

3. Snarey, M. *Molecular Mechanics*, 1995.

4. *Discover User Guide version 2.9*, Biosym Technologies, San Diego, 1993, Part 1, 1/2-1 - 1/2-119.

5. Born, M.; Oppenheimer, J.R. *Ann Physik* **1927**, *84*, 457.

6. Sutcliffe, M.J. *NMR of Macromolecules: A Practical Approach* (Roberts, G.C.K., ed.), Oxford University Press, Oxford, **1993**, 259-392.

7. Blaney, J.M.; Crippen, G.M. *DGEOM, distance geometry program.*

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

9. Havel, T.F.; Wüthrich, K. *Bull. Math. Biol.* **1984**, *46*, 673.

10. McCammon, J.A.; Harvey, S.C. *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, 1987.

11. Havel, T.F.; Kuntz, I.D.; Crippen, G.M. *Bull. Math. Biol.* **1983**, *45*, 665-720.

12. Kuszewski, J.; Nilges, M.; Brunger, A.T. *J. Biomolec. NMR* **1992**, *2*, 33.

13. Braun, W.; Go, N. *J. Mol. Biol.* **1965**, *186*, 611.

14. Hockney, R.W. *Methods Comput. Phys.* **1970**, *9*, 136.

15. Guntert, P.; Braun, W.; Wüthrich, K. *J. Mol. Biol.* **1991**, *217*, 517.

16. Guntert, P.; Wüthrich, K. *J. Biomolec. NMR* **1991**, *1*, 447.

17. Metzler, W.; Hare, D.B.; Pardi, A. *Biochemistry* **1989**, *28*, 7045.

18. Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765-784.

19. *MacroModel Handbook*, Department of Chemistry, Columbia University, New York, 1990.

20. Mohamadi, F.; Richards, N.G.J.; Guida, W.C.; Liskamp, R.; Lipton, M.; Caufield, C.;

Chang, G.; Hendrickson, T.; Still, W.C. *J. Comput. Chem.* **1990**, *11(4)*, 440-467.

21. Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.O., States, D.J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

22. *Discover Manual version 2.8*, Biosym Technologies Inc, San Diego, CA, 1992.

23. van Gunsteren, W.F.; Berendsen, H.J.C. *Groningen Molecular Simulation (GROMOS) Library Manual.* BIOMOS B.V., Nienborgh 16, Groningen, The Netherlands, 1987.

24. Brünger, A.T. *XPLOR Manual version 3.0*, Yale University, New Haven, CT, 1992.

25. Braun, W. *Q. Rev. Biophys.* **1987**, *19*, 115.

26. James, T.L.; Basus, V.J. *Annu. Rev. Phys. Chem.* **1991**, *42*, 501-542.

27. Allinger, N.L. *J. Am. Chem. Soc.* **1977**, *99*, 8127

28. Allinger, N.L.; Yuh, Y.H. *MM2 Manual*, San Diego Supercomputer center center, GA Technologies, San Diego, CA, 1988.

29. Allinger, N.L.; Yuh, Y.H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551.

30. Laidler, K.J.; Meiser, J.H. *Physical Chemistry*, The Benjamin/Cummings Publishing Company, Inc., Menlo Park, California, 1982.

31. Saunders, M.; Houk, K.N.; Wu, Y.-D.; Still, W.C.; Lipton, M.; Chang, G.; Guida, W.C. *J. Am. Chem. Soc.* **1990**, *112*, 1419-1427.

32. Chang, G.; Guida, W.C.; Still, W.C. *J. Am. Chem. Soc.* **1990**, *111*, 4379-4386.

33. Karplus, M.; Petsko, G.A. *Nature* **1990**, *347*, 631-639.

34. Wille, L.T. *Letters to Nature* **1986**, *324*, 46-48.

35. Ku, H.; Karimi, I. *Ind. Eng. Chem. Res.* **1991**, *30*, 163-169.

36. Press, W.A; Flannery, B.P.; Teukolsky, S.A.; Vetterling, W.T.; *Numerical Recipes (The Art of Scientific Computing)*, Cambridge University Press, New York, 1986.

37. Metropolis, N.; Rosenbluth, A.; Rosenbluth, M.; Teller, A.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.

38. Tidor, B. *J. Chem. Phys.* **1993**, *97*, 1069-1073.

39. Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P. *Science*, **1983**, *220(4598)*, 671-680.

40. Bohachevsky, I.O.; Johnson, M.E.; Stein, M.L. *Technometrics*, **1986**, *28(3)*, 209-217.

41. Correa, P.E.; *Proteins: Struct., Funct., Genet.* **1990**, *7*, 366.

42. Lee, C.; Subbiah, S. *J. Mol. Biol.* **1991**, *217*, 373.

43. Homans, S.W. *NMR of Macromolecules: A Practical Approach* (Roberts, G.C.K., ed.), Oxford University Press, Oxford, **1993**, 289-313.

44. Barsukov, I.L.; Lian, L.-Y. *NMR of Macromolecules: A Practical Approach* (Roberts, G.C.K., ed.), Oxford University Press, Oxford, **1993**, 315-357.

45. Post, C.B.; Meadows, R.P.; Gorenstein, D.G. *J. Am. Chem. Soc.* **1990**, *112*, 6796-6803.

46. Meadows, R.P.; Post, C.B.; Gorenstein, D.G.; Luxon, B.A. *MORASS User Guide*, Purdue University, West Lafayette, USA, 1994.

47. Borgias, B.A.; James, T.L. *J. Magn. Res.* **1990**, *87*, 475.

48. Boelens, R.; Konig, T.M.G.; Kaptein, R. *J. Mol. Struct.* **1988**, *173*, 299.

49. Nikonowicz, E.P.; Meadows, R.P.; Gorenstein, D.G. *Biochemistry* **1990**, *29*, 4193-4204.

50. Gonzalez, C.; Rullmann, J.A.C.; Bonvin, A.M.J.J.; Boelens, R.; Kaptein, R. *J. Magn. Res.* **1991**, *91*, 659-664.

51. Thomas, P.D.; Basus, V.J.; James, T.L. *Proc. Natl. Acad. Sci.* **1991**, *88*, 1237-1241.

CHAPTER 4

LOM-AKH-I: RESULTS AND DISCUSSION

Lom-AKH-I was the first of the Lom peptide series to be isolated and sequenced by Stone *et al.* [1] in 1976 (Figure 4.1). This hydrophobic decapeptide (five of its ten amino acids exhibit a high degree of hydrophobicity) is the only one of the three Lom peptides to have been found in both *Locusta migratoria* and *Schistocerca gregaria*. A 0.0017M (2mg/ml) solution of the peptide dissolved in d^6-dimethylsulphoxide (d^6-DMSO) was prepared. Although a 90%H_2O:10%D_2O solution would have simulated *in vivo* conditions more accurately, d^6-DMSO was used because the hydrophobic nature of these peptides rendered them virtually insoluble in water. The advantage of using this solvent was that problems associated with the extensive solvent suppression required to generate adequate spectra for solutes dissolved in water were avoided. The NMR study commenced with recording variable temperature ^1H NMR spectra which showed maximum spectral resolution to occur at 35°C. This temperature seemed relatively high with respect to peptide stability. However, no denaturing was observed, and thus all subsequent proton and two dimensional (2D) NMR experiments were carried out at this elevated temperature. Even though these experiments were carried out in d^6-DMSO, the amount of residual water in the solution was significant enough to warrant some use of solvent suppression. Presaturation involved the use of the decoupler in the COSY experiment and the transmitter in the TOCSY and NOESY experiments.

Integration of the proton spectrum allowed some general assignments to be made, however, two dimensional spectra were necessary for analysis of extensively overlapped regions. Intraresidue assignments can be made through identification of the characteristic amino acid coupling patterns generated in COSY and TOCSY spectra. Since the COSY spectrum (Figure 4.2) contained sequential scalar couplings, side chain protons sometimes generated crosspeaks in the region of the diagonal, where insufficient peak resolution prevented conclusive peak

Lom-AKH-I: pGlu-Leu-Asn-Phe-Thr-Pro-Asn-Trp-Gly-ThrNH2

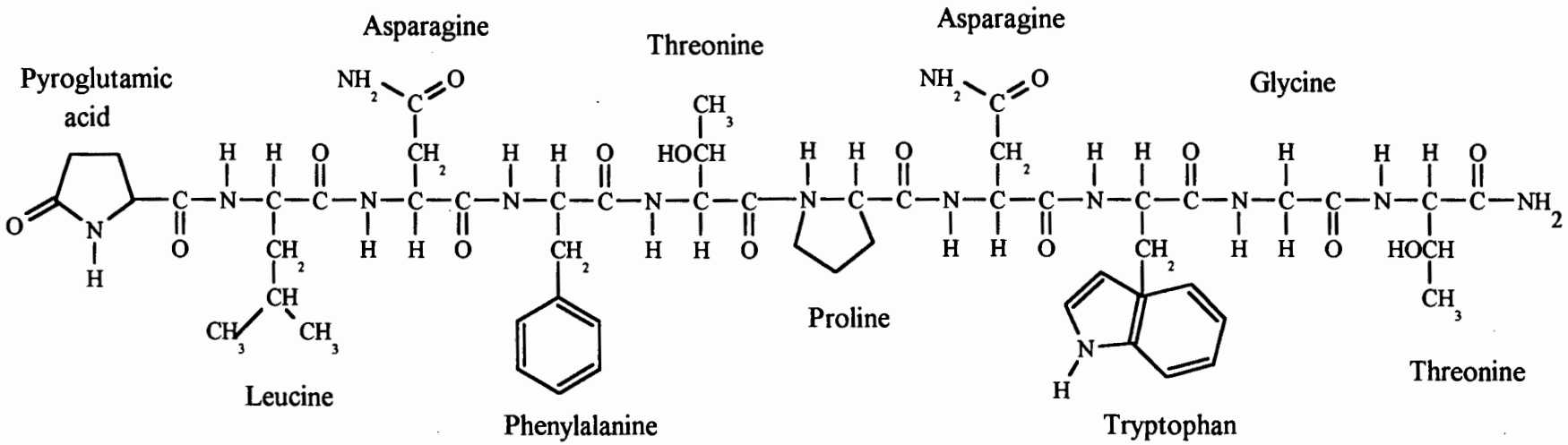


Figure 4.1: Primary sequence recorded for Lom-AKH-I.

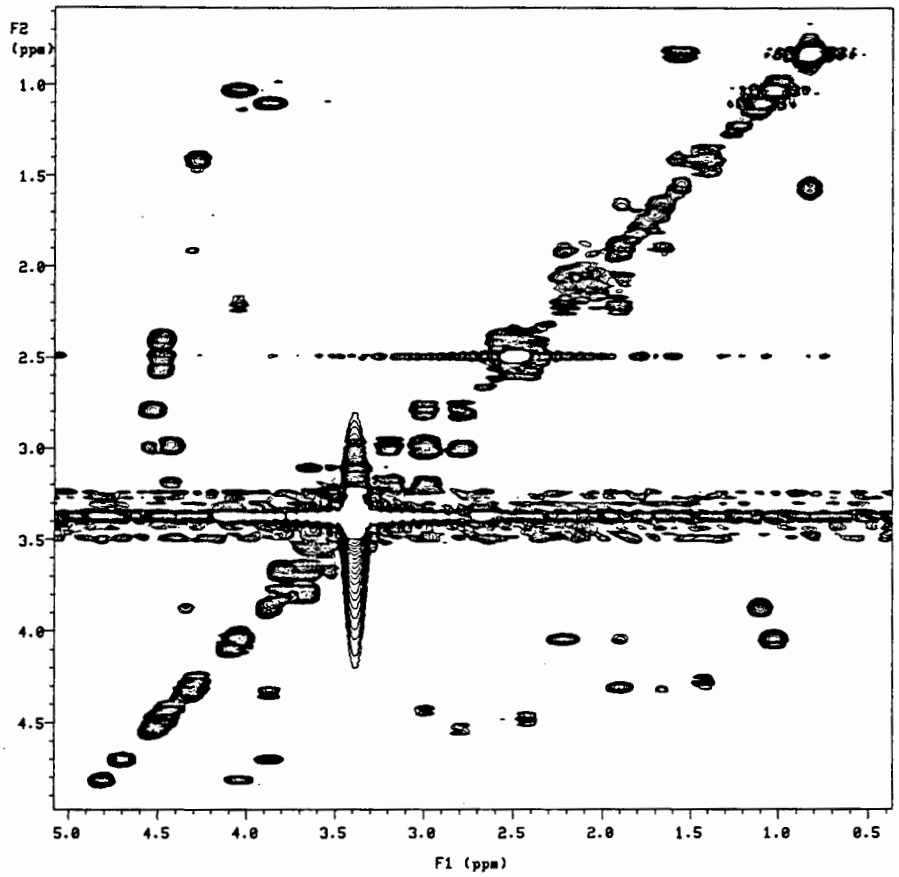


Figure 4.2: C_α-side chain region of the COSY spectrum recorded in d^6-DMSO for Lom-AKH-I.

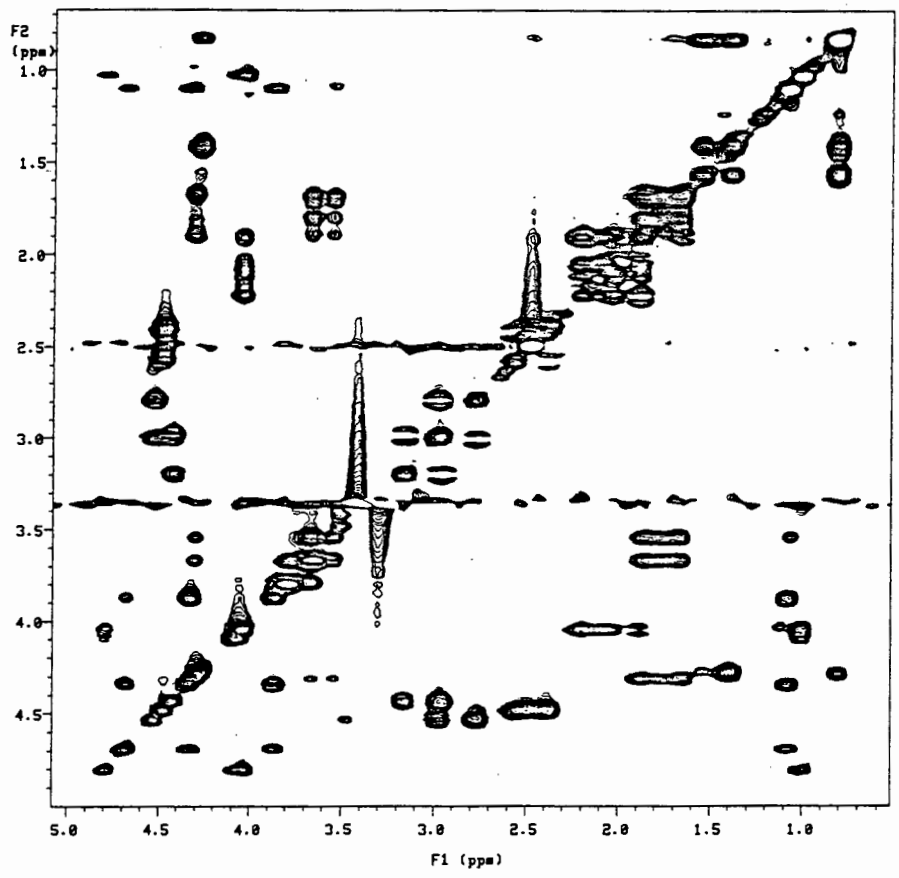


Figure 4.3: C_α-side chain region of the TOCSY spectrum recorded in d^6-DMSO for Lom-AKH-I.

assignments from being made. The TOCSY spectrum (Figure 4.3) showing total correlation, contained additional crosspeaks corresponding to these protons in regions away from the diagonal. The superior resolution of these peaks improved the accuracy of their assignments. These two experiments are thus interdependent and are both necessary for complete assignment of the peptides discussed here.

Regarding the assignment of Lom-AKH-I, the glycine, two Thr and Leu residues were readily identifiable through their very distinct AX, A_3MX and A_3B_3MPTX spin systems, respectively (Figure 4.3 and 4.4). Location of the two Thr and the Leu residues was facilitated further by the upfield position of the four associated methyl groups.

Overlapping peaks often indicate a marginal difference in chemical shift values between the contributing protons. In such cases, to distinguish between chemical shift differences of less than 0.01ppm, it was necessary to take a trace through the relevant crosspeak in the spectrum and make assignments based on trace peak maxima. This was necessary to distinguish the two β-protons of the Leu residue in the COSY spectrum, which in this case showed superior resolution to the TOCSY.

Although crosspeaks corresponding to side chain protons of the Pro (in particular the δ protons) and the pGlu (the γ protons) residues were identifiable in the COSY spectrum, their close proximity to the diagonal resulted in limited resolution. However, the region of the TOCSY (recorded with a mixing time of 80ms) showing relay scalar connectivities between the α- and side chain protons had superior resolution of crosspeaks corresponding to these protons and thus enhanced assignment accuracy.

Repeat residues often have similar chemical shift values, making the assignment process more difficult. Although the two Thr residues were readily distinguishable, the same was not true for the two Asn residues. Careful assessment of peak maxima and appropriate traces was required to make the relevant assignments.

The ring protons of the Phe residue evaded assignment due to their near degenerate chemical shift values. Assignment of the Trp ring was, however, possible from the ring couplings

observed in the COSY and TOCSY spectra (Figure 4.5). Their assignments to specific ring atoms were based on the corresponding literature values recorded for this residue. Two of the three amine groups were readily located in the TOCSY spectrum, while the third, which occurred in the same chemical shift range as the Phe aromatic protons, was unassignable. The absence of nOe crosspeaks between any of these protons and other side chain or backbone protons prevented their assignment to any specific residue.

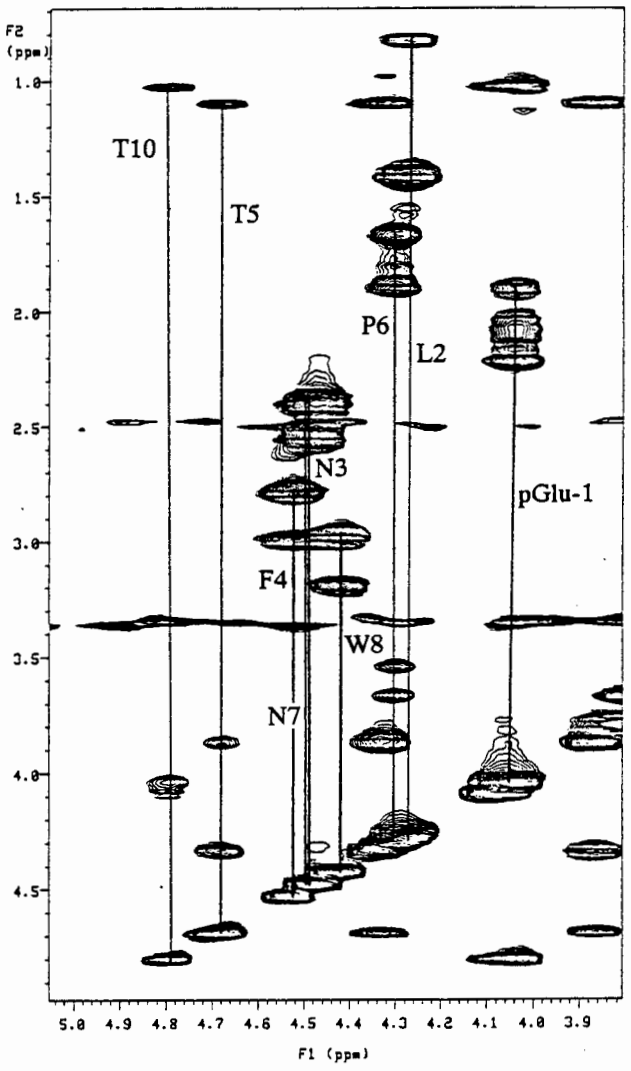


Figure 4.4: TOCSY spectrum showing the relay scalar connectivities between the C_α and sidechain protons of Lom-AKH-I.

Four of the ten residues, viz. the two Asn, Phe and Trp residues, had AMX spin systems. Although the coupling patterns were readily identifiable and assignable, to differentiate them on a residue basis was not possible. The two Asn residues could tentatively be distinguished from the other two residues due, not only, to their similar chemical shift values, but also by the upfield positioning of their β-protons - the β-protons of residues containing aromatic systems will be subject to greater deshielding effects and therefore resonate further downfield than those positioned adjacent to carbonyl groups, as is the case in Asn residues. Confirmation of these assignments required dipolar coupling information obtained from the NOESY experiment carried out with a mixing time of 150ms. Dipolar coupling was also needed to distinguish between the two Asn residues (at positions 3 and 7 in the sequence) and the two Thr residues (at positions 5 and 10 in the sequence) - see later.

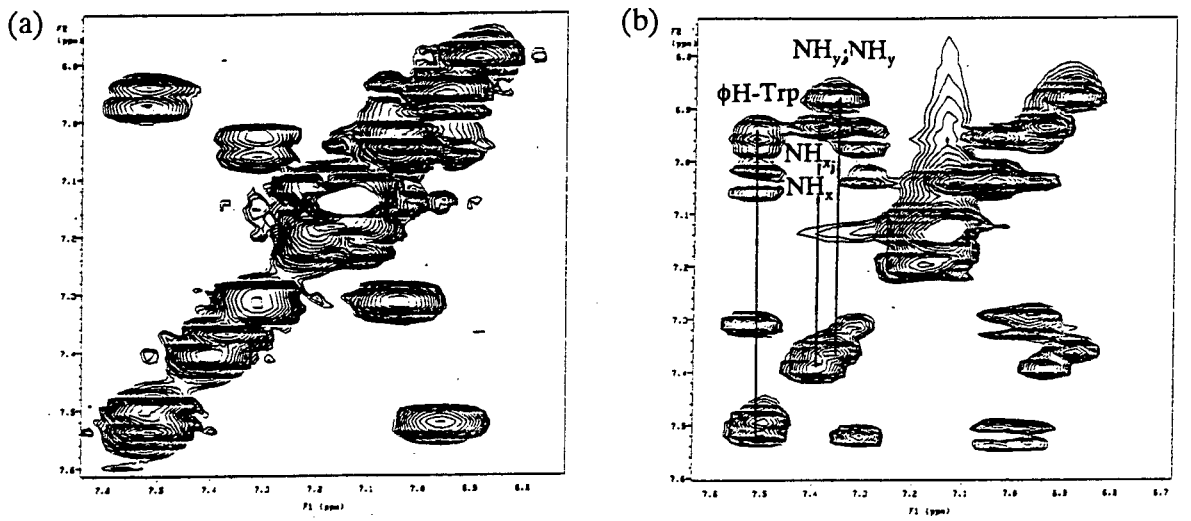


Figure 4.5: Region of (a) the COSY and (b) the TOCSY experiments showing the definite coupling observed between the tryptophan aromatic protons which allowed their accurate assignment.

All NOESY values for Lom-AKH-I were recorded with the alpha region being read off the F2 high resolution axis. This choice was the result of predominant overlap occurring in the alpha region of the spectrum. Only one side of the diagonal was considered, since the relaxation pathways in the two dimensions were different, and averaging could have led to inaccurate intensity data. The primary sequence of the peptide was used in the residue differentiation process, i.e. peak assignment in the NOESY fingerprint region was based on the probability of certain interproton interactions generating significant nOe's. In this manner the Trp and Phe residues were differentiated from each other by the presence of a Gly-Trp peak. Similarly, a Phe-Thr_5 peak distinguished the two Thr residues, and Asn_7-Pro and Asn_3-Leu peaks distinguished the two Asn residues. The complete set of chemical shift values assigned to Lom-AKH-I are recorded below in table 4.1 - see also Figure 4.6. These values are in good agreement with corresponding literature values for random coil amino acids also recorded in d^6-DMSO (Table 4.2) [2].

The possibility that Pro residues may undergo *cis-trans* isomerism makes it necessary to ascertain which isomer is present for a particular peptide, and if both are present, to determine their respective ratios. Considering an -X-Pro- sequence, evidence for the existence of a *trans* isomer lies in the presence of nOe peaks between the two Pro δH's and the αH of X, the preceeding residue. An nOe crosspeak between the αH's of Pro and X indicates a *cis* conformation [3]. If both isomers are present the proton spectrum will contain a double set of resonance peaks - one corresponding to the *trans* conformation and the other to the *cis* [4].

Table 4.1: Chemical shift values recorded for Lom-AKH-I in d^6-DMSO.

Residue	NH	αH	βH	βH'	Other
pGlu	7.710	4.065	2.230	1.920	2.190 γH 2.060 γH'
Leu	8.032	4.289	1.478	1.412	1.564 γH 0.852 δH 0.832 δH'
Asn_3	8.098	4.492	2.500	2.390	
Phe	7.748	4.540	3.000	2.800	6.98-7.24 φH's
Thr_5	8.082	4.352	3.890		1.100 γH 4.699 OH
Pro		4.320	1.916	1.660	3.670 γH 3.560 γH' 1.880 δH 1.810 δH'
Asn_7	8.046	4.494	2.560	2.420	
Trp	7.999	4.446	3.170	2.980	10.720 φNH 7.198 C2H 7.530 C4H 6.959 C5H 7.041 C6H 7.320 C7H
Gly	8.261	3.799			3.701 αH'
Thr_{10}	7.520	4.095	4.060		4.820 OH 1.030 γH

Only two of the three NH_2 groups could be assigned in the spectrum with the relevant crosspeaks occurring at 7.39ppm; 6.95ppm and 7.36ppm; 6.90ppm. It was impossible to make any conclusive assignments of these groups to specific residues. The third group occurred in the same resonance frequency range as the phenylalanine aromatic protons and therefore were unresolved and unassignable.

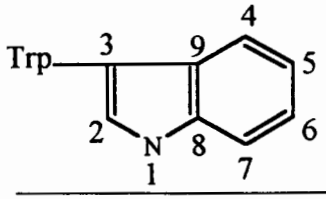
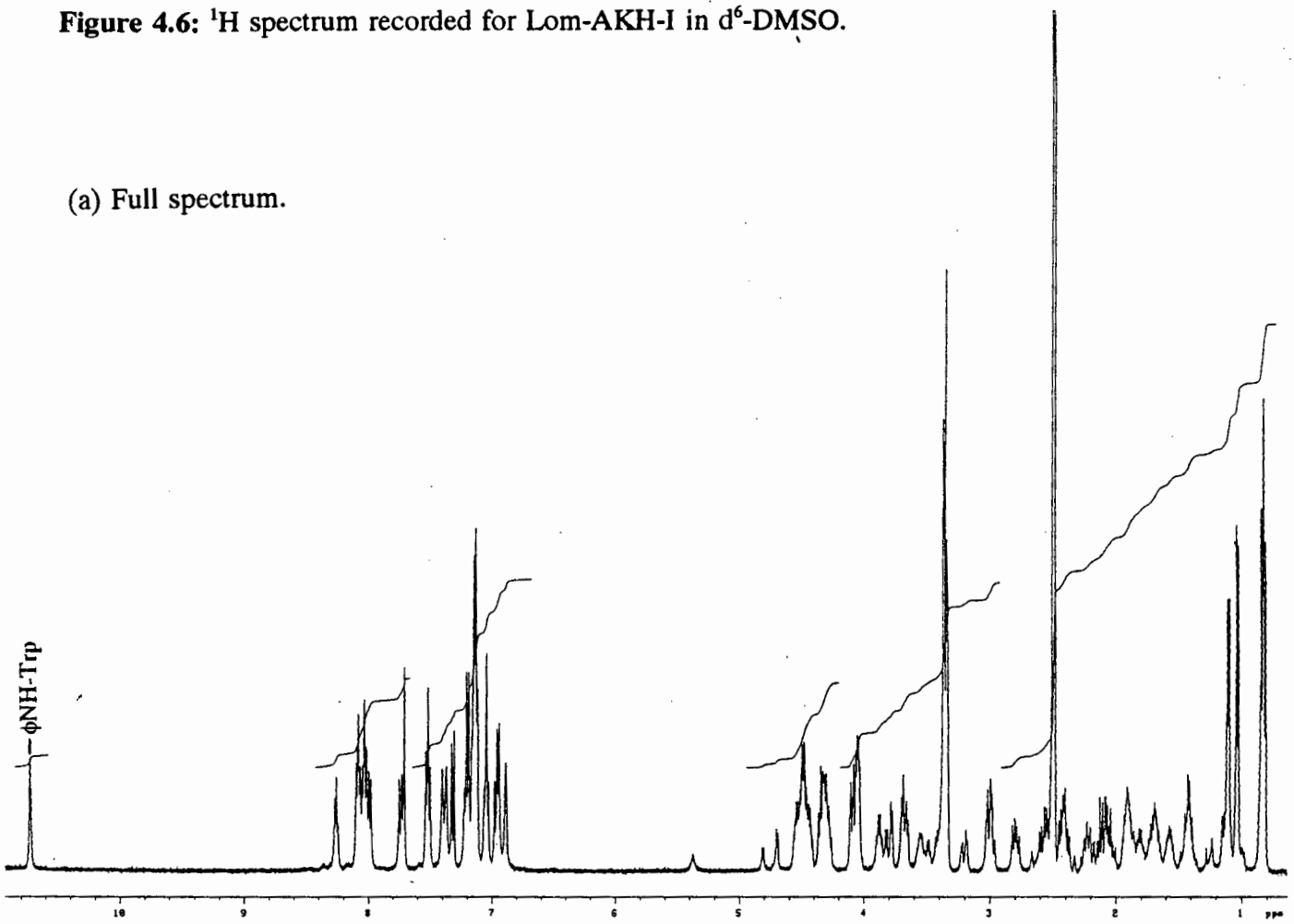
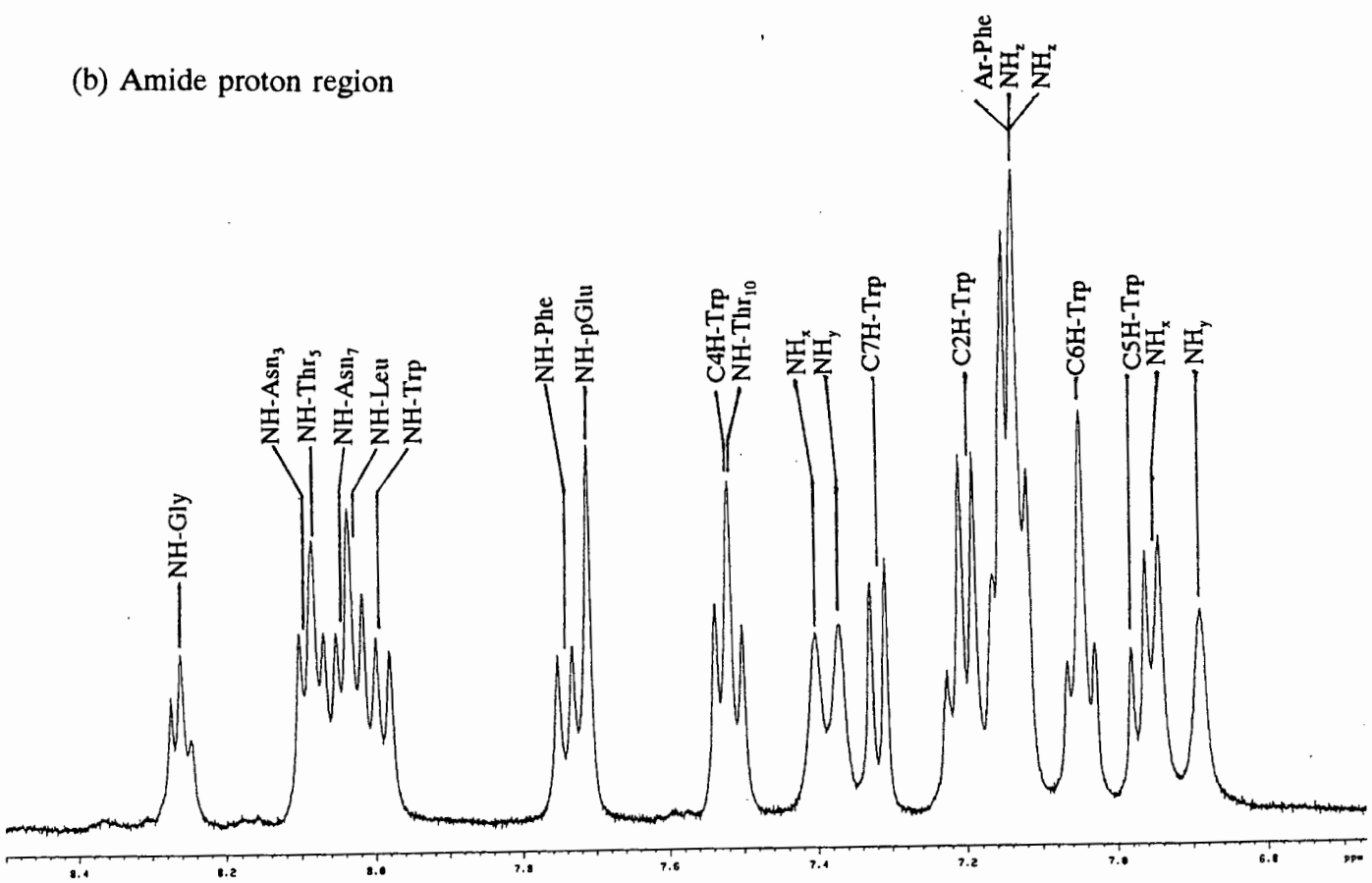


Figure 4.6: ^{1}H spectrum recorded for Lom-AKH-I in d^6-DMSO.

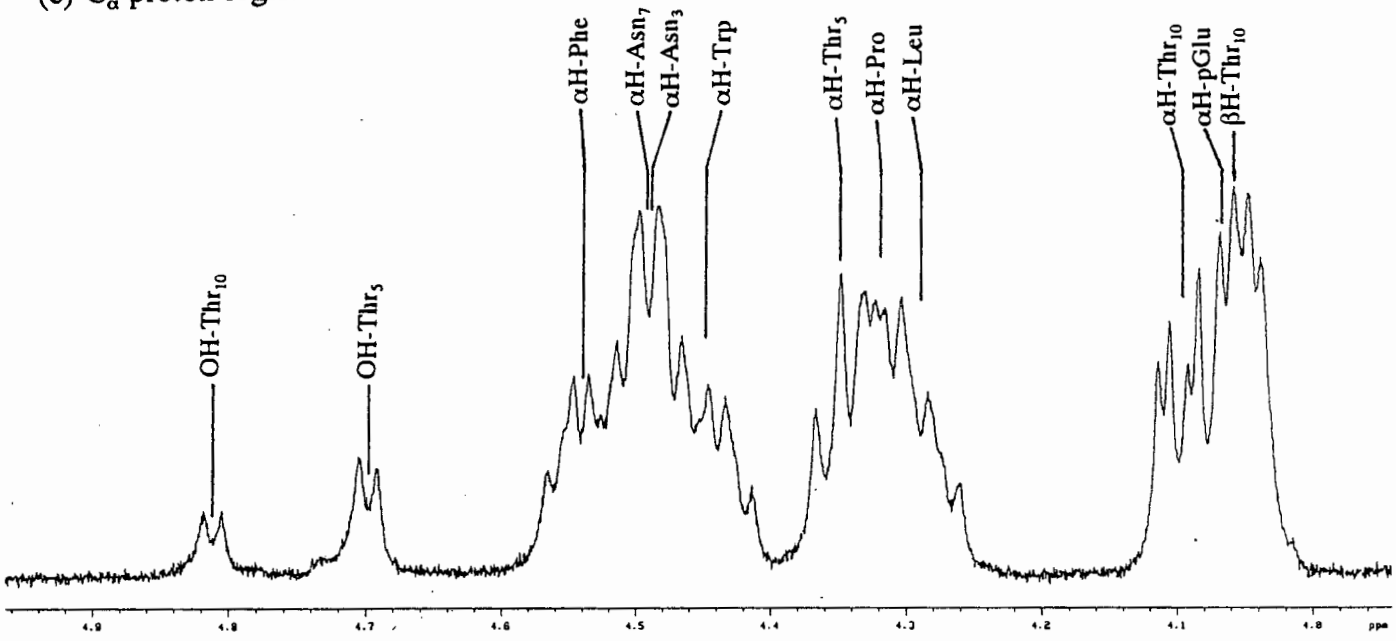
(a) Full spectrum.



(b) Amide proton region



(c) C_α-proton region



(d) Sidechain proton region.

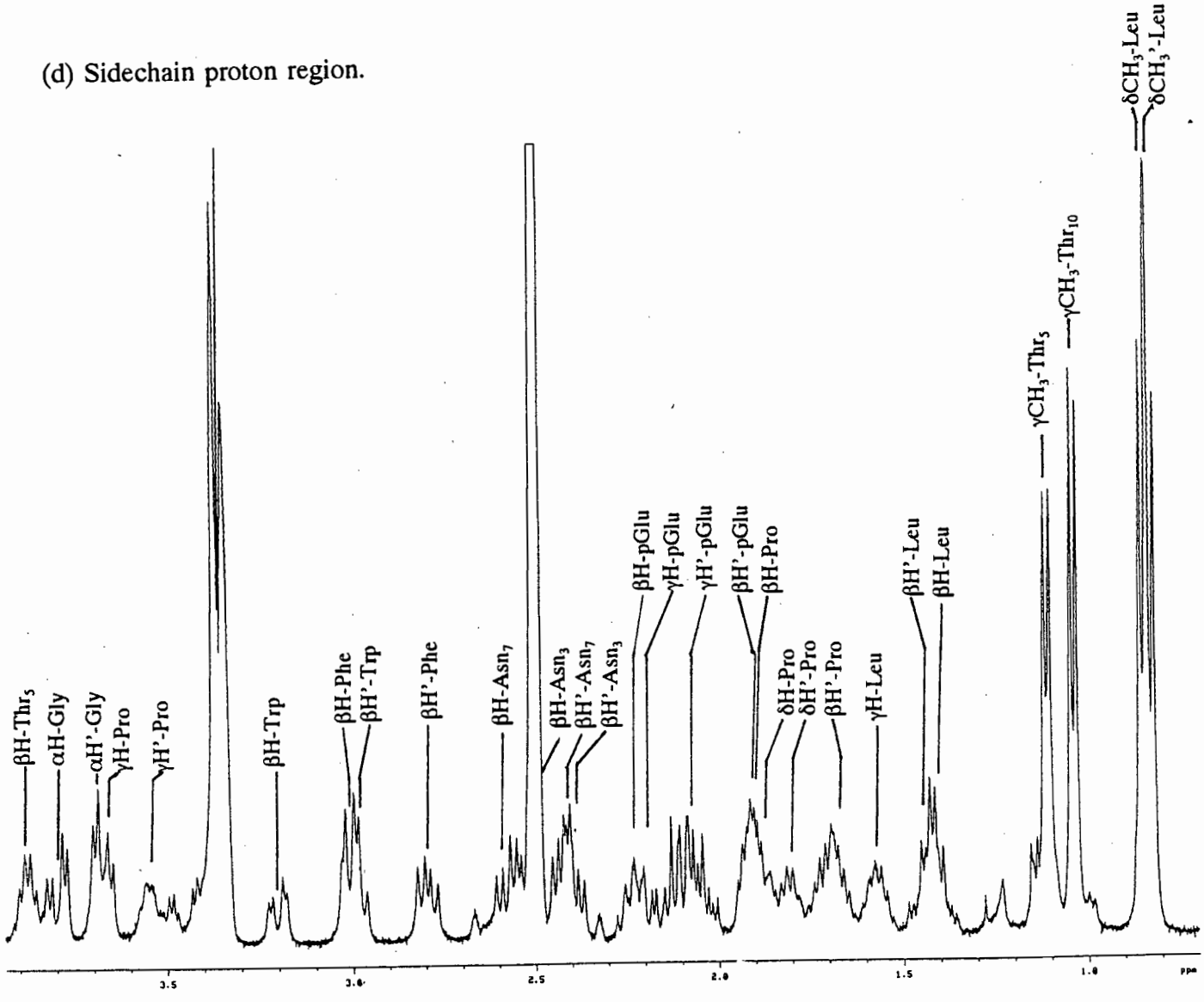


Table 4.2: Literature chemical shift values for amino acid residues found in Lom-AKH-I, II and III recorded in d^6-DMSO (Taken from Wüthrich [2]).

Residue	NH	αH	βH	βH'	Other
Ser	7.98	4.34	3.60	3.60	5.92 OH
Leu	7.95	4.37	( )	( )	( ) 0.89 δCH$_3$ 0.85 δCH$_3$
Asn	8.14	4.61	(2.60)	(2.38)	7.28 CONH$_2$ 6.88
Phe	8.14	4.56	3.03	2.75	7.26 φH
Thr	7.83	4.22	3.93		1.07 γH 4.79 OH
Ala	8.04	4.34			1.22 βCH$_3$
Trp	8.09	4.58	3.16	2.93	10.76 NH 7.17 C2H 7.65 C4H 7.01 C5H 7.10 C6H 7.35 C7H
Gly	8.15	3.76			3.76 αH'

No data was recorded for proline. The bracketed chemical shift values indicate limited spectral resolution did not allow unambiguous assignments to be made.

A study of the NOESY spectrum generated for Lom-AKH-I revealed distinct peaks between the α-proton of Thr_5 and the two δ-protons of the Pro ring, indicating that the *trans* Pro isomer was present. Since a double set of peaks in the proton spectrum was not observed, the sole existence of the *trans* isomer was concluded.

The most common problems associated with running NOESY experiments are the relatively low signal to noise ratios at short mixing times (these can have a significant effect on peak integration values) and the effects of spin diffusion experienced at longer mixing times. For this reason a series of experiments were run at different mixing times, viz. 150, 200, 250 and 300ms, and the resultant data assessed in terms of maximum spectral resolution. The 150ms spectrum showed sufficient nOe build-up, i.e. sufficiently high signal to noise ratios, for

generation of all the relevant crosspeaks. Being the experiment having the shortest mixing time, it was least affected by spin diffusion effects. This led to the 150ms spectrum being the most resolved of the four recorded spectra. Thus, although the fingerprint region of all four spectra was integrated, only the data collected from the 150ms experiment was used in subsequent distant constraint estimations and MORASS calculations. Unfortunately, the poor resolution of crosspeaks in close proximity to the diagonal prevented the amide region (usually used in conjunction with the fingerprint region for secondary structure determinations) of the spectrum from being integrated. The integration values recorded from the fingerprint regions of the various NOESY spectra are listed below in Table 4.3 - see Figure 4.7 for the corresponding peaks.

Table 4.3: NOESY integration values recorded for Lom-AKH-I.

Peak	Integrated Volumes recorded for different Mixing Times (ms)			
	150	200	250	300
A	0.19	0.10	0.24	0.25
B	0.20	0.15	0.20	0.25
C	0.31	0.33	0.65	0.69
D	0.17	0.24	0.33	0.34
E	1.58	1.38	2.98	4.51
F	0.36	0.28	0.53	0.76
G	1.91	1.63	3.84	4.79
H	0.60	0.66	1.38	1.92
I	0.09	0.08	0.22	0.16

The NOESY fingerprint region consisted of eight major peaks, of which two showed severe overlap - to the extent of preventing any accurate assignments or even estimations being made. All other peaks in the spectrum were the result of either NH(i)-αH(i) or αH(i)-NH(i+1) interactions, i.e. no long range interactions were observed. Solving the problem of assigning the overlapping peaks required extensive molecular mechanics calculations. These were used in conjunction with MORASS to generate a structure whose theoretical NOESY showed

significant agreement with that of the experimental one. This enabled predictions of possible interproton interaction sets contributing to each of these peaks, to be made.

One of the primary problems experienced with generating a crude set of distance constraints from the experimental data was the absence of a suitable reference peak in the NOESY spectrum. An ideal peak would have been one generated from the cross relaxation between two geminally coupled protons. All such peaks were, however, located in regions near the diagonal, and thus were extensively overlapped and unusable. Integrated volumes were severely underestimated and therefore insufficient to satisfy the isolated spin pair approximation (ISPA), in which interproton distances are estimated in terms of a reference peak and its corresponding distance (in the case of geminal protons this reference distance would be 1.76Å) - see Chapter 2. As expected, direct substitution into ISPA equation (equation (2.3)) generated distances which violated the van der Waals minimum distance of approach rule [5]. The other drawback was the lack of consideration for spin diffusion effects within the system at these longer mixing times.

Morass takes spin diffusion effects into account when generating a theoretical data set. An overall scaling factor, calculated using the quotient of the sum of the experimental and theoretical data sets, was used to improve accuracy in data scaling. (This was shown by *R*-factors calculated using the **overall** scaling factor which were lower than those calculated using a single reference peak from the NOESY spectrum.) The problem of using a single peak to scale all the data is that the error in any one given integrated peak is unknown and if large, could adversely affect all subsequently scaled data, thus generating an unrealistic data set. MORASS requires a molecular correlation time for calculation of a theoretical NOESY data set. Stokes law [6] makes this possible using the following equation:

$$\tau_c = \frac{4\pi a^3 \eta}{3k_B T} \qquad (4.1)$$

where a is the sphere radius, η, the viscosity, k_B, the Boltzmann constant and T, the temperature. The calculation of τ_c has limited accuracy due to the high degree of flexibility associated with small peptides, which introduces a significant degree of ambiguity in the determination of the molecular radius. Although nOe intensity has been shown to be directly

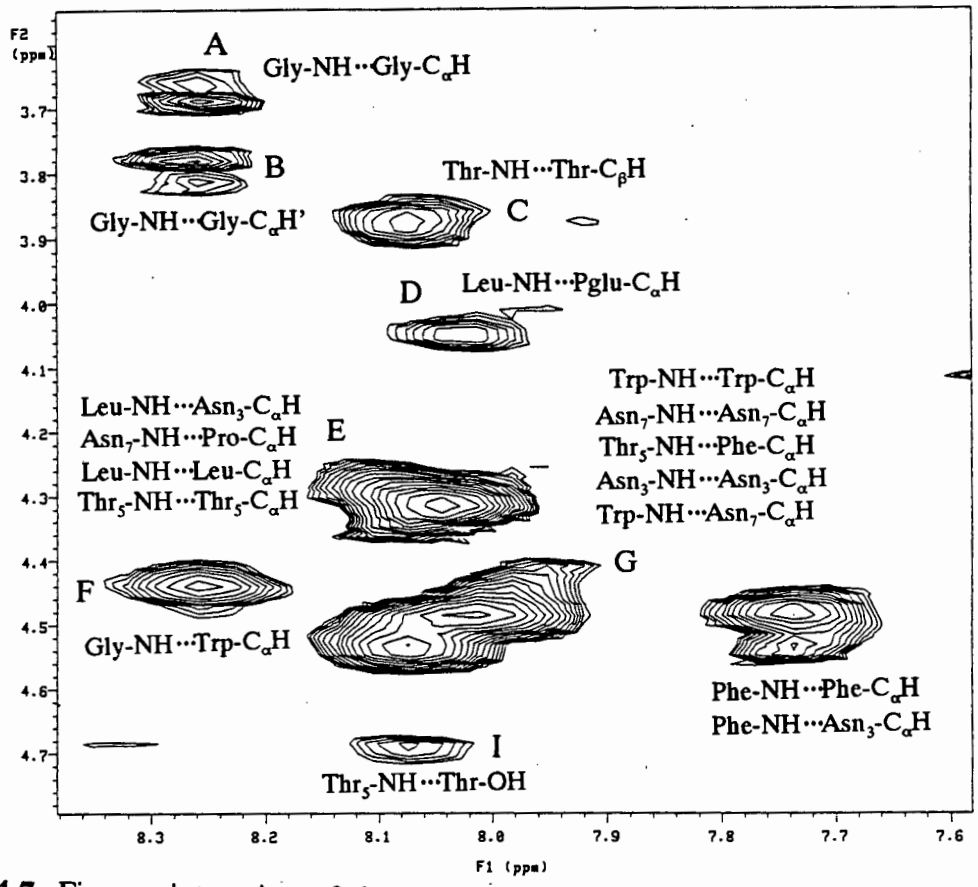


Figure 4.7: Fingerprint region of the NOESY spectrum recorded for Lom-AKH-I in d^6-DMSO.

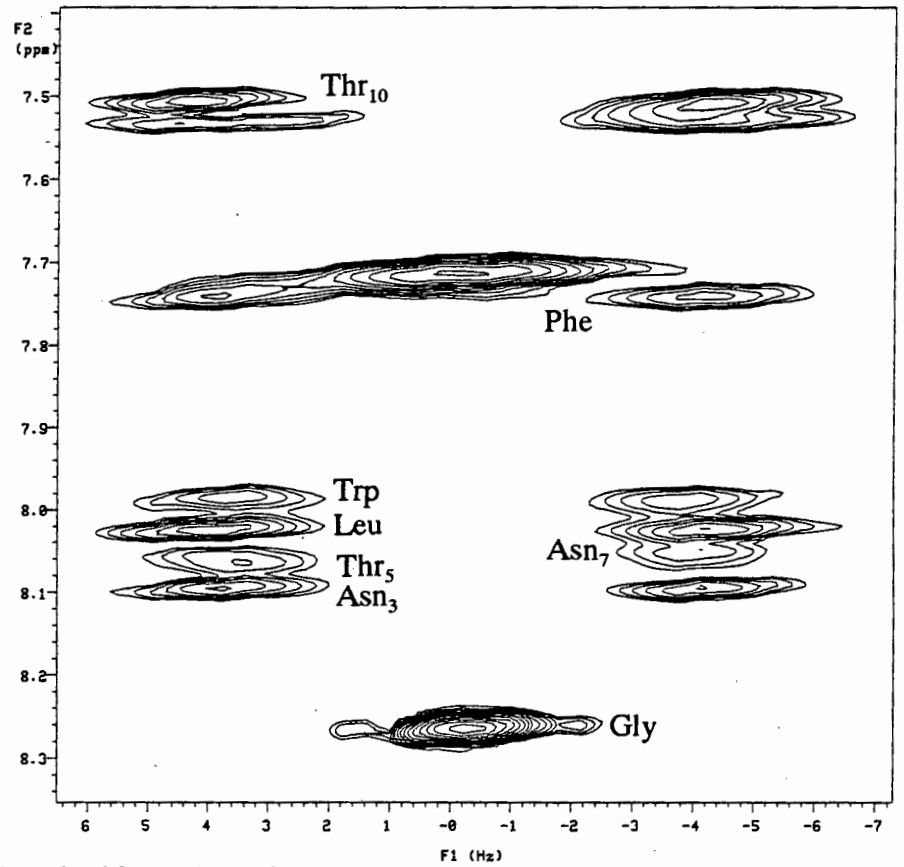


Figure 4.8: Amide region of the Homonuclear 2DJ-Resolved spectrum recorded for Lom-AKH-I in d^6-DMSO.

proportional to τ_c (equation 2.2), high percentage errors in its calculation could be tolerated, since the relative and not the absolute nOe values were of importance in subsequent *R*-factor calculations. For Lom-AKH-I, τ_c was calculated to be 0.88ns.

To generate the ϕ torsion angle constraints, vicinal αH-NH spin-spin coupling constants, $^3J_{NH\alpha}$ had to be determined. These were obtained from the amide region of a Homonuclear 2DJ-Resolved experiment (Figure 4.8), and subsequently substituted into the appropriately modified form of the Karplus equation developed specifically for calculation of ϕ dihedrals by Bystrov (1976) [7] and later by Pardi et al. (1984) [8]. The equation they derived is as follows:

$$.^3J_{HN\alpha} = 6.4\cos^2\theta - 1.4\cos\theta + 1.9 \qquad \textbf{(4.2)}$$

where

$$\theta = |\phi - 60^\circ|. \qquad \textbf{(4.3)}$$

The experimental coupling constants and their corresponding torsion angles are listed below in Table 4.4.

The torsion angle values and initial distance estimates both acquired from the NMR data were submitted as constraints to the various molecular mechanics calculations. An iterative approach was adopted to generate a tighter set of distance constraints i.e. runs were continuously performed with significant manipulation of distance constraints, until reasonable agreement between the experimental and theoretical data was attained. Structural determination of Lom-AKH-I was carried out using both the MacroModel and Biosym packages.

Structures generated from MacroModel were the product of Monte Carlo conformational searches using Batchmin. A total of 3 150 structures were minimised during a 38 CPU hour run. Convergence of the Monte Carlo search was indicated by the high occurrence of the lowest energy conformer in the product structure set. For the duration of the run an energy window of 50kJ/mol was implemented, i.e. only structures having energies within 50kJ/mol of the current global minimum were stored, to prevent collection of a large number of higher

energy, and therefore insignificant, structures. A use-directed search method was implemented in the MC runs, i.e. structures used a minimum number of times were taken as starting geometries (see Chapter 3). The product structures were minimised using the Conjugate Gradient method until a convergence criterion of less than 0.01kJ/Å-mol was satisfied. The 50 lowest energy conformers had energies of between -360.7 and -330.2kJ/mol.

Table 4.4: Coupling Constant and Torsion Angles calculated for Lom-AKH-I.

Residue	Coupling Constant (Hz)	Torsion Angle, ϕ
Leu	8.3	-146.7
Asn_3	7.9	-149.8
Phe	8.1	-148.3
Thr_5	7.3	-155.4
Asn_7	8.0	-148.8
Trp	7.5	-153.8
Gly	5.6	-168.5
Thr_{10}	8.6	-143.1

Biosym's Discover module offered only Molecular Dynamics (MD) as a means of conformational space searching. The optimised straight chain molecule was subjected to steepest descent and conjugate gradient minimisations, followed by restrained MD. The resultant structure was analysed using MORASS, and an *R*-factor was generated. As mentioned above, extensive overlap within the fingerprint region made it impossible to accurately assign distance constraints, or even conclusively predict the contributing interactions to these peaks. There was thus room for a significant amount of manipulation with regards to both the proton pairs contributing to any one given peak, and the actual distance of approach. Constraint manipulation was carried out until a reasonable *R*-factor was obtained, implying acceptable agreement between theoretical and experimental data. The *R*-factor only takes into account peaks which are present in the spectrum. It does not take into account that the absence of a peak from the spectrum implies a minimum interproton distance of 4.5 - 5Å. Thus the viability of each structure produced was also checked in terms of all interactions generated from interproton distances of less than 4.5Å being in agreement with the NOESY spectrum. This above mentioned theoretical minimum distance is often much shorter in practice. In the case of NOESY spectra recorded for Lom-AKH-I the signal to noise ratio was such that the maximum distance of approach that would generate significant peaks in the spectrum was between 3.47 and 3.60Å. The background noise for the 150ms spectrum was recorded to be 0.02 units of integration with a standard deviation, σ, of 0.05. The

theoretical volumes generated by Morass for a particular interaction are structure dependent rather than distance dependent, since the relaxation rate matrix is specific to a structure. Two sets of constraints were thus active in all molecular mechanics runs - those generated from peaks present in the NOESY spectrum and those generated by forcing certain protons further apart to alleviate their theoretical contribution to the spectrum.

Once the *R*-factor had been sufficiently reduced, the relevant constraints were saved and final runs performed. These included three runs all following the same simulated annealing protocol in which the system was initially equilibrated at 450K for 10ps and thereafter structures were collected at 5ps intervals at 300K. One hundred structures were generated in each run, resulting in an overall total of three hundred being collected. Each structure was subsequently minimised as described above. A starting temperature of 300K is relatively low, and it would seem likely that there has been (a) minimal coverage of conformational space and (b) possible trapping of the molecule in a local minimum due to insufficient energy to overcome energy barriers. However, although the first two runs were connected in that the starting structure for the second run was the final structure of the first, a third run was done in order that two of the three runs be started from a straight chain structure. Since each dynamics run started with a different random seed number, the occurrence of similar product structures ceases to be a local minimum problem. Further evidence to support the fact that these structures are indeed low energy conformers, and not just trappings, is that they exhibited very similar structure to the lowest energy conformers generated by the MacroModel Monte Carlo search. A separate dynamics run was also performed at a higher temperature of 600K; however this run yielded structures having very high energy and unlikely conformations.

The 200 structures generated from the two consecutive runs were subsequently divided up into families, and a rigid body *RMS* comparison carried out for each using a sample structure as the template for the calculation (Table 4.5). The first two structure sets were manually divided into families, while the third structure set made use of a more recently acquired "family" module, which grouped structures according to a user specified maximum *RMS* value. An *RMS* value of 0.8 was chosen which generated nine families, of which one was discarded because all members violated the van der Waals minimum distance of approach rule. A sample structure from each family was submitted to MORASS, and their respective *R*-factors

Table 4.5: *RMS* values recorded for the families generated from the first two consecutive MD runs.

Family	*RMS* (Backbone atoms)	*RMS* (Turn region - thr_5 - Asn_7)
A	0.60	0.30
B	0.46	0.15
C	0.72	0.30
D	0.94	0.48
E	0.95	0.38
F	0.45	0.20
G	0.86	0.49
H	0.14	0.12
I	0.18	0.15
J	0.31	0.22
K	0.49	0.58
L	0.12	0.10
M	0.48	0.13
N	0.09	0.07
O	0.31	0.17
P	0.62	0.35

calculated. Each structure was also assessed with respect to NOESY violations. Short interproton distances involving the amide protons of the terminal pGlu and Thr_{10} residues were not considered to be NOESY violations even though MORASS predicted significant nOe's to exist between the relevant proton pairs. The reason for this is that the correlation time of these flexible regions could well be different to that of the bulk of the molecule. This assumption is to a certain extent validated by the complete absence of a NOESY peak corresponding to the pGlu NH-αH interaction. This distance is somewhat fixed by the cyclic nature of the amino acid, and thus should theoretically produce a peak in the NOESY fingerprint region.

Tables 4.6 and 4.7 give a general overview of the results obtained. This table lists the *R*-factors, the energies, NOESY violations, if any, and type and number of H-bonds. The hydrogen bonds were defined as having a maximum donor-acceptor distance of 3.2Å (Pimental and McClellan [9] defined the probable range as being 2.55 to 3.04Å and the average as 2.93Å) and a donor-hydrogen···acceptor angle that lies between 120° and 180°. Peak D labelled in the fingerprint region of the NOESY spectrum (Figure 4.7) corresponds directly to peak D mentioned in Table 4.5. This peak had an integrated volume of 0.17. The four protons possibly contributing to this peak were the α-protons of the Thr_{10} and pGlu residues, and amide protons of the Leu and Asn_7 residues. A detailed study of the interproton distances within the molecule showed which of these interactions existed in a given structure, and thus was the contributor to that nOe. In

Table 4.6: Comparative table of results allowing assessment of the extent of agreement between experimental and theoretical data (first 200 related structures).

Structure	R_1	R_2	Energy (kcal/mol)	# NOESY viol.	Peak D		# H-bonds	hydrogen bond type			D-A dist.	D-H-A angle
					interact.	distance		Donor	Hydrogen	Acceptor		
A	0.13	0.03(1)	199.687	0	pgl-leu	3.61	1	asn_7-N	asn_7-HN	thr_5-O	2.98	120.5
B	0.20	0.04(7)	199.798	1	pgl-leu pgl-asn_7	3.61 3.72	1	asn_7-N	asn_7-HN	thr_5-O	2.91	127.8
C	0.17	0.03(2)	199.716	0	pgl-leu	2.98	0	-	-	-	-	-
D	0.15	0.03(3)	199.646	1	pgl-leu	3.62	1	asn_7-N	asn_7-HN	thr_5-O	2.94	131.9
E	0.17	0.03(6)	200.477	0	pgl-leu	3.63	2	asn_7-N_{sc} thr_{10}-O_{sc}	asn_7-HN thr_{10}-HO	thr_5-O thr_{10}-O	3.14 3.01	143.2 132.4
F	0.17	0.03(6)	199.169	1	pgl-leu	3.62	1	asn_7-N	asn_7-HN	thr_5-O	2.88	142.6
G	0.16	0.03(5)	202.543	0	pgl-leu	3.62	0	-	-	-	-	-
H	0.18	0.03(5)	194.080	2	thr_{10}-leu	2.99	2	pgl-N asn_7-N	pgl-HN asn_7-HN	gly-O thr_5-O	3.17 2.91	161.2 147.2
I	0.18	0.03(3)	200.021	1	thr_{10}-leu	2.99	2	pgl-N asn_7-N	pgl-HN asn_7-HN	gly-O thr_5-O	3.21 2.91	141.8 141.7
J	0.18	0.03(3)	197.769	2	thr_{10}-leu	2.99	2	pgl-N asn_7-N	pgl-HN asn_7-HN	gly-O thr_5-O	3.12 2.90	155.6 143.8
K	0.18	0.03(5)	196.392	2	thr_{10}-leu	3.00	1	asn_7-N	asn_7-HN	thr_5-O	2.95	122.9
L	0.20	0.04(1)	196.035	0	pgl-leu thr_{10}-leu	2.98 2.99	2	pgl-N asn_7-N	pgl-HN asn_7-HN	asn_7-O_{sc} thr_5-O	2.90 3.21	145.7 125.6
M	0.17	0.03(2)	205.579	1	pgl-leu	2.99	0	-	-	-	-	-
N	0.19	0.03(3)	201.290	2	thr_{10}-leu	3.00	0	-	-	-	-	-
O	0.16	0.03(4)	205.403	0	pgl-leu	3.90	0	-	-	-	-	-
P	0.16	0.03(6)	199.871	0	thr_{10}-leu	3.62	0	-	-	-	-	-

Table 4.7: As in table 4.6 except applicable to structures from the third MD run.

Structure	R_1	R_2	Energy (kcal/mol)	# NOESY viol.	Peak D		# H-bonds	hydrogen bond type			D-A dist.	D-H-A angle
					interact.	distance		Donor	Hydrogen	Acceptor		
A1	0.23	0.04(5)	200.479	0	pgl-leu	3.62	1	asn_7-N	asn_7-HN	thr_5-O	2.87	149.8
B1	0.20	0.03(6)	200.117	0	pgl-leu	3.47	2	asn_7-N pgl-N	asn_7-HN pgl-HN	thr_5-O asn_7-O	2.99 3.10	138.6 126.6
C1	0.21	0.04(1)	200.369	0	pgl-leu	3.61	1	asn_7-N	asn_7-HN	thr_5-O	2.89	139.5
D1	0.26	0.04(9)	200.850	0	pgl-leu	3.62	2	asn_7-N thr_{10}-O_{sc}	asn_7-HN thr_{10}-HO	thr_5-O gly-O	2.79 2.79	155.3 149.6
E1	0.27	0.04(8)	200.989	0	pgl-leu	3.61	1	asn_7-N_{sc}	asn_7-HN	thr_5-O	2.94	125.8
F1	0.21	0.04(3)	201.130	0	pgl-leu	3.63	0	-	-	-	-	-
G1	0.25	0.04(7)	203.200	0	pgl-leu	3.62	0	-	-	-	-	-
H1	0.26	0.04(8)	207.500	0	pgl-leu	3.62	1	thr_{10}-O_{sc}	thr_{10}-HO	gly-O	2.71	150.5

certain cases, the distance was found to be greater than the maximum distance allowed for peak generation, i.e. the theoretical NOESY volume was less than the noise level, and therefore should have been discarded. However, the peak did exist, and those were the only interactions which could have generated that peak, and therefore they were included in MORASS calculations. A similar situation was witnessed for peak I in the spectrum which integrated to 0.09 (just distinguishable above the noise level). This peak was the product of cross relaxation between the hydroxyl and amide protons of Thr_5. Again this peak does exist, and so this interaction was also included in the MORASS calculations, even though the distance often exceeded the maximum. In all other cases, only those interactions which generated significant nOes were included.

From Tables 4.6 and 4.7 it was deduced that sample structure A generated the most favourable agreement with the experimental data, and all discussion regarding the structure of Lom-AKH-I will thus be based on this structure (Figure 4.9). The lack of interstrand interactions indicates a total absence of any β sheet structure, parallel or antiparallel, within the peptide. Although a type-I β-turn is the most commonly predicted turn for a Thr-Pro-XY type sequence [3], it is unlikely to be the turn type witnessed in Lom-AKH-I for two reasons: (i) hairpin tight turns are a feature usually associated with antiparallel β sheets [10] and (ii) there is no agreement between the experimental and predicted ϕ torsion angles for such a turn. The predicted angles for a type-I β-turn are approximately -91.4°, -63.3°, and -71.3° for the Thr, Pro and X residues, respectively [11], while the experimental values for Thr, Pro and Asn of Lom-AKH-I were -152.71, -70.67 and -148.34, respectively. As can be seen from Figure 4.9, the turn formed is very open, and the expected Pro-αH···HN-Trp interaction, one of the key interactions required for type-I turn formation, is absent from the NOESY spectrum. The geometrical positioning of the Phe residue at position four in the sequence is one of the major factors contributing to the looseness of the turn region. A side view of the structure shows that it adopts an additional concave conformation (Figure 4.10). From this angle it is also possible to observe the hydrophobic side chain grouping on the convex surface and the hydrophillic grouping on the concave surface (Figure 4.10). This turn region was not only stabilised by an Asn-N-H ... O-Thr_5 hydrogen bond, but also extremely well preserved relative to the flexible peptide termini. This was confirmed by both the structure overlay of the family corresponding structure A (Figure 4.11) and a hundred structure overlay containing

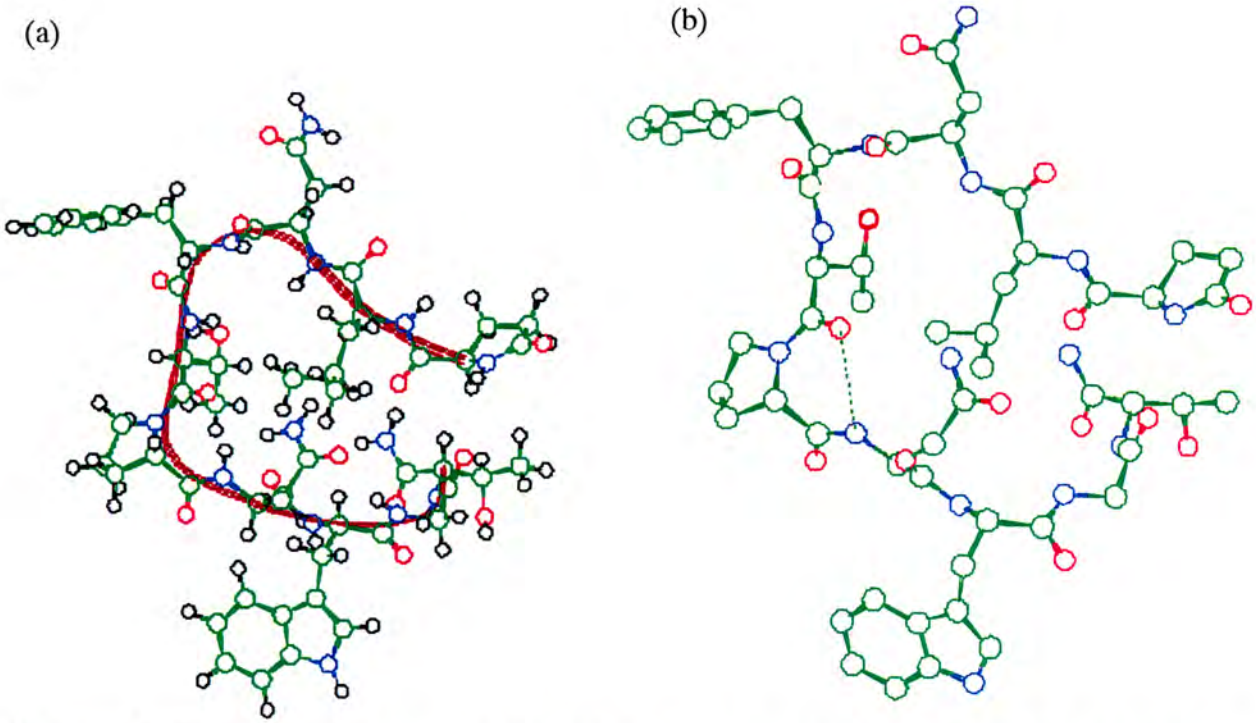


Figure 4.9: Overview of Structure A (a) with the hydrogens (backbone region is highlighted by a brown ribbon), and (b) without the hydrogens, illustrating the open turn adopted by the backbone of Lom-AKH-I.

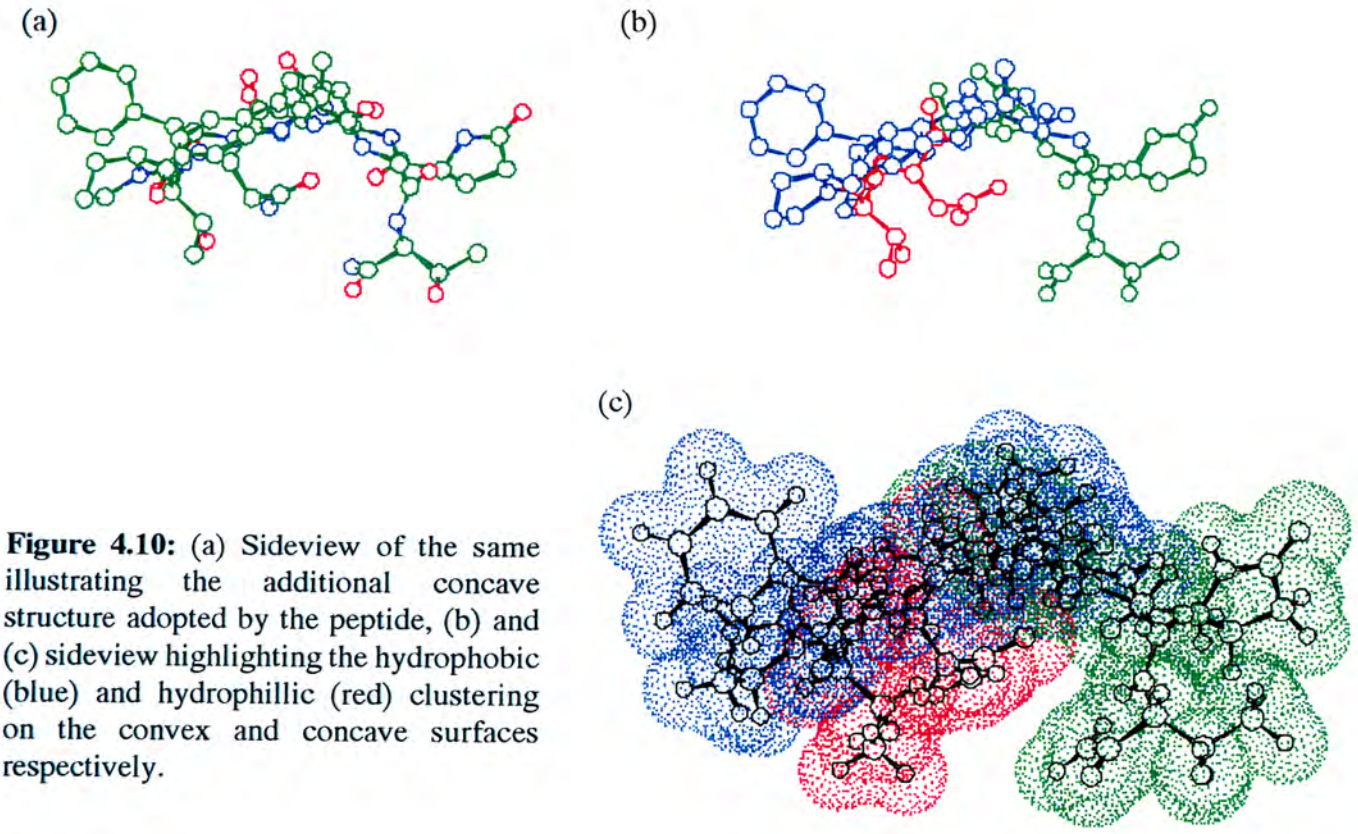


Figure 4.10: (a) Sideview of the same illustrating the additional concave structure adopted by the peptide, (b) and (c) sideview highlighting the hydrophobic (blue) and hydrophillic (red) clustering on the convex and concave surfaces respectively.

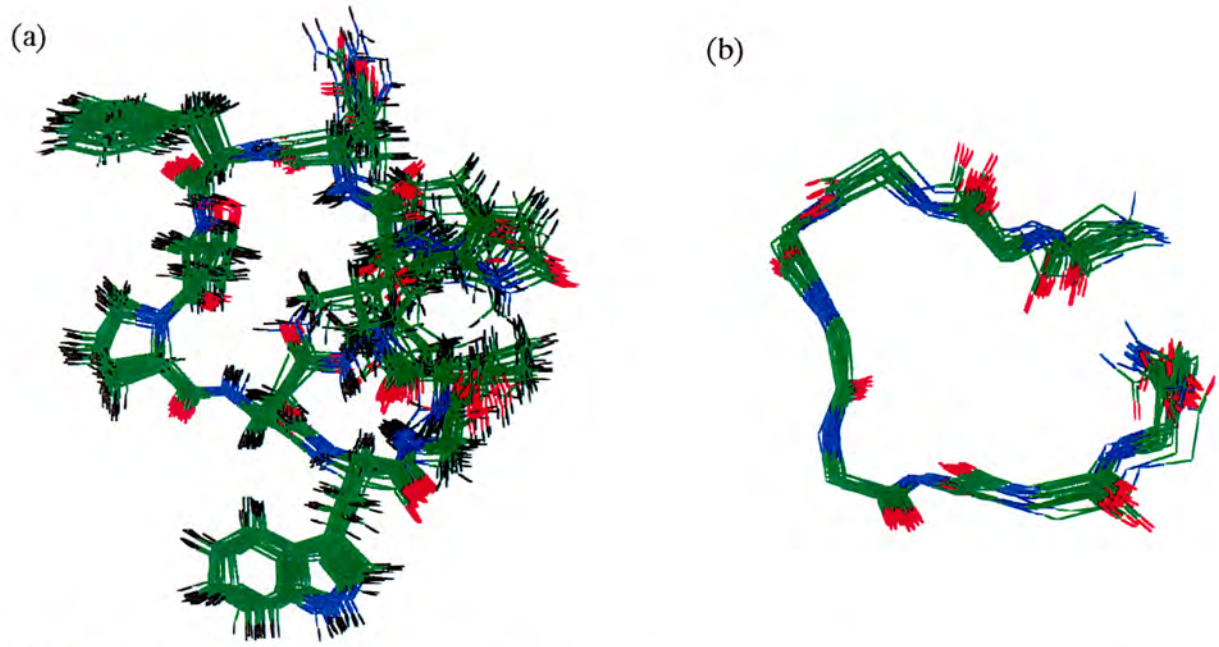


Figure 4.11: Overlay corresponding to the family associated with structure A showing the high conservation of the turn region relative to the flexible termini - (a) full structure overlay, and (b) backbone overlay.

Figure 4.12: One hundred structure overlay for the first MD run. This again illustrates the extent of turn conservation relative to the flexible termini.

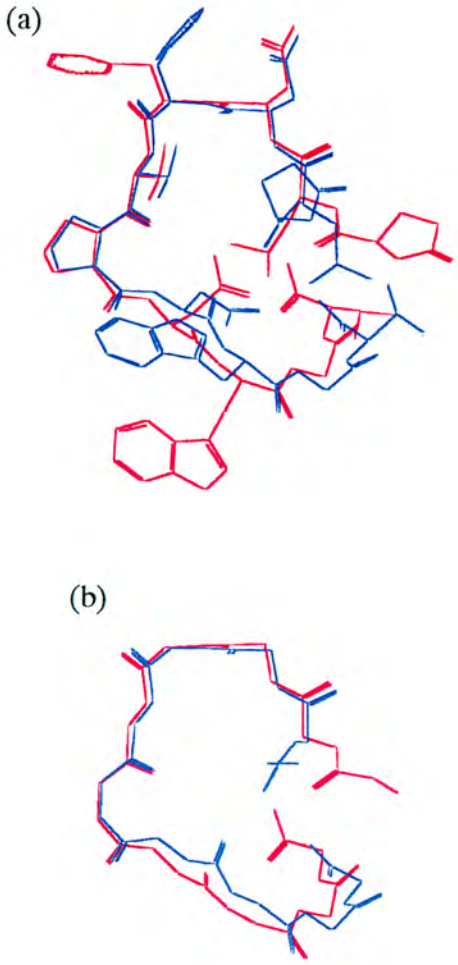


Figure 4.13: Overlays of (a) the full structures (excluding hydrogens) and (b) the backbones of structures A (red) and B1 (blue) showing the good agreement and conservation of secondary structure between the results obtained from the first and third MD runs.

structure A (Figure 4.12). Again the hydrogen bond fails to be consistent with the one expected in a type-I turn (between the carbonyl oxygen of residue i and the amide proton of residue (i+3) of the turn). The coupling constants obtained for this peptide further support the idea that this peptide, although highly structured, does not exhibit any form regular secondary structure.

Table 4.8: RMS values generated for various backbone sections.

Region	RMS value
Thr_5 - Asn_7	0.31
Phe - Trp	0.36
Asn_3 - Gly	0.42
Leu - Thr_{10}	0.54
pGlu - NH_2	0.85

A progressive increase in *RMS* value observed when calculated from the turn region outwards for the family associated with structure A (Table 4.8) indicated good conservation of the turn region relative to the high flexibility observed on approaching the termini. Further evidence for preservation of this turn, this time in structures generated during the third MD run (in which the original straight chain structure constituted the starting structure), was illustrated by the low *RMS* values recorded in the turn region between structures A and B1 (the "best" structure generated in the run of interest (Figure 4.13)). The *RMS* deviation in aligned positions for the backbone was 0.31 from the Thr_5 to Asn_7 residues and 0.47 from the Phe to Trp residues. Figure 4.14 contains an overview and sideview of structure B1 illustrating the similarity between this and structure A.

The MORASS calculations allowed complete assignment of the overlapping peaks in the fingerprint region of the NOESY spectrum. Solution sets are governed by the *R*-factor, and thus there exists no unique solution since a variety of structures may exist, all of which give rise to satisfactory *R*-factors. This was demonstrated by the two different solution sets generated by structure A and the lowest energy MacroModel conformer. Table 4.9 gives a full breakdown of the NOESY fingerprint region for structure A, in terms of what interactions contribute as well as the extent the extent of their contribution to any one given peak. It also lists the corresponding interproton distances. Table 4.10 illustrates the good agreement between the experimental and theoretical torsion angle data. Thus, although structure A may not necessarily be the lowest energy structure, it shows satisfactory agreement with all the available experimental data.

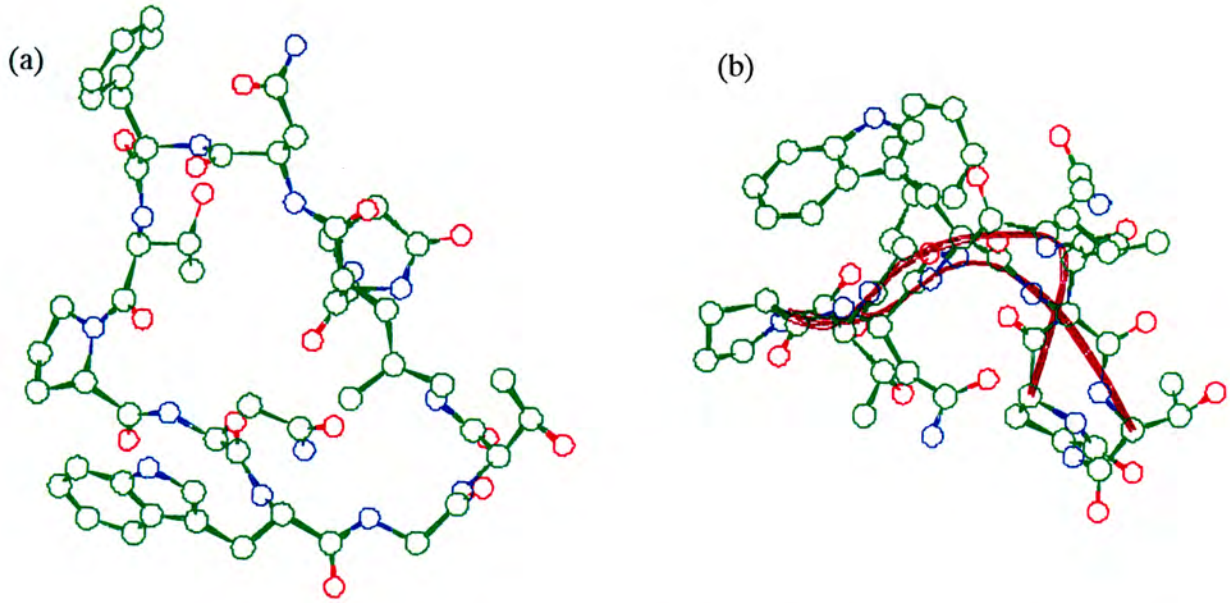


Figure 4.14: (a) Overview and (b) sideview of B1 (backbone highlighted by the brown ribbon) illustrating the retention of concavity and hydrophobic and hydrophillic clustering on the concave and convex surfaces, respectively.

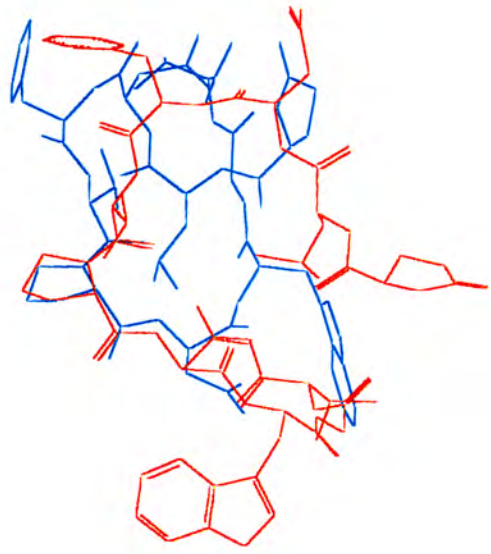


Figure 4.15: Overlay of the lowest energy Macromodel structure (blue) and structure A (brown) showing relatively good agreement in the direct turn region (Thr_5 to Asn_7) but significant deviation of the former from A towards to termini. Both structures have the same H-bonding scheme in the turn region.

Table 4.9: NOESY breakdown in accordance to results obtained from Morass for structure A.

Peak	Interaction	Dist.	NOESY Integration		
			exp.	theor.	diff.
	Thr_5-αH ··· Pro-$δH_1$	2.75	0.36	0.30	0.06
	Thr_5-αH ··· Pro-$δH_2$	2.39	0.57	0.59	-0.02
A	Gly-NH ··· Gly-$αH_1$	2.99	0.19	0.19	0.00
B	Gly-NH ··· Gly-$αH_2$	2.79	0.20	0.26	-0.06
C	Thr_5-NH ··· Thr_5-βH	2.60	0.31	0.42	-0.11
D	Leu-NH ··· pGlu-αH	3.61	0.17	0.06	0.11
E	Asn_3-NH ··· Leu-αH	2.42	0.53	0.64	-0.11
	Asn_7-NH ··· Pro-αH	2.43	0.64	0.66	-0.02
	Leu-NH ··· Leu-αH	3.01	0.20	0.18	0.02
	Thr_5-NH ··· Thr_5-αH	3.03	0.20	0.17	0.03
F	Gly-NH ··· Trp-αH	2.95	0.36	0.21	0.15
G	Trp-NH ··· Asn_7-αH	2.42	0.65	0.65	0.00
	Thr_5-NH ··· Phe-αH	2.45	0.65	0.64	0.00
	Asn_3-NH ··· Asn_3-αH	2.77	0.30	0.30	0.00
	Asn_7-NH ··· Asn_7-αH	3.00	0.17	0.18	-0.01
	Trp-NH ··· Trp-αH	3.06	0.13	0.16	-0.03
H	Phe-NH ··· Asn_3-αH	2.51	0.48	0.56	-0.07
	Phe-NH ··· Phe-αH	3.00	0.20	0.20	0.00
I	Thr_5-NH ··· Thr_5-OH	3.72	0.09	0.06	0.03

Structure A was also used in *R*-factor calculations for the different NOESY mixing times (Table 4.11). The satisfactory results obtained further emphasise an overall good agreement between the experimental and theoretical data.

The lowest energy conformation generated by MacroModel (Figure 4.14), A2, was submitted to MORASS, and *R*-factors were subsequently calculated. Both the *R*-factor of 0.27

($R^{1/6}$ factor = 0.05(2)) and the absence of any NOESY violations indicated satisfactory agreement between the theoretical and experimental data. This conformation showed some definite structural similarities to those generated by Biosym. The above discussed open turn was again well preserved - a comparison between structures A and A2 showed the *RMS* deviation in aligned positions for the backbone from the Thr_5 to Asn_7 residues to be 0.42 (Figure 4.15). Two hydrogen bonds were observed in this structure. The one occurred between the Asn_7 amide and the Thr_5 carbonyl groups (donor-acceptor distance = 2.92Å and the N-H ··· O angle = 146.8°) and was thus analogous to that observed to maintain the turn about the Pro residue in the Biosym structures. The other occurred between the pGlu amide the Gly carbonyl groups (donor-acceptor distance = 3.18Å and N-H ··· O bond angle = 152.5°). Although the concave nature of the molecule was conserved, the hydrophobic and hydrophillic clusters were largely absent (only the Trp and Pro side chains were located on the convex surface) due to an alternative arrangement of the more flexible terminal regions of the peptide (Figure 4.16). An overlay of the ten lowest energy MacroModel structures is presented in Figure 4.17.

Table 4.10: A listing of the theoretical torsion angles for both structures A and the lowest energy structure generated by MacroModel (A2).

Residue	Torsion Angle, ϕ		
	Exp.	Theor. (A)	Theor. (A2)
Leu	-146.7	-146.2	-145.2
Asn_3	-149.8	-151.4	-153.1
Phe	-148.3	-147.1	-145.4
Thr_5	-155.4	-152.7	-154.2
Asn_7	-148.8	-148.3	-148.9
Trp	-153.8	-147.7	-145.1
Gly	-168.5	-166.9	-168.1
Thr_{10}	-143.1	-140.5	-140.6

The solution set generated by Morass for this structure showed significant differences in the regions of overlap to that generated for the Biosym structures. This confirms the fact that it is impossible to generate a unique solution to any given overlapping peak, and that, even in the event of there being regions of structural conservation, there was is space for ambiguity. Table 4.12 gives a full solution to the NOESY fingerprint region for the MacroModel

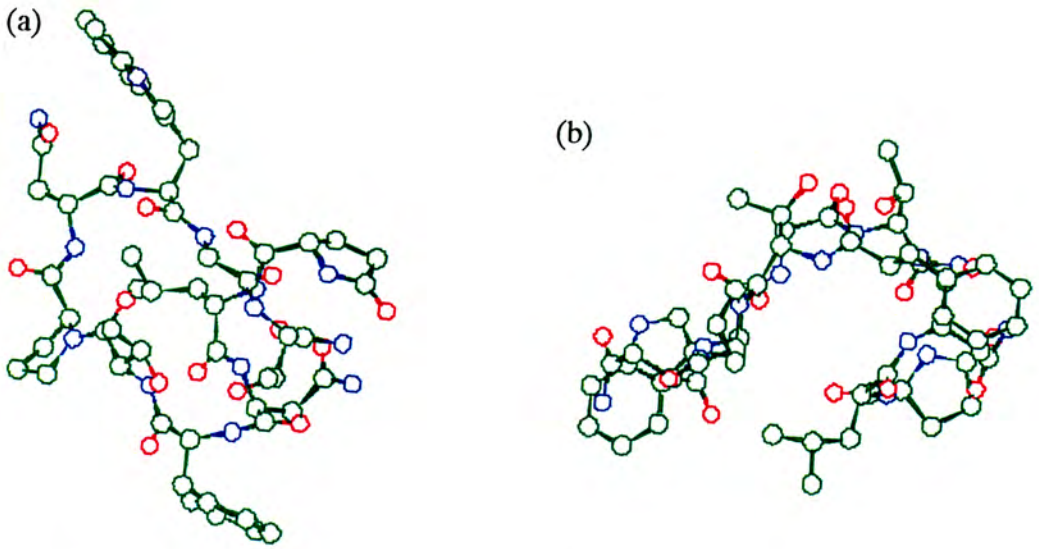


Figure 4.16: (a) Overview and (b) sideview of the lowest energy Macromodel structure; the latter showing the discontinued existence of the large hydrophobic clustering on the convex surface due to extensive conformational changes in the backbone regions away from the turn.

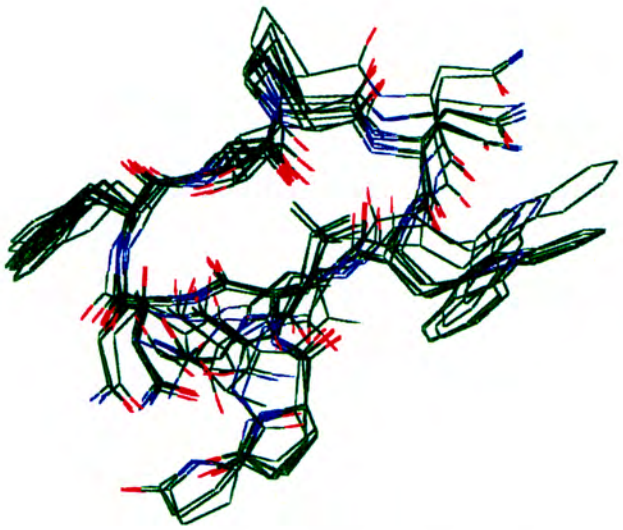


Figure 4.17: Overlay of the ten lowest energy structures produced by the Macromodel Monte Carlo search.

Table 4.11: R-factors generated at different NOESY mixing times for structure A.

Mixing Time (ms)	R-Factor	$R^{1/6}$-Factor
150	0.14	0.03(5)
200	0.23	0.05(6)
250	0.23	0.05(4)
300	0.21	0.04(7)

structure. The good agreement observed between the theoretical and experimental torsion angle data is illustrated in Table 4.9, and provides further evidence for the legitimacy of this structure.

The existence of close key structural similarities between the Biosym and MacroModel output structures indicates that the search has not been limited by the relatively low temperature dynamics used in Biosym. The MC search provided an effective means of searching conformational space and therefore the problem regarding the trapping of a structure in a local minimum may be discounted. It may thus be concluded that the overall features generated for Lom-AKH-I are indeed the product of an extensive conformational space search. The fact that it was possible to generate a single structure which satisfied all the imposed experimental constraints, is an indication of possible high peptide reactivity. Dose response curves generated for Lom-AKH-I have indeed shown this to be the case, and an extremely high reactivity has been reported for this peptide [12] - see later.

Table 4.12: NOESY breakdown in accordance to Morass for structure A2.

Peak	Interaction	Distance	NOESY Integration		
			Exp.	Theor.	Diff.
	Pro-δH_1 ··· Thr_5-αH	2.36	0.36	0.65	-0.29
	Pro-δH ··· Thr_5-αH	2.40	0.57	0.59	-0.02
A	Gly-NH ··· Gly-αH_1	2.90	0.19	0.21	-0.02
B	Gly-NH ··· Gly-αH_2	2.78	0.20	0.28	-0.08
C	Thr_5-NH ··· Thr_5-βH	2.84	0.31	0.23	0.08
D	Leu-NH ··· pGlu-αH	3.61	0.17	0.07	0.10
E	Asn_3-NH ··· Leu-αH	2.75	0.44	0.30	0.14
	Thr_5-NH ··· Pro-αH	3.21	0.10	0.12	-0.02
	Asn_7-NH ··· Pro-αH	2.48	0.45	0.53	-0.08
	Thr_5-NH ··· Leu-αH	3.43	0.16	0.08	0.08
	Leu-NH ··· Leu-αH	2.93	0.21	0.20	0.01
	Thr_5-NH ··· Thr_5-αH	2.90	0.21	0.21	-0.00
F	Gly-NH ··· Trp-αH	2.52	0.36	0.52	-0.16
G	Trp-NH ··· Asn_7-αH	3.49	0.10	0.08	0.02
	Asn_3-NH ··· Phe-αH	3.44	0.20	0.08	0.12
	Thr_5-NH ··· Phe-αH	2.25	0.83	0.94	-0.11
	Asn_7-NH ··· Trp-αH	2.96	0.16	0.19	-0.04
	Asn_3-NH ··· Asn_3-αH	2.95	0.20	0.20	0.00
	Asn_7-NH ··· Asn_7-αH	2.89	0.22	0.23	-0.01
	Trp-NH ··· Trp-αH	2.74	0.23	0.31	-0.08
H	Phe-NH ··· Asn_3-αH	3.36	0.30	0.10	0.20
	Phe-NH ··· Phe-αH	2.97	0.28	0.19	0.09
I	Thr_5-NH ··· Thr_5-OH	3.56	0.09	0.07	0.02

REFERENCES

1. Stone, J.V.; Mordue, W.; Batley, K.E.; Morris, H.R. *Nature*, **1976**, *263*, 207-211.

2. Wüthrich, K. *NMR in Biological Research: Peptides and Proteins*, North-Holland Publishing Company, Amsterdam, 1976.

3. Suzuki, M.; Gernstein, M.; Johnson, T. *Protein Engineering* **1993**, *6(6)*, 565-574.

4. Evans, P.A.; Dobson, C.M.; Kautz, R.A.; Hatfull, G.; Fox , R.O. *Letters to Nature* **1987**, *329*, 266-268.

5. Homans, S.W. *NMR of Macromolecules: A Practical Approach* (Roberts G.C.K, ed.); Oxford University Press, Oxford, 1993, 289-314.

6. Atkins, P.W. *Physical Chemistry*, Oxford University Press, Oxford, 1982.

7. Bystrov, V.F. *Progr. Nucl. Magn. Reson. Spectros.* **1976**, *10*, 41-82.

8. Pardi, A.; Billeter, M.; Wüthrich, K. *J. Mol. Biol.* **1984**, *180*, 741-751.

9. Pimentel, G.L.; McClellan *The Hydrogen Bond*, Freeman San Francisco, 1960, 255-295.

10. Wüthrich, K. *NMR of Proteins and Nucleic Acids*, Wiley: New York, 1986.

11. Suzuki, M.; Yagi, N. *Proc. R. Soc. Lond.* **1991**, *246*, 231-235.

12. Goldsworthy, G.J.; Mallison, K.; Wheeler, C.H. *J. Insect Physiol.* **1986**, *32(1)*, 92-101.

CHAPTER 5

LOM-AKH-II: RESULTS AND DISCUSSION

In 1979 Carlson *et al.* [1] isolated a second peptide exhibiting adipokinetic activity from the *corpus cardiacum* of both *Locusta migratoria* and *Schistocerca gregaria*. This was confirmed by Gäde in 1981 [2]. Sequencing of this new peptide revealed two octapeptides differing by one amino acid substitution [3,4]. The one isolated from *Locusta migratoria* was subsequently added to the Lom peptide series as Lom-AKH-II (Figure 5.1), while the other, in which there was a Thr-Ser substitution, was named Scg-AKH-II. It was Lom-AKH-II that was further investigated with regard to secondary structure determination. The major difference between this peptide and the other series members was the distinct absence of a proline residue. A 0.0034M solution of this peptide dissolved in d^6-DMSO was prepared. The temperature arrayed proton experiment yielded maximum resolution at 298K and all subsequent NMR experiments, COSY, TOCSY, NOESY and Homonuclear 2DJ-Resolved experiments, were performed at this temperature.

Assignment of the proton spectrum was carried out using COSY (Figure 5.2) and TOCSY (Figures 5.3 and 5.4) spectra. The whole assignment process was facilitated not only by the chemical shift assignments recorded for Lom-AKH-I, but also by the fact that all residues in this peptide were unique and generated different coupling patterns, with the exception of the three AMX spin systems of the Asn, Phe and Trp residues respectively. Thus there was reduced peak overlap in the various 2D experiments. The β-protons of the Asn residue were assumed to resonate further upfield than those of the Phe and Trp residues, and the resultant complete chemical shift set obtained for this residue was in close agreement with that obtained for Asn_3 of Lom-AKH-I and corresponding literature values (Tables 4.1 and 4.2). Indisputable differentiation between the Phe and Trp residues proved to be impossible due to solvent suppression in the TOCSY using the transmitter, which avoided inclusion of the ϕNH-Trp proton in the spectrum in order to enhance spectral resolution. Even if this region had

LOM-AKH-II: pGlu-Leu-Asn-Phe-Ser-Ala-Gly-TrpNH2

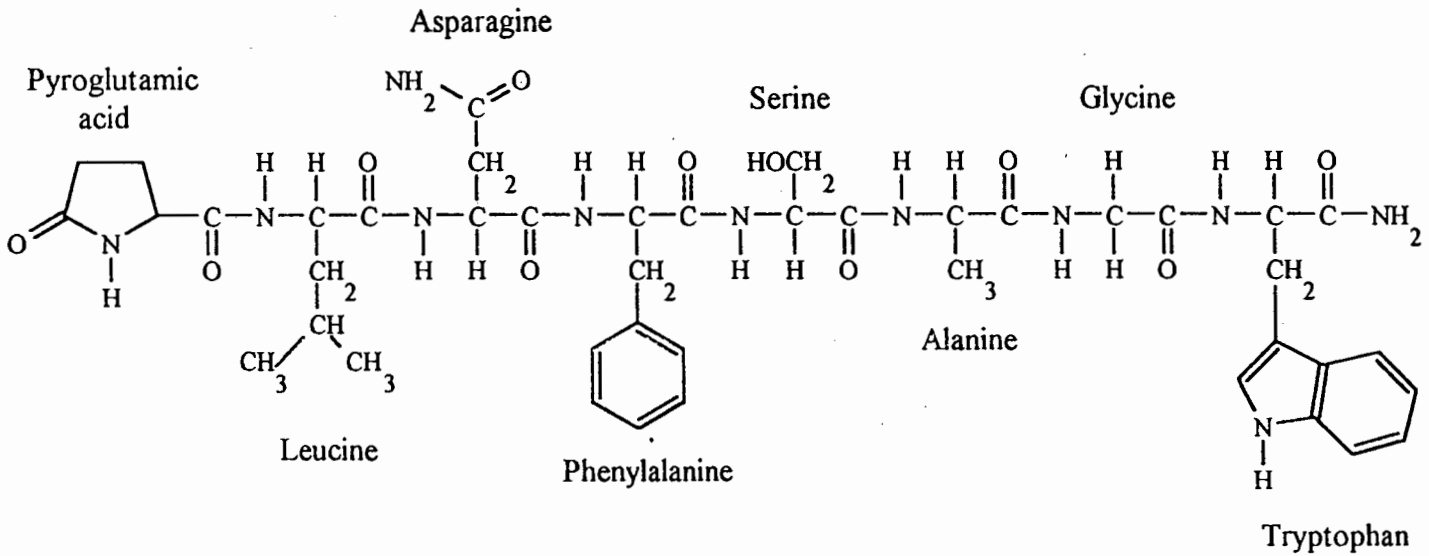


Figure 5.1: Primary sequence recorded for Lom-AKH-II.

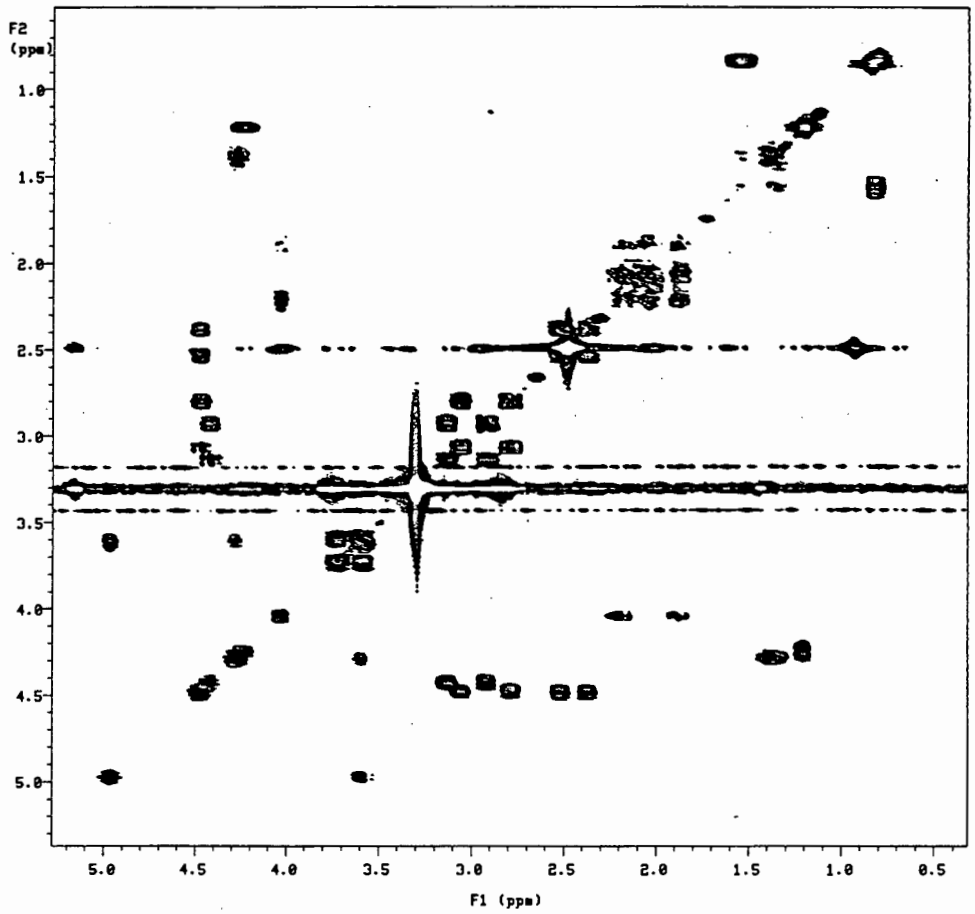


Figure 5.2: C_α-side chain region of the COSY spectrum recorded in d^6-DMSO for Lom-AKH-II.

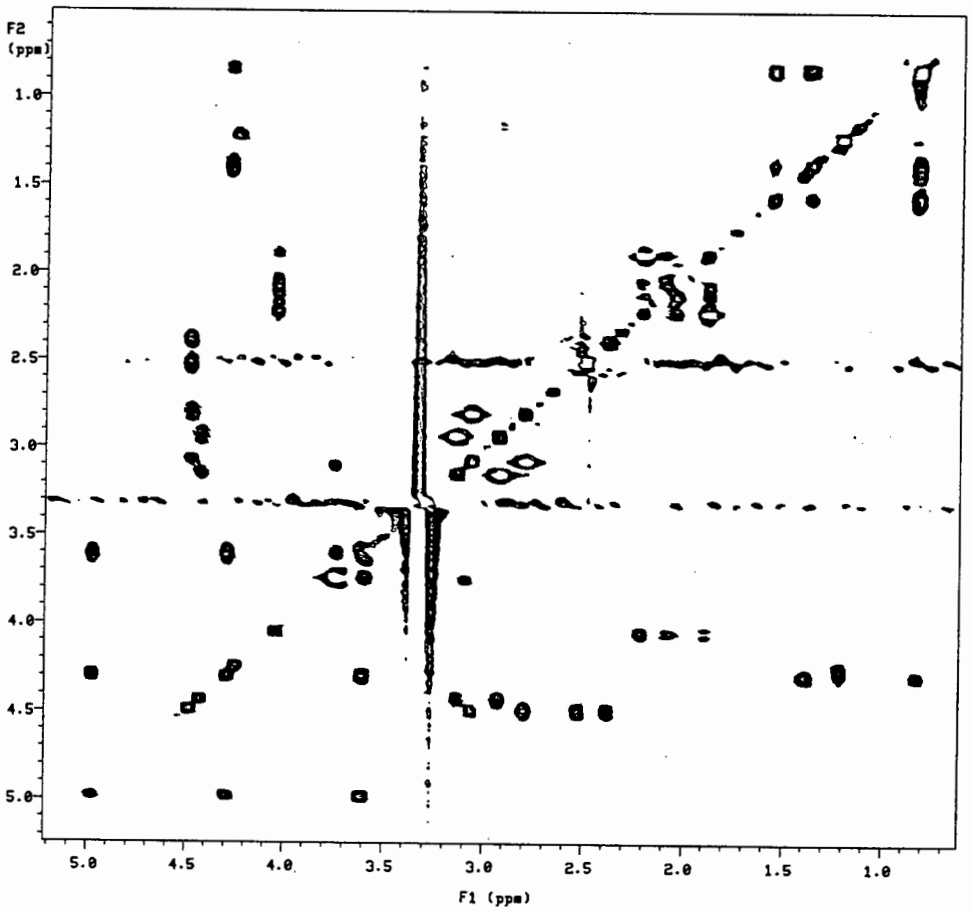


Figure 5.3: C_α-side chain region of the TOCSY spectrum recorded in d^6-DMSO for Lom-AKH-II.

been included, there is no guarantee that the relevant coupling would have been observed (the TOCSY generated for Lom-AKH-III did include this region but showed no coupling of the ϕNH-Trp protons to any of the relevant NH, C_α or C_β protons). The first possible assignment was based on a comparison with the unambiguous amide chemical shift values obtained for Phe and Trp of Lom-AKH-I, and the second on more feasible interactions in the NOESY spectrum. Although in retrospect the latter seems more reasonable, at the time of assignment only the former was considered, and all further studies were therefore carried out using these assignments.

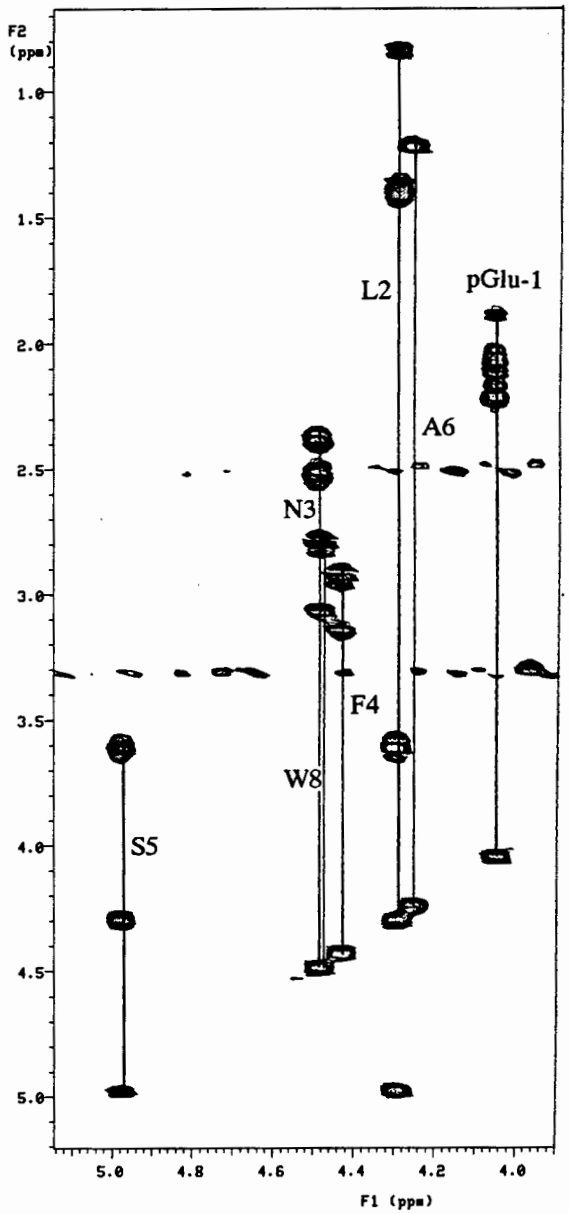


Figure 5.4: TOCSY spectrum (recorded with a mixing time of 80ms) showing the relay scalar connectivities between the C_α- and side chain protons of Lom-AKH-II.

The methyl groups associated with Leu and Ala generated distinctive upfield peaks and therefore made the location of these residues immediately apparent. Their different spin systems, A_3B_3MPTX and A_3X for Leu and Ala respectively, further allowed them to be distinguished from each other. The Ser hydroxyl proton, resonating downfield of the α-proton region, showed relay scalar connectivities (present in the TOCSY spectrum) to the α- and β-protons, and allowed not only location of the Ser residue, but also distinguished it from the other three AMX spin systems. The simple AX Gly spin system was readily identifiable in both the COSY and TOCSY spectra, as were the couplings associated with the pGlu residue (the TOCSY spectrum showed superior resolution for accurate assignment of the pGlu protons). The pGlu, Leu and Gly proton chemical shifts all showed good agreement with those obtained for Lom-AKH-I and the literature values.

Table 5.1: Chemical shift values recorded for Lom-AKH-II in d^6-DMSO.

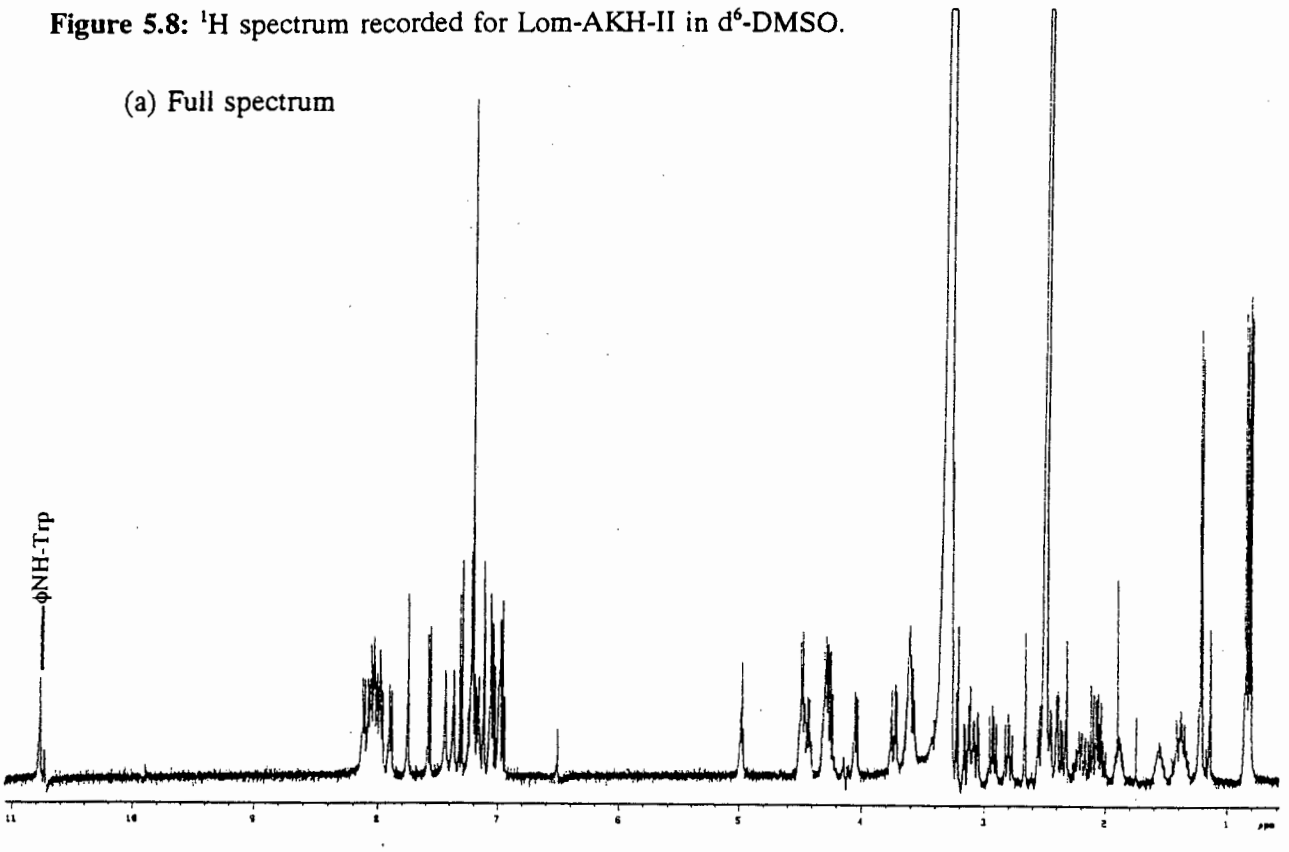
Residue	Amide	$C_\alpha H$	$C_\beta H$	$C_\beta H'$	Other
pGlu	7.757	4.021	2.208	1.872	2.110 γH 2.050 γH'
Leu	8.034	4.275	1.357	1.357	1.562 γH 0.858 δCH_3 0.822 δCH_3'
Asn	8.115	4.498	2.519	2.370	
Phe	7.894	4.429	3.126	2.932	7.12-7.26 φH's
Ser	8.068	4.290	3.620	3.596	4.98 OH
Ala	7.960	4.239			1.213 CH_3
Gly	8.020	3.710			3.562 $C_\alpha H'$
Trp	7.984	4.488	3.059	2.800	10.880 φNH 7.120 C2H 7.580 C4H 6.965 C5H 7.045 C6H 7.310 C7H

The four protons of the two NH_2 groups were assignable at 7.38;7.06 and 7.44;7.00, respectively. It was impossible to assign these two groups to any specific residue.

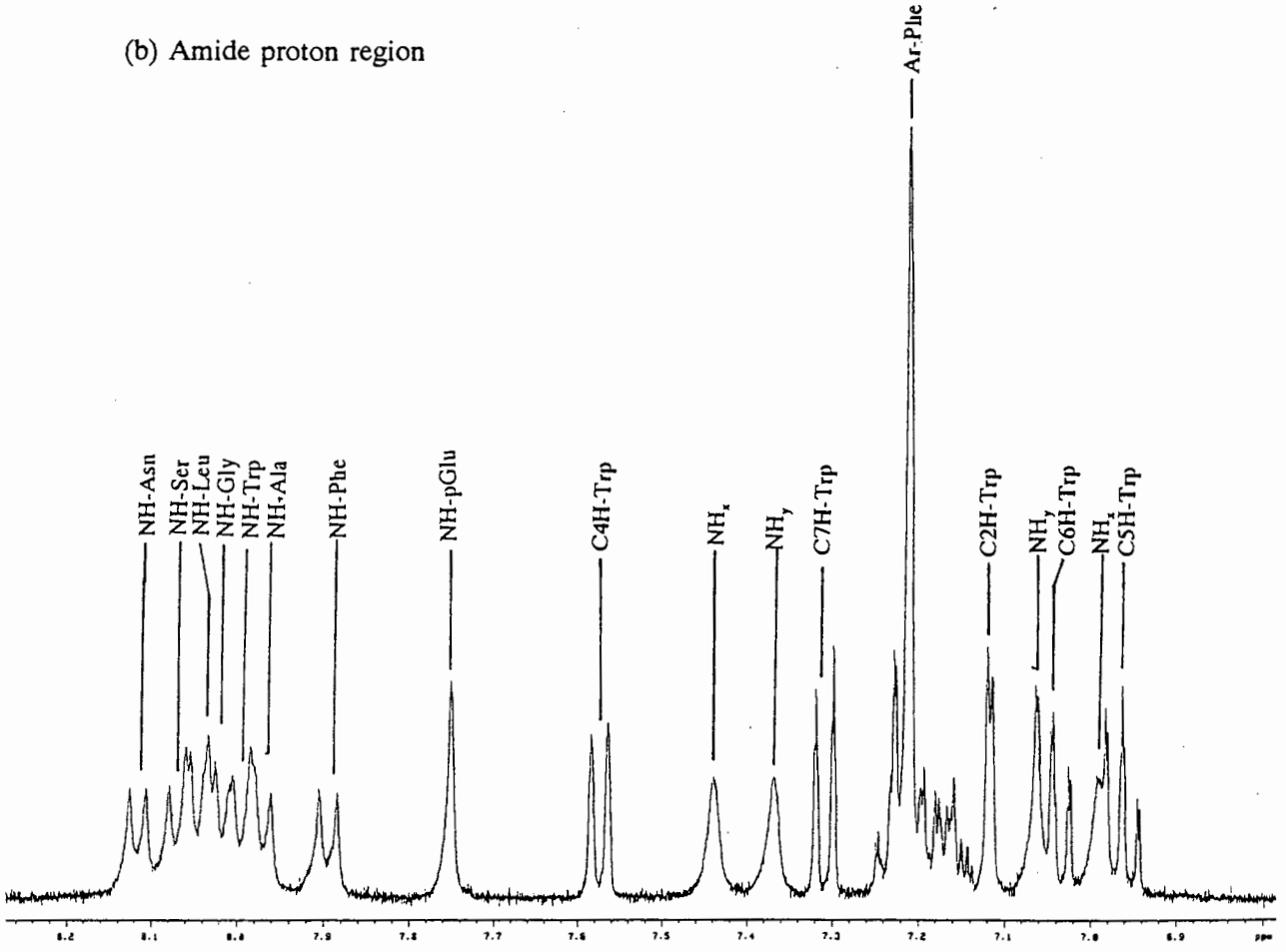
Problems encountered when differentiating between the very similar chemical shift values of the Asn and Trp α-protons were solved when a trace through the β region of the TOCSY experiment revealed a more upfield shift of the Trp α-proton. The same approach was required to distinguish the overlapping Gly α- and Ser β-protons from each other (Figure 5.5). The NOESY spectrum contained crosspeaks in the fingerprint region, which enabled the two Ser β-protons to be successfully differentiated (Figure 5.6). Neither the TOCSY or COSY spectra showed sufficient resolution to differentiate between the two Leu β protons. The clear coupling between the Trp phenyl ring protons, observed in both the Homonuclear 2DJ-Resolved and TOCSY experiments, permitted accurate assignments which were in agreement with both the Lom-AKH-I and literature values (Figure 5.7 (a) and (b)). All chemical shift values are listed above in Table 5.1; see also Figure 5.8.

Figure 5.8: ^{1}H spectrum recorded for Lom-AKH-II in d^6-DMSO.

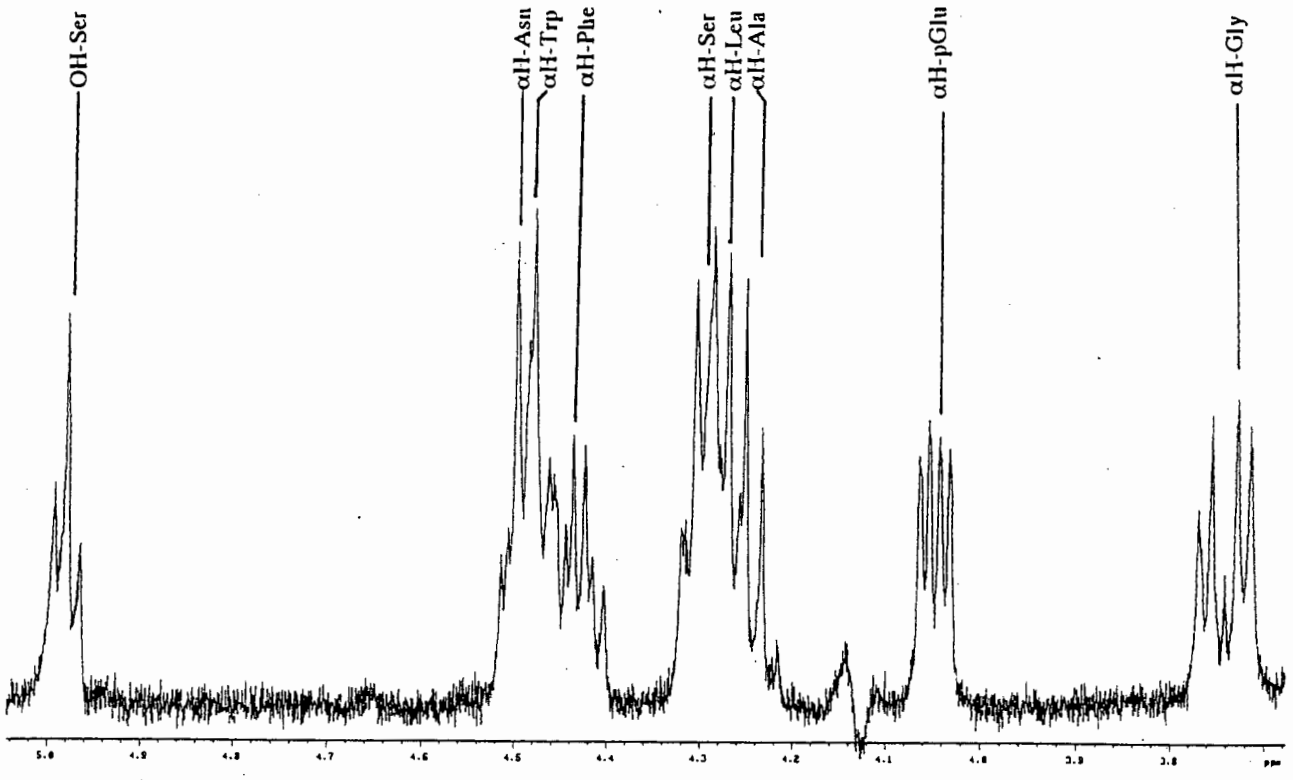
(a) Full spectrum



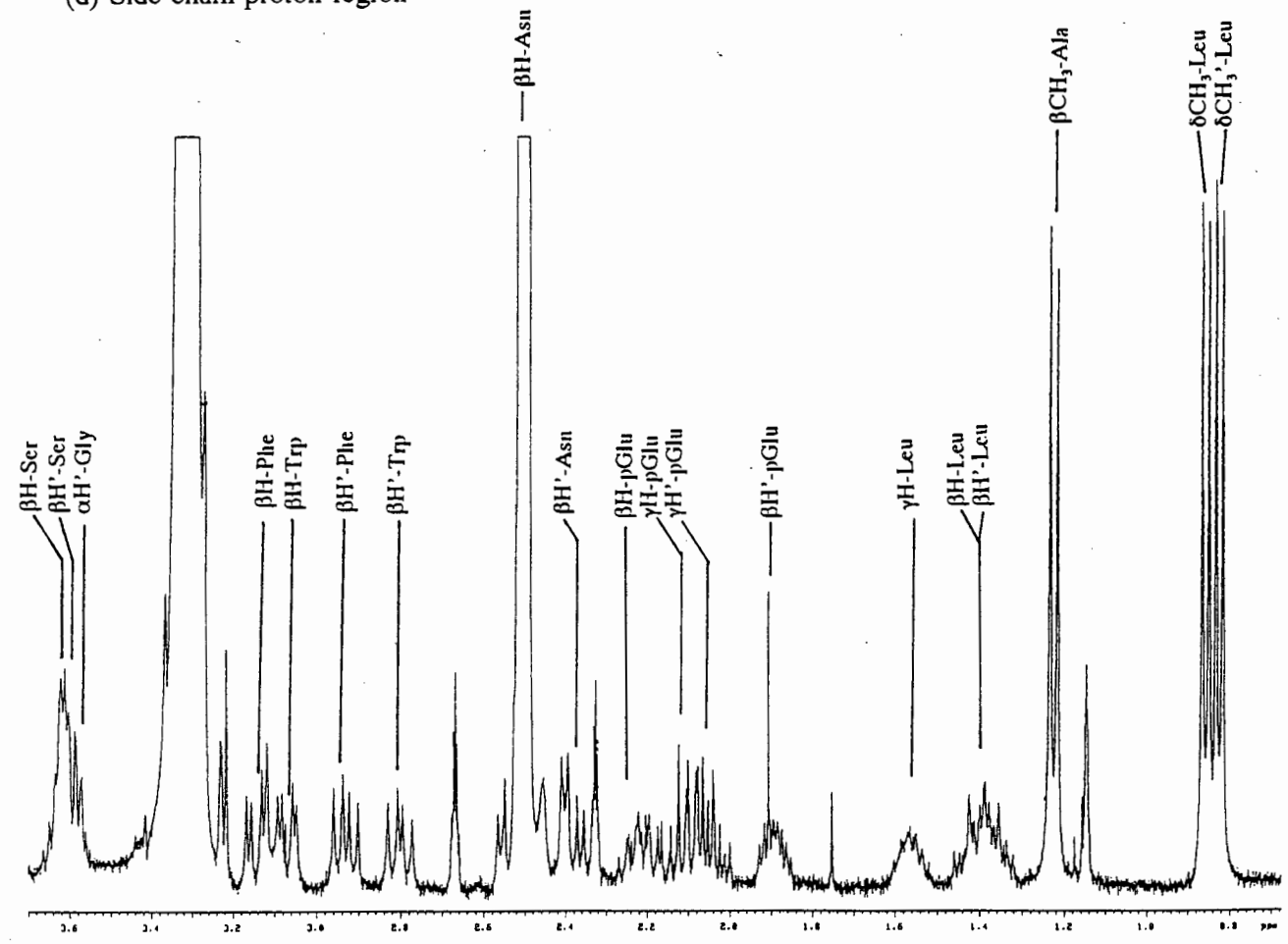
(b) Amide proton region



(c) C_α-proton region



(d) Side chain proton region



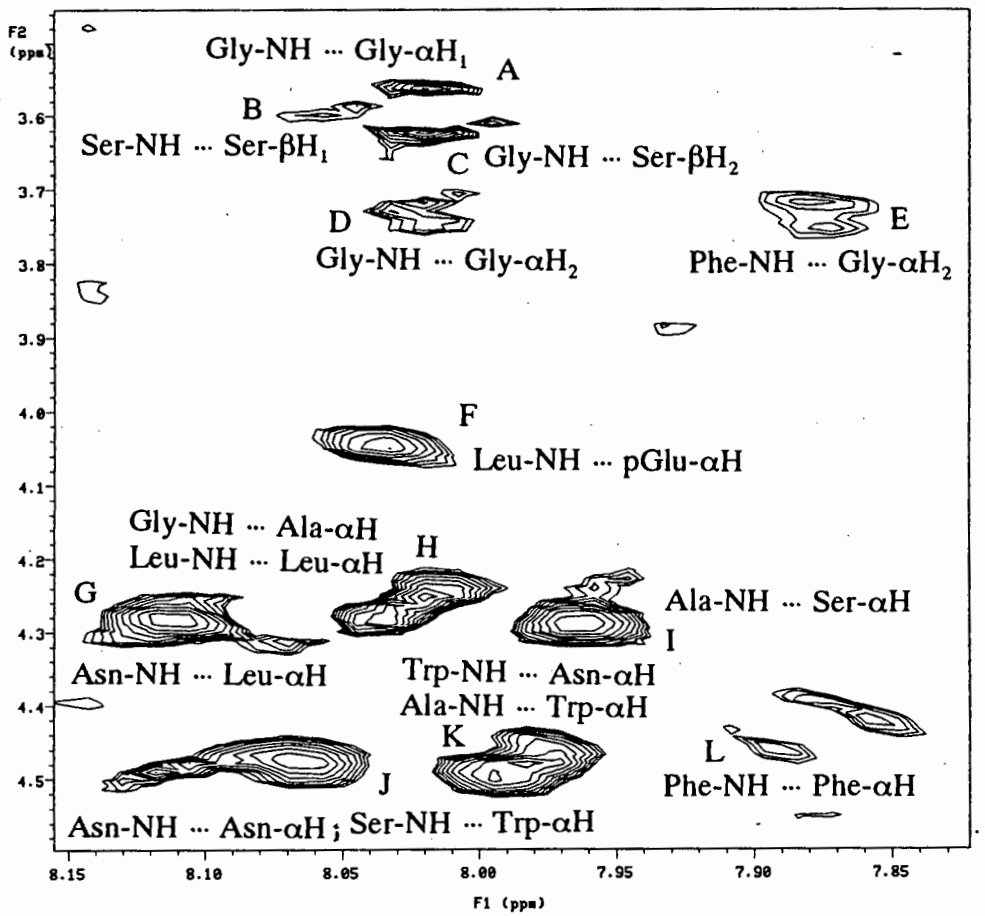


Figure 5.6: Fingerprint region of the NOESY spectrum recorded for Lom-AKH-II in d^6-DMSO.

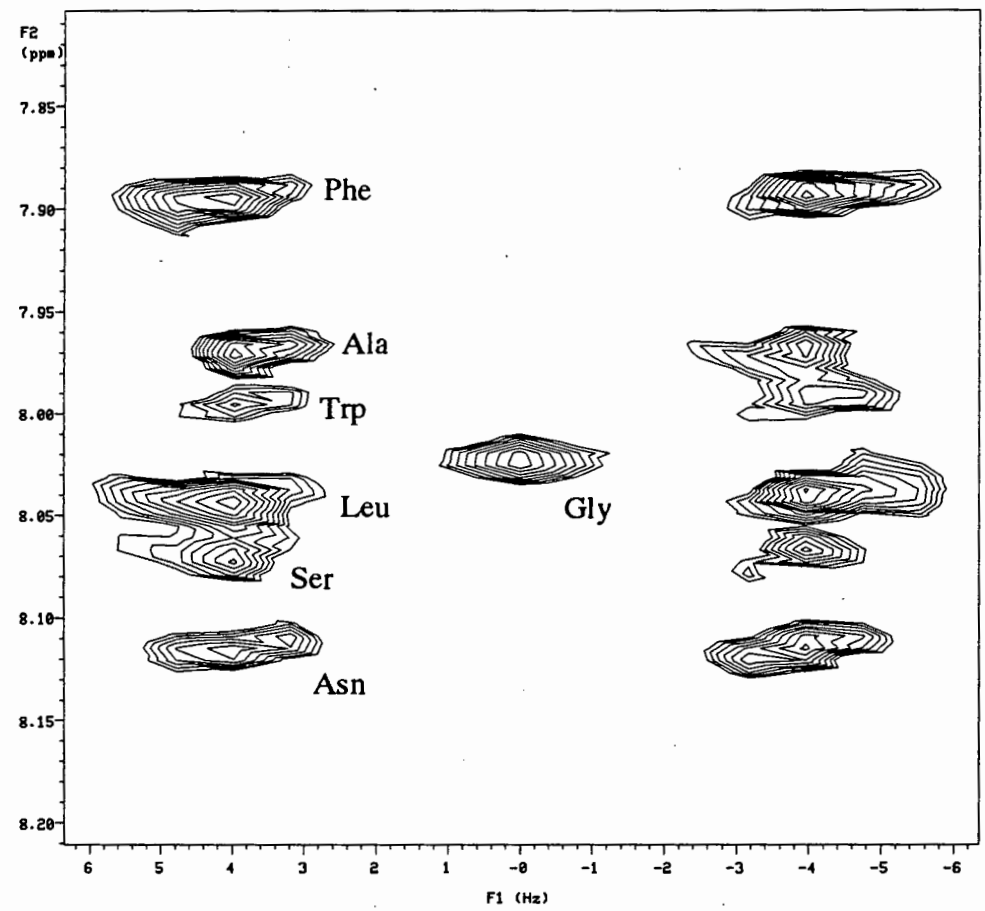


Figure 5.9: Amide region of the Homonuclear 2DJ-Resolved spectrum recorded for Lom-AKH-II in d^6-DMSO.

With regards to the two primary amide groups - although the couplings were clearly visible in the COSY and TOCSY spectra and accurate chemical shift assignment of the proton pairs was possible (Figure 5.7(a)), subsequent assignment, using the NOESY spectra, of either group to the terminus or Asn residue proved to be impossible.

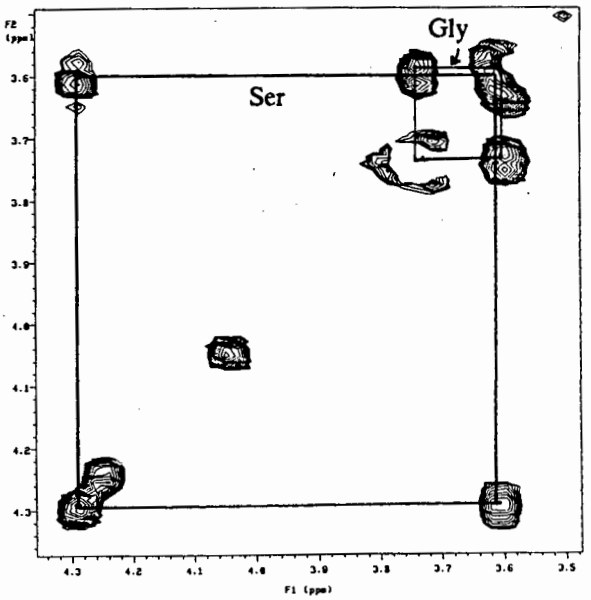


Figure 5.5: TOCSY spectrum showing the close chemical shift values of the Ser β- and Gly α-protons.

Extensive overlap in the amide region of the proton spectrum meant that coupling constant calculations were again largely dependent on the Homonuclear 2DJ-Resolved experiment, which showed satisfactory resolution in this region (Figure 5.9). The only exceptions involved the Phe and Gly residues, where higher resolution in the proton spectrum meant that for this residue the proton, and not the Homonuclear 2DJ-Resolved, spectrum was used. In the case of Gly, coupling constants

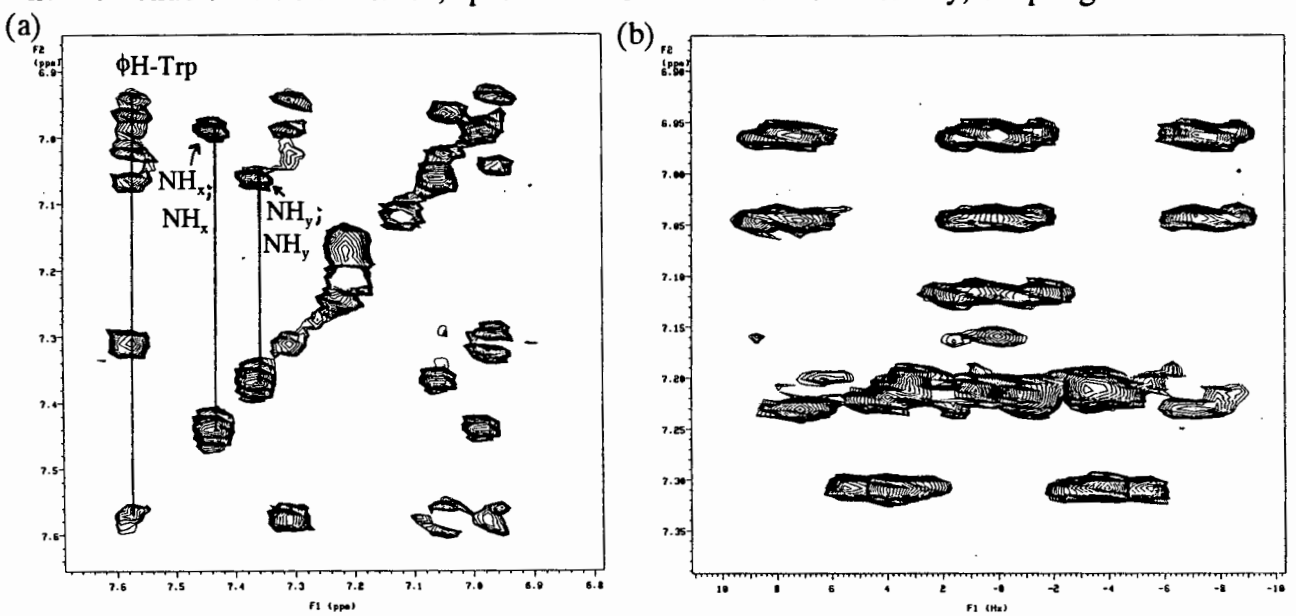


Figure 5.7: (a) TOCSY spectrum showing the relay scalar coupling between the Trp ring protons. (b) Homo. 2DJ-Resolved spectrum indicating the same.

were calculated from the AB coupling patterns generated by the α-protons. The Karplus

equation was subsequently used to calculate the corresponding torsion angles (Table 5.2).

NOESY spectra were recorded at mixing times of 150, 200 and 250ms. A study of the resultant spectra showed the 200ms spectrum, which had a signal:noise ratio higher than that of the 150ms spectrum, to be the most resolved. The problem of increased occurrence of spin diffusion effects witnessed at these longer mixing times was to a certain extent alleviated in

Table 5.2: Coupling Constant and torsion angle values recorded from the proton and Homo. 2DJ-Resolved spectra for Lom-AKH-II in d^6-DMSO.

Residue	Coupling Constant, J, in Hz		Torsion Angle, ϕ
	Homo-2DJ	Proton	
Leu	7.9	-	-150.1
Asn	7.8	7.8	-151.3
Phe	8.0	8.2	-147.2
Ser	7.9	7.9	-150.1
Ala	7.9	-	-150.1
Gly	-	5.6	88.4/-168.6
Trp	7.9	-	-150.1

subsequent calculations through the use of MORASS. Due to the absence of certain critical peaks from the fingerprint region of the 250ms NOESY spectrum, only the integration values for the 200ms spectrum have been given below in Table 5.3. Again the ISPA was not applied due to a lack of a suitable reference peak, and distance constraints were generated through data manipulation in conjunction with results generated by MORASS.

Assignment of the NOESY peaks was a fairly straightforward process. Although accurate integral values for overlapping peaks could not be reported, the interactions contributing to any one given peak were conclusive (Figure 5.6). The initial dynamics runs performed to generate a reasonable distant constraints set corresponding to an acceptable *R*-factor proved

Table 5.3: NOESY (200ms) integration values recorded for Lom-AKH-II.

Peak	Integrated Volumes recorded from the 200ms NOESY spectrum.
A	0.062
B	0.048
C	0.075
D	0.100
E	0.141
F	0.168
G	0.333
H	0.322
I	0.285
J	0.391
K	0.402
L	0.040

very problematic. It soon became obvious that no single structure could possibly obey all the imposed constraints. This was due to the fact that there existed three different long range interactions, which were impossible to co-exist in any one given structure. These interactions, shown in Figure 5.10, included a Trp_8-NH ··· Asn_3-αH, an Ala_6-NH ··· Trp_8-αH and a Ser_5-NH ··· Trp_8-αH interaction. The assumption was therefore made that the molecule was in fact fluctional, and that a minimum of three different structures must therefore have co-existed at any one given time in solution. The above three interactions were thus split up and three different structures (A, B and C respectively) were generated, which when combined in a 1:1:1 ratio generated an *R*-factor of less than 30%. All other interactions were assumed to make equal contributions to each of the three structures, i.e. the constraints corresponding to these interactions were common to all three run types, and the theoretical nOes generated for each were averaged. Two of the MD runs were performed at a temperature of 900K to ensure maximal coverage of conformational space. The third was run at a lower temperature of 700K to reduce the occurrence of distance violations at the higher temperature.

To generate these three structures it was necessary to formulate a theoretical NOESY data set corresponding to an average of the three example structures. This was done as follows:

$$\textit{Overall theor. peak intensity, } V = \frac{theor.(V)_A}{3} + \frac{theor.(V)_B}{3} + \frac{theor.(V)_C}{3} \qquad \textbf{(5.1)}$$

where A, B and C refer, respectively, to the three co-existing structures. The product

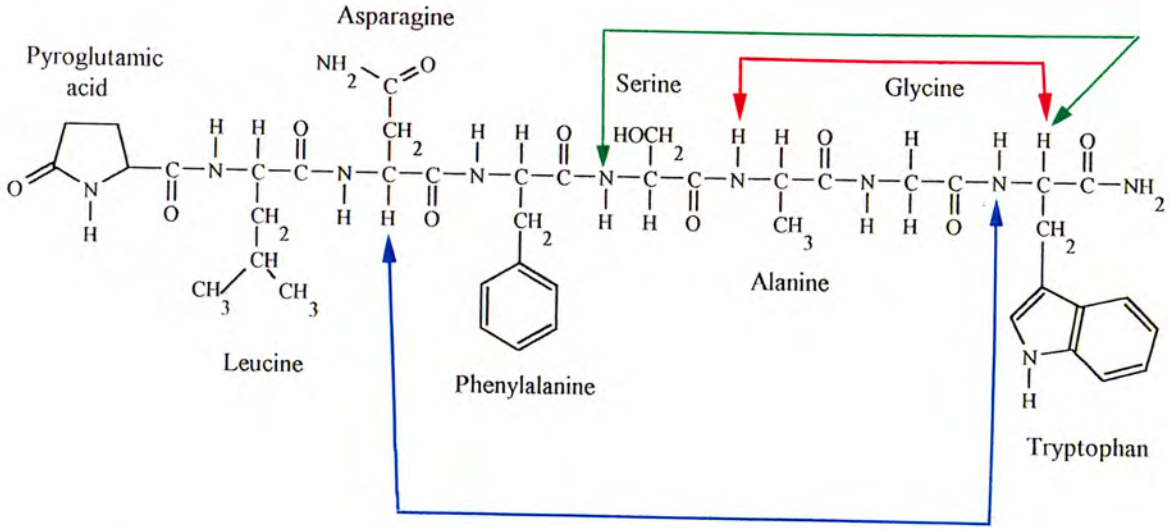


Figure 5.10: The three long range interactions which required a minimum of three structures to be satisfied.

theoretical data set was scaled down to the experimental data as previously described for Lom-AKH-I, and the *R*-factor calculated.

The choice of structure from each of the key MD runs was simply restricted to structures possessing no distance violations, and the sample structures A, B and C, were thus representative of the whole run, i.e. there was no sub-division of structures into families due to the impracticalities which would have arisen in the extensive regrouping of the three different structural types. The ensuing discussion will thus be based on structures representative of the three different structural types and not on variations within each of these individual groups.

A complete theoretical NOESY breakdown for all three structures is given in Table 5.4 along with the overall scaled theoretical data set. The one interaction which could never be satisfied was that between the Phe-NH ··· Gly-αH_1. A study of the trace taken through this region showed this peak to be very broad and therefore probably integrating to far more than the true value, i.e. the protons in reality were much further apart. Depending on the processing (reduced use of the shifted sine bell), this peak would disappear below the noise levels (background noise = 0.002 ± 0.028). The fact that no theoretical nOe was predicted

Table IV: Theoretical NOESY results recorded for structures A, B and C.

Peak	Interaction	Structure A		Structure B		Structure C		Exp.	Theor. (scaled)	Diff.
		Dist.	Theor. (unscaled)	Dist.	Theor. (unscaled)	Dist.	Theor. (unscaled)			
A	Gly-NH ··· Gly-αH_1	3.17	0.3800	2.91	0.581	3.05	0.443	0.062	0.105	-0.043
B	Ser-NH ··· Ser-βH_1	2.90	0.520	3.30	0.354	3.33	0.421	0.048	0.097	-0.049
C	Gly-NH ··· Ser-βH_2	-	-	3.17	0.382	3.56	0.382	0.075	0.058	0.017
D	Gly-NH ··· Gly-αH_2	3.02	0.428	3.16	0.489	3.06	0.441	0.100	0.101	-0.001
E	Phe-NH ··· Gly-αH_2	-	-	-	-	-	-	0.141	-	0.141
F	Leu-NH ··· pGlu-αH	2.82	0.918	2.83	0.938	2.82	0.911	0.168	0.207	-0.039
G	Asn-NH ··· Leu-αH	2.41	1.578	2.47	1.786	2.45	1.589	0.333	0.370	-0.037
H	Gly-NH ··· Ala-αH	3.46	0.329	2.91	0.751	3.11	0.446	0.138	0.114	0.024
	Leu-NH ··· Leu-αH	2.77	0.792	2.71	1.047	2.72	0.974	0.184	0.210	-0.026
I	Ala-NH ··· Ser-αH	2.76	0.987	2.46	1.692	2.50	1.662	0.285	0.324	-0.039
J	Asn-NH ··· Asn-αH	3.09	0.425	3.02	0.585	3.09	0.462	0.096	0.110	-0.014
	Ser-NH ··· Trp-αH	-	-	-	-	2.27	2.660	0.296	0.200	0.096
K	Ala-NH ··· Trp-αH	-	-	2.32	2.222	-	0.031	0.201	0.169	0.032
	Trp-NH ··· Asn-αH	2.25	2.900	-	-	-	-	0.201	0.217	-0.016
L	Phe-NH ··· Phe-αH	3.24	0.355	3.23	0.412	3.23	0.380	0.040	0.086	-0.046

for this peak was therefore not assumed to be a major NOESY violation. Additionally, structures A and C showed two and structure B three other NOESY violations. The overall agreement with the experimental data was demonstrated by the calculated *R*-factor values of 0.262 and 0.080 for R_1 and R_2 respectively. No hydrogen bonding was observed in any of the three sample structures - this would be expected in the light of Lom-AKH-II being a highly flexible peptide.

The three structures are shown below in Figure 5.11. All three of these structures contain some form of turn region. In Lom-AKH-I and III, the presence of a proline residue automatically provided a degree of stability to the turn region even in the absence of any hydrogen bonding. Lom-AKH-II, lacking a proline, therefore contains no inherent stability to a turn region, and hence, in the absence of hydrogen bonds, will be highly flexible. Looking at the three structures, it would appear that the turn could possibly be involved in continuous 180° flipping, i.e. structure A shows a key Asn_3 - Trp interaction which induces a prominent turn around the Phe, Ser and Ala residues. Structure C, however, shows the amide terminus to have flipped completely around (effectively generating a "mirror image" of structure A) in order to satisfy the Ser-Trp interaction which also generates a substantial turn, this time involving the Ser, Ala and Gly residues. Structure B shows an intermediate between the two extreme states. The turn here is formed around the last three residues in the sequence (the key interaction is one involving the Trp and Ala residues), and is therefore very shallow relative to the previous two described above. The direction of this turn is the same as described for structure A. It was therefore proposed that the major flexibility of this peptide involved continuous backwards and forwards flipping of the amide region of the peptide. The pGlu and Leu residues were excluded from any long range interactions and therefore had random flexibility restricted only by the associated torsion angles - an average of their motion was being observed (Figure 5.12).

The torsion angle results for the three structures are given below in Table 5.5. Structures A and B showed significantly better overall agreement with the experimental data than structure C. (This included agreement with the reported Gly torsion angle, which, unlike previously in the case of Lom-AKH-I, was not constrained in the MD runs.)

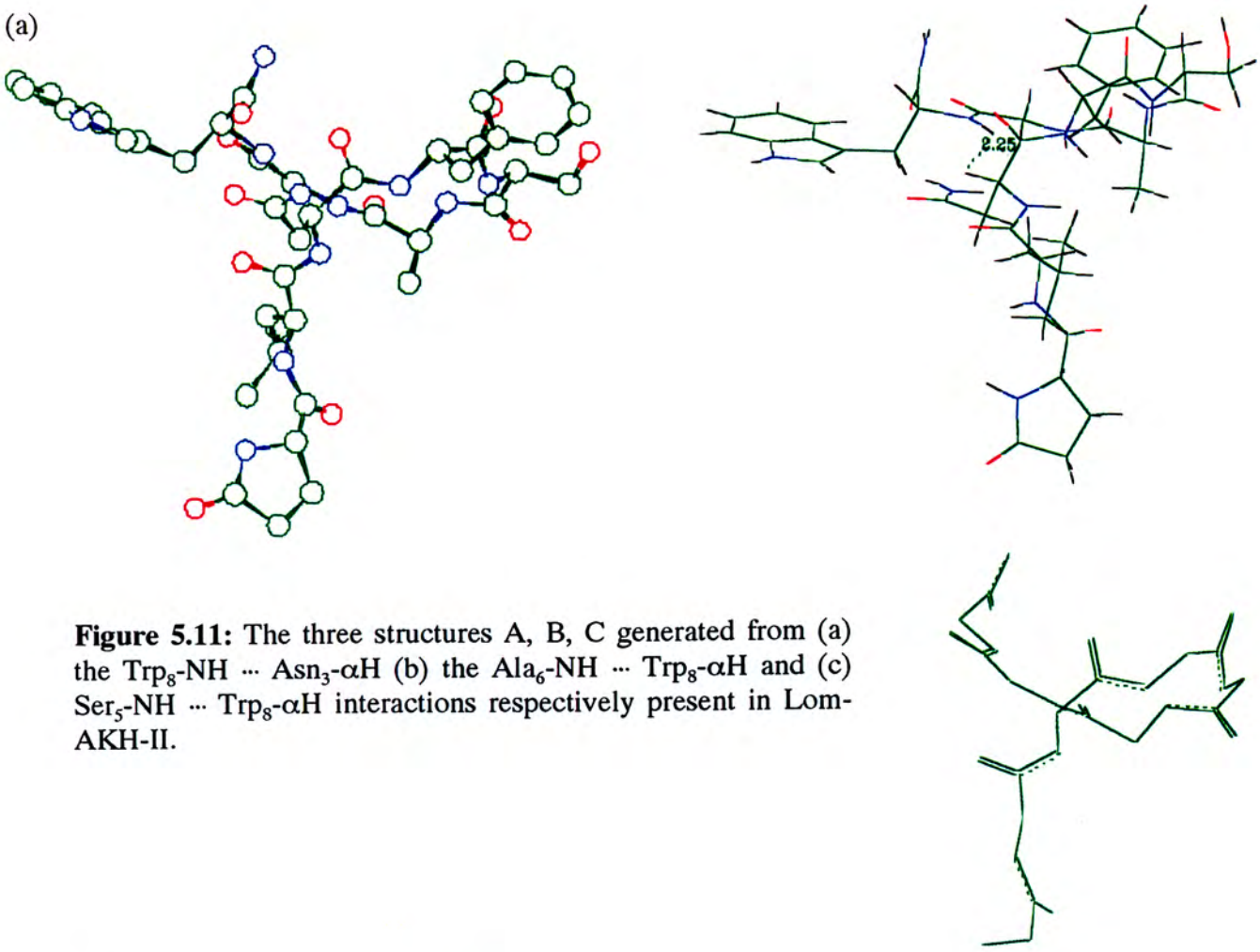


Figure 5.11: The three structures A, B, C generated from (a) the Trp_8-NH ··· Asn_3-αH (b) the Ala_6-NH ··· Trp_8-αH and (c) Ser_5-NH ··· Trp_8-αH interactions respectively present in Lom-AKH-II.

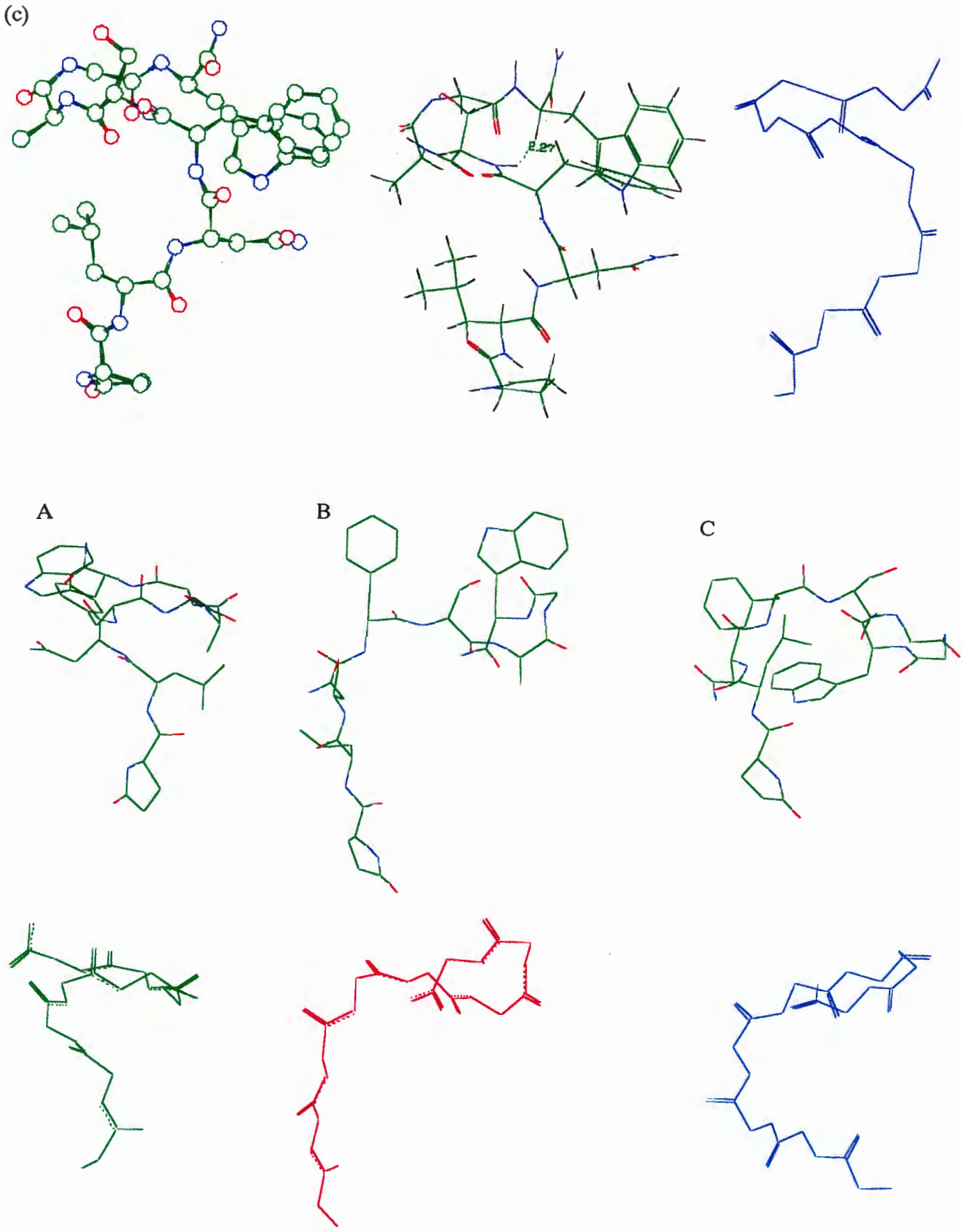


Figure 5.12: Illustration of the 180° flipping of the amide terminus. Structures A (Trp at the back) and C (Trp in the front) represent the two extreme states ("mirror images"), while structure B (Trp at the back) represents the intermediate "shallow" turn state.

Table 5.5: A comparison between the theoretical and experimental torsion angle values.

Residue	Torsion Angle, ϕ			
	Structure A	Structure B	Structure C	Exp.
Leu	-149.6	-151.1	-151.0	-150.1
Asn	-151.8	-151.0	-150.3	-151.3
Phe	-142.7	-146.1	-145.0	-147.2
Ser	-143.2	-148.4	-145.8	-150.1
Ala	-150.3	-141.8	-137.8	-150.1
Gly	-160.7	77.0	-111.1	88.4/-168.6
Trp	-147.5	-148.7	-152.2	-150.1

The energies of the three structures A, B and C were 270.45, 311.85 and 312.94 respectively. Although these energies were distinct, the structures were still averaged in a 1:1:1 ratio. Ideally, it would have been more accurate to calculate the probability of the three different structures from their partition functions. However, such calculations are beyond the scope of currently available computer software.

Averaging was considered acceptable on the grounds of the above study being the simplest model for a highly complex problem. All attempts to proceed as before with Lom-AKH-I were rejected in the initial stages of the calculations. It became obvious that there was no solution set that would generate reasonable agreement between the experimental and theoretical data. A careful review of the long range interactions present in the NOESY revealed there to be three interactions which could not be satisfied in one or even two structures. A minimum of three were essential for the generation of a reasonable *R*-factor.

In conclusion, therefore, Lom-AKH-II has been shown to be highly fluctional - there was no possibility of generating a single structure which showed satisfactory agreement with all the experimental data. For this reason the activity of this peptide would be predicted to be relatively low. This has been reported to be the case with regards to certain, but not all, processes within the insect [5] - see later. It should be mentioned here that the behaviour of

this molecule could be drastically affected by the presence of a suitable substrate, i.e. in such a situation the peptide could be more likely to form the appropriate conformation required for effective binding, thereby inducing a rapid response.

Preliminary results obtained with the alternative assignments for the Phe and Trp residues are given in Appendix 1.

REFERENCES

1. Carlsen, J.; Herman, W.S.; Christensen, M.; Jefferson, L. *Insect Biochem.* **1979**, *9*, 497-501.

2. Gäde, G. *Zool. Jb. Physiol.* **1981**, *85*, 266-267.

3. Siegert, K.J.; Morgan, P.J.; Mordue, W. *Biol. Chem. Hoppe-Seyler* **1985**, *366*, 723-727.

4. Gäde, G.; Schaffer, M.H.; Cook, J.C.; Rinehart, K.L. *Biochem. Biophys. Res. Commun.* **1986**, *134*, 723-730.

5. Goldsworthy, G.J.; Mallison, K.; Wheeler, C.H. *J. Insect Physiol.* **1986**, *32(1)*, 92-101.

CHAPTER 6

LOM-AKH-III: RESULTS AND DISCUSSION

In 1991 a third peptide exhibiting adipokinetic activity was isolated from *Locusta migratoria* and sequenced first by Ouderjans *et al.* [1] and later by Heerma *et.al.*. Due to its complete absence in *Schistocerca gregaria* [2], it was added to the Lom peptide series as Lom-AKH-III. The primary sequence bore a strong resemblance to that of Lom-AKH-I (Figure 6.1) with the major differences being that it was an octa- and not a decapeptide, and that Asn was replaced by Trp at position seven in the sequence. It was the first of the adipokinetic hormones to have a Trp residue at position seven in the sequence [1]. This substitution accounted for the extremely hydrophobic nature of this peptide - five of the eight residues exhibited a high degree of hydrophobicity.

A 0.0070M solution of the peptide dissolved in d^6-DMSO was prepared. The temperature arrayed proton experiment showed maximum resolution at 298K, and thus all subsequent proton, COSY, TOCSY and NOESY (150, 200 and 250ms) experiments were carried out at this temperature. Assignment of the proton spectrum was carried out as previously described - being largely dependent on spin system identification in the 2D COSY (Figure 6.2) and TOCSY (Figure 6.3 and 6.4) experiments and was again facilitated by chemical shift assignments previously made in Lom-AKH-I and Lom-AKH-II.

The characteristic upfield shifts of the three methyl groups of the Leu and Thr residues meant that their respective A_3B_3MPTX and A_3MX spin systems were readily located in the COSY and TOCSY spectra. Extensive streaking in the TOCSY spectrum by the residual solvent peak made it difficult, in certain cases, to make accurate peak assignments. The Pro and pGlu residues were both completely assigned using the COSY (Figure 6.5) spectrum which showed good resolution of the corresponding sequential coupling patterns. Although the δ-protons of the Pro residue were completely masked in the TOCSY, this spectrum did show some

Lom-AKH-III: pGlu-Leu-Asn-Phe-Thr-Pro-Trp-Trp-NH2

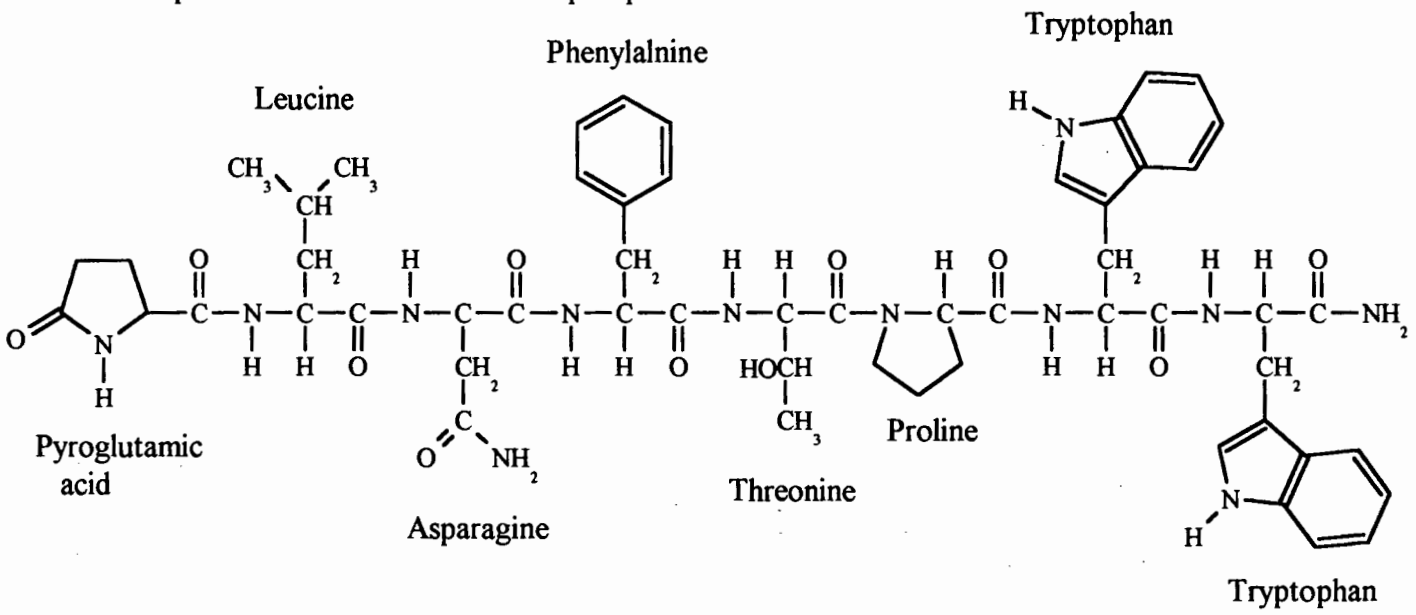


Figure 6.1: Primary sequence recorded for Lom-AKH-III.

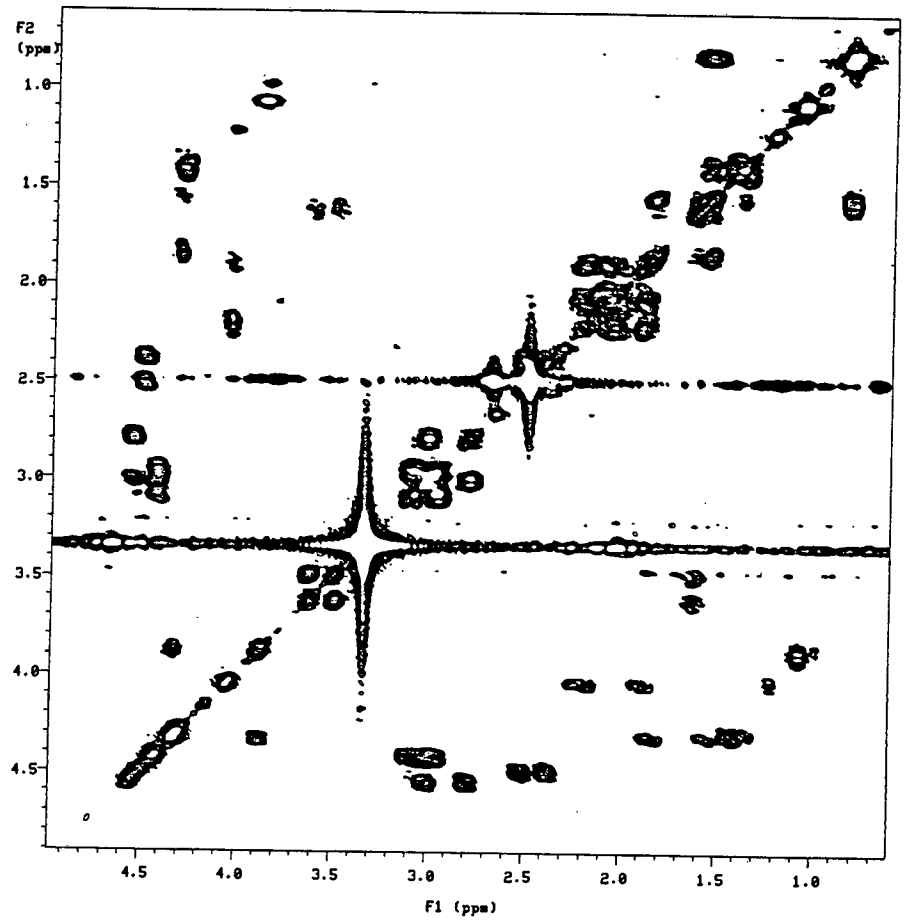


Figure 6.2: C_α-side chain region of the COSY spectrum recorded in d^6-DMSO for Lom-AKH-III.

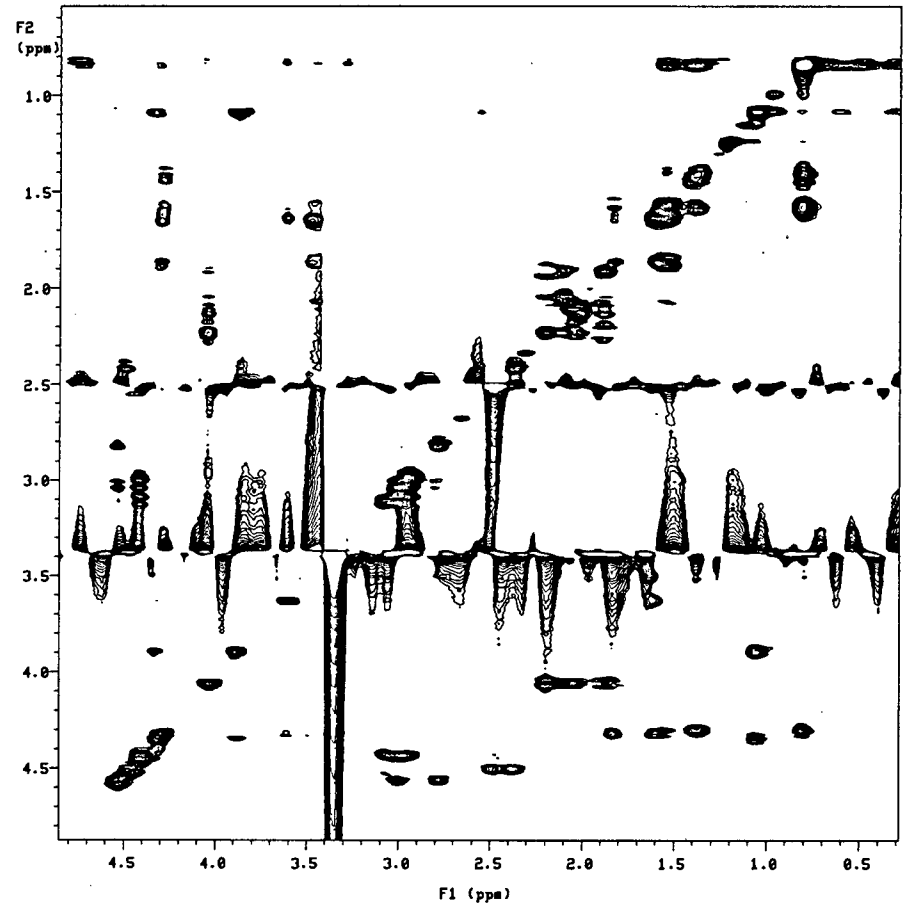


Figure 6.3: C_α-side chain region of the TOCSY spectrum recorded in d^6-DMSO for Lom-AKH-III.

coupling in the α-proton-sidechain region away from the diagonal. Although the TOCSY and COSY spectra (Figure 6.6) were cross referenced to make approximate assignments, the inferior resolution of the former meant that the actual chemical shift data was only obtained using the latter.

The remaining four Asn, Phe, and the two Trp residues, all having AMX systems, generated readily assignable coupling patterns in the COSY and TOCSY spectra; however, direct assignment of these systems to particular residues was impossible. The assumption that the Asn β-protons would resonate further upfield made it possible to distinguish this residue from the others. The corresponding assignments made were in good agreement with those recorded in Lom-AKH-I and Lom-AKH-II. Phenylalanine could be tentatively distinguished from the two Trp residues due to the similarity, with the exception of the amide protons, in chemical shift values of the latter two. These preliminary assignments were later confirmed by crosspeaks generated in the fingerprint region of the NOESY spectrum. These included both Thr-NH···Phe-αH and Phe-NH···Asn-αH interactions, which distinguished the Phe residue from the two Trp residues. A distinct Trp_7NH···Pro-αH interaction allowed differentiation of the two consecutive Trp residues. The proton spectrum integration revealed the absence of the Thr hydroxyl proton - most probably the result of very rapid exchange with the residual water peak. Based on the Trp ring proton assignments in Lom-AKH-I and II, it was possible to generate approximate assignments for the two Lom-AKH-III Trp rings. Unlike the previous two cases, where high

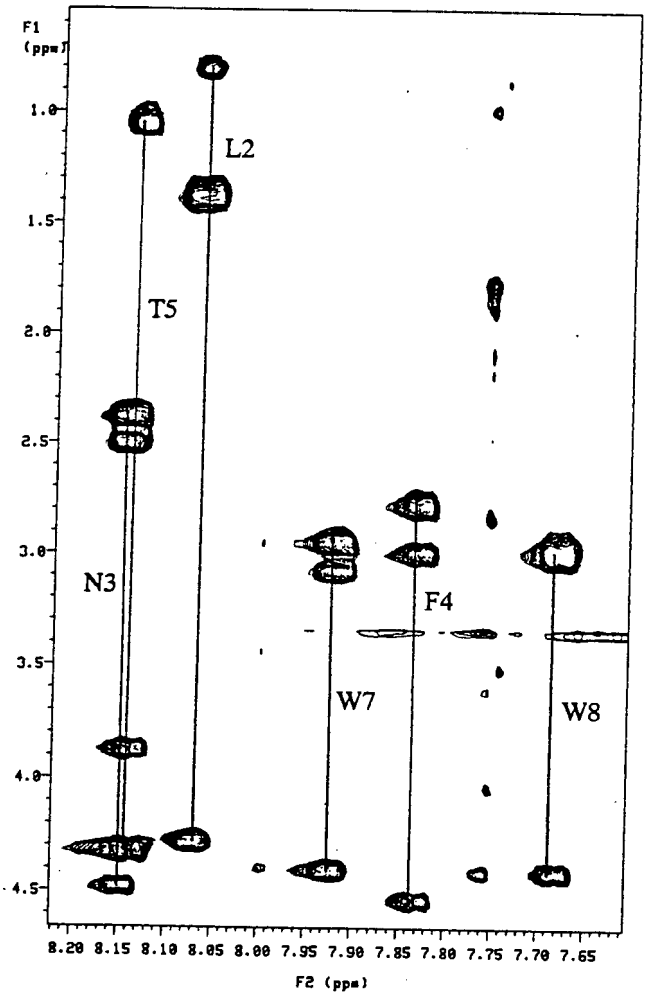


Figure 6.4: TOCSY spectrum (recorded with a mixing time of 80ms) showing the relay scalar connectivities between the C_α- and side chain protons of Lom-AKH-III.

resolution existed in the ring proton region of the Homonuclear 2DJ-Resolved spectrum, significant overlap in this region prevented accurate proton assignment. The final chemical shift values obtained for Lom-AKH-III are listed above in Table 6.1 - see also Figure 6.7. Although assignment of the two NH_2 groups to either the Asn residue or the terminus proved impossible, the scalar coupling within each proton pair was clearly visible in the TOCSY experiment. The fact that the two Trp ring amide protons showed no correlation in the TOCSY spectrum to any of the relevant β-, backbone amide or α-protons meant that no distinction could be made between these two protons.

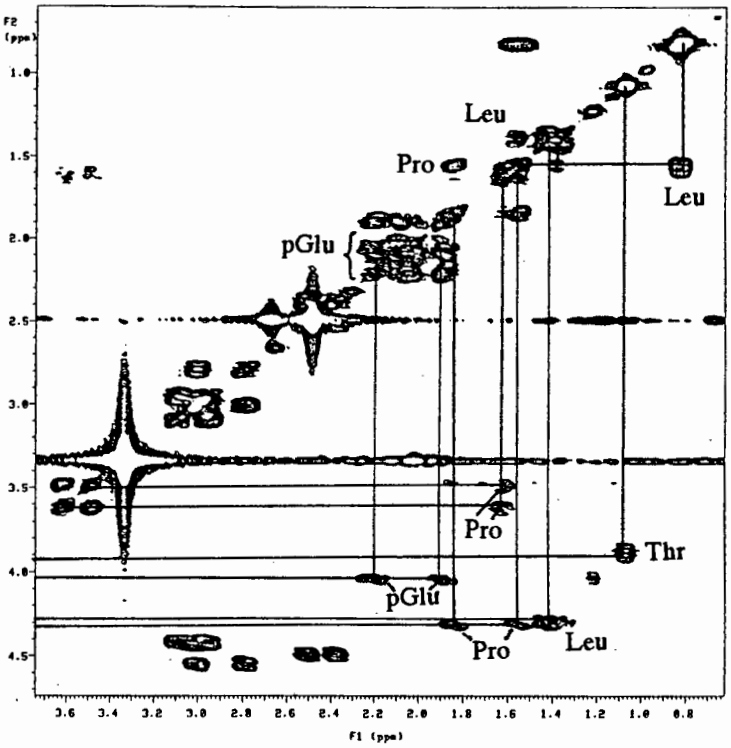


Figure 6.5: COSY spectrum showing the good resolution of the pGlu and Pro residues.

Coupling constant values for the previous two peptides were obtained almost exclusively from the Homonuclear 2DJ-Resolved experiment. However, in the case of Lom-AKH-III, the amide region of the proton experiment showed such superior resolution (particularly on application of the sine bell window functions) as to justify its sole use in torsion angle calculations. The coupling constant values recorded below in Table 6.2. show, in certain cases, the significant differences that existed between the two data sets. The amide region of the Homonuclear 2DJ-Resolved is given in Figure 6.8. The major problem with this

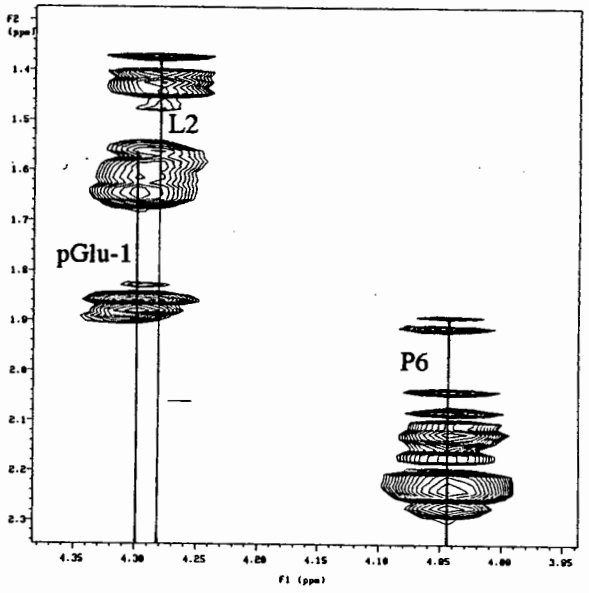


Figure 6.6: Sidechain region away from the diagonal illustrating inferior resolution of the pGlu and Pro residues in the TOCSY spectrum.

Table 6.1: Chemical shift values generated for Lom-AKH-III in d^6-DMSO.

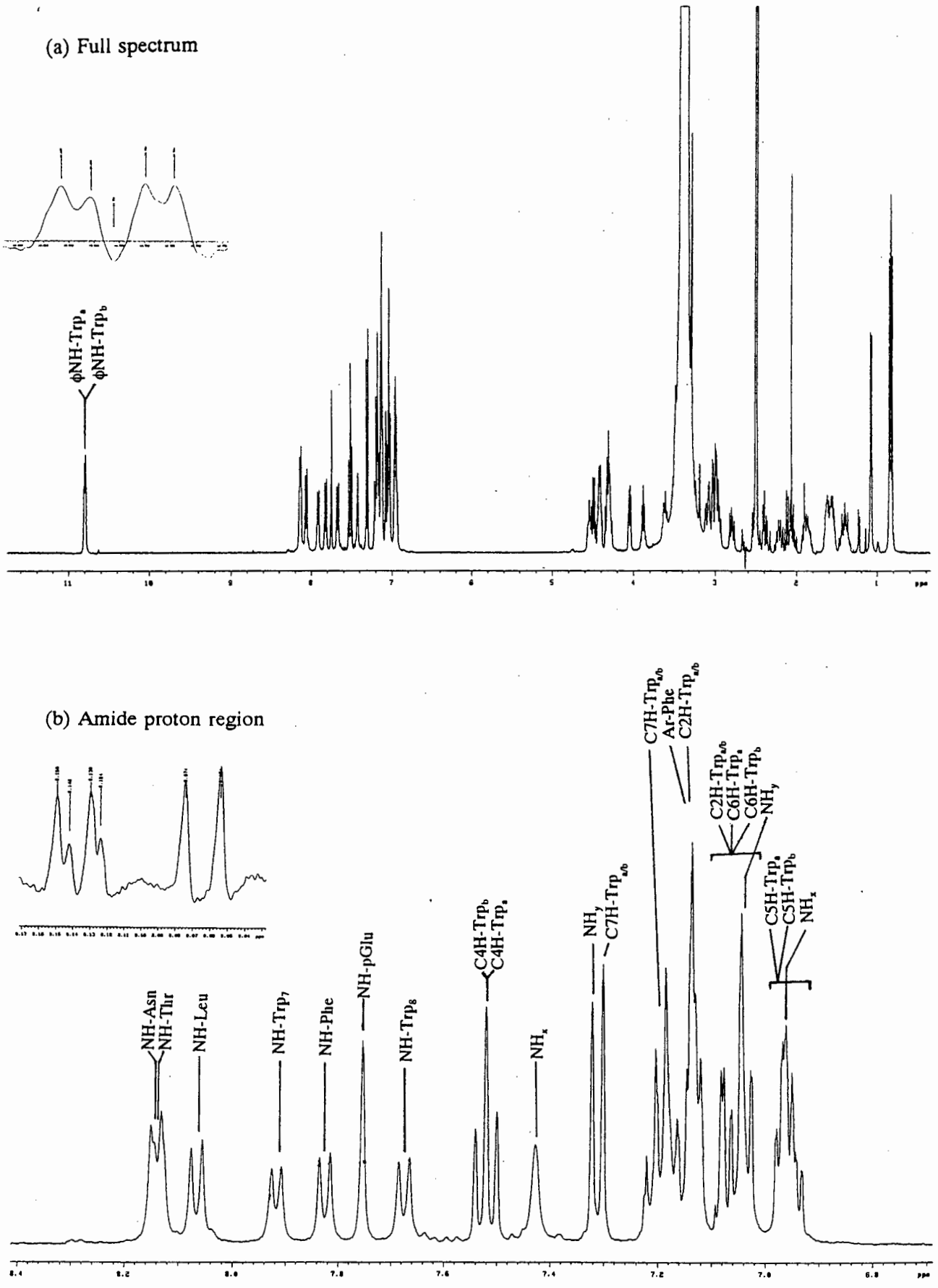
Residue	Amide	$C_\alpha H$	$C_\beta H$	$C_\beta H'$	Other
pGlu	7.753	4.041	2.211	1.885	2.139 γH 2.058 γH'
Leu	8.064	4.286	1.428	1.386	1.555 γH 0.843 δCH_3 0.821 δCH_3'
Asn	8.140	4.471	2.501	2.376	
Phe	7.824	4.536	3.001	2.795	7.115-7.172 ϕH's
Thr	8.130	4.322	3.882		1.078 γCH_3 - OH
Pro		4.300	1.842	1.550	1.639 γH's 3.620 δH 3.510 δH'
Trp_7	7.916	4.410	3.092	2.964	7.500-7.540 C4H 6.925-6.985 C5H 7.020-7.092 C6H
Trp_8	7.674	4.425	3.038	2.974	7.500-7.540 C4H 6.925-6.985 C5H 7.020-7.092 C6H

Although the two amide groups at (7.428;6.970) and (7.040;7.311) were assignable, they were not distinguishable. Other indistinguishable protons included the two ϕNH-Trp protons at 10.798 and 10.792, and the two C2H and C7H of the Trp ring. Of the former, the one occurred between 7.020 and 7.092 and the other between 7.113 and 7.160. Regarding the latter, the one occurred between 7.160 and 7.231 and the other at 7.311.

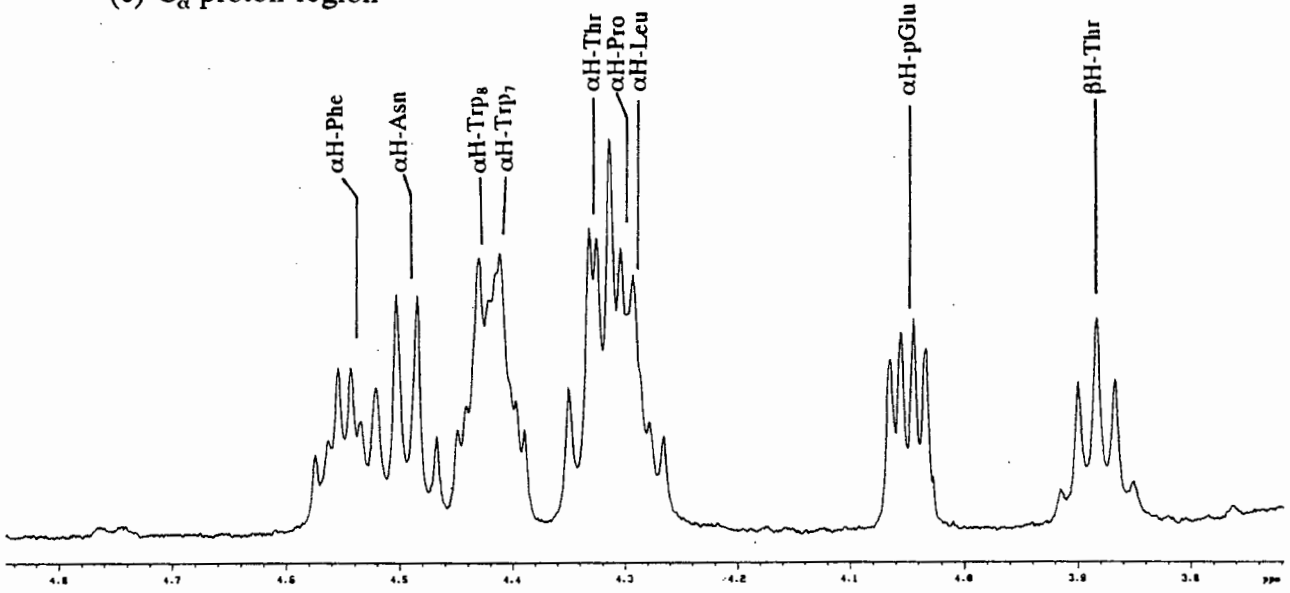
spectrum was that the peaks became disjointed on rotation due to the fact that the digital resolution in the F1 and F2 dimensions was not the same. The Karplus equation was subsequently applied to the relevant data for generation of ϕ torsion angle values.

NOESY spectra were run at mixing times of 150, 200 and 250ms. These were processed and assessed in terms of viability for use in the ensuing molecular mechanics calculations for structure generation. The 150ms spectrum had the advantage of lower spin diffusion effects

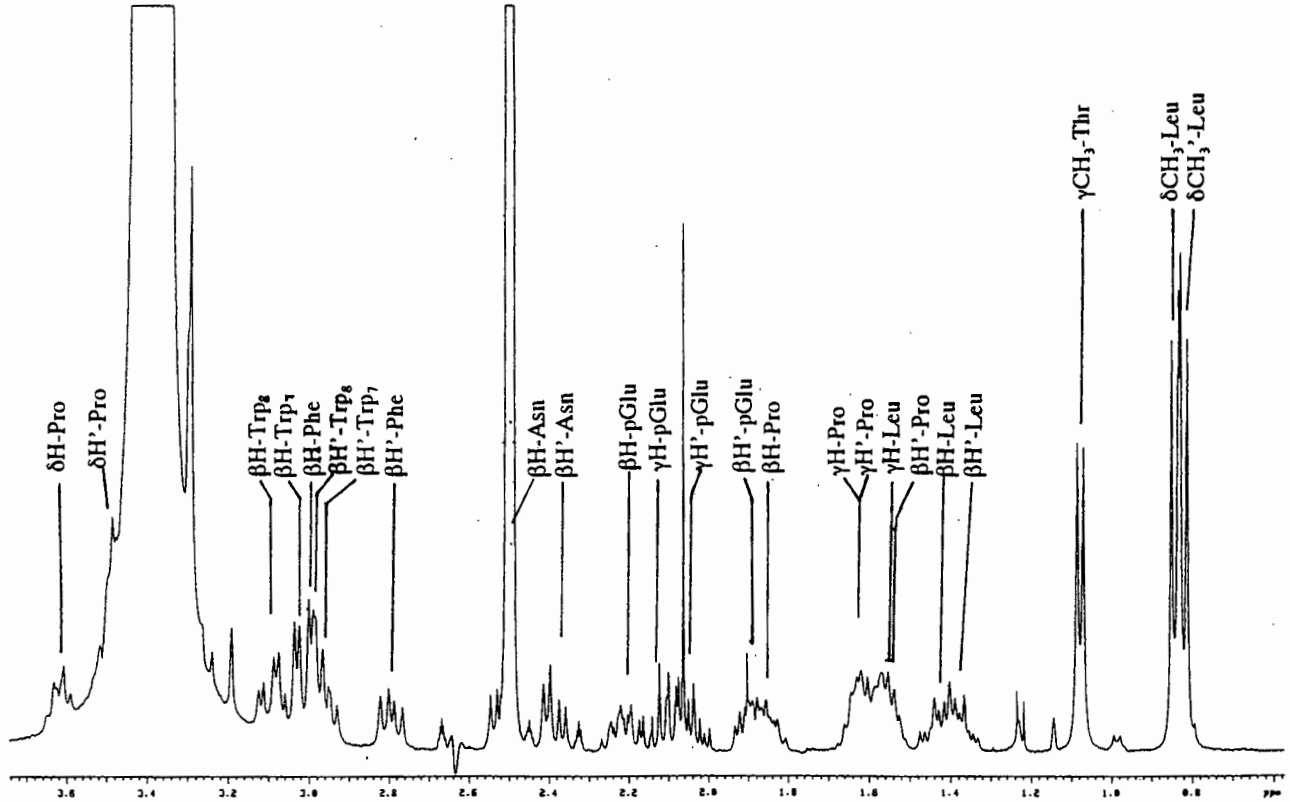
Figure 6.7: ^{1}H spectrum recorded for Lom-AKH-III in d^6-DMSO.



(c) C_α-proton region



(d) Side chain proton region



within the spectrum, however a study of the fingerprint region showed high noise levels, thus preventing accurate peak integration. Conversely, both the 200 and 250ms NOESY spectra had higher signal:noise ratios and displayed significantly higher resolution in the region of interest. Because lower spin diffusion effects would be experienced in the experiment of shorter mixing time, the 200ms spectrum was used for all subsequent molecular mechanics calculations. Both the integration (Table 6.3) and peak assignment processes were facilitated by the fact the NOESY spectrum had relatively little overlap in the fingerprint region (Figure 6.9).

The presence of the Pro residue immediately opened the question of *trans-cis* isomerism around the Pro imide bond. The proton spectrum showed only one set of NH resonances, and from the nOe's between the Thr-αH and the two Pro-δH's the imide bond could be assigned to the *trans* rotomer.

The diffuse nature of the water peak prevented accurate integration of a suitable reference peak, thus making application of the ISPA impossible. This prevented distance calculations for generation of a crude distance constraints set and made distance constraint manipulation once again a necessity.

Table 6.2: Coupling constant and torsion angle values recorded for Lom-AKH-III.

Residue	Coupling Constant		Torsion Angle, ϕ
	Proton	Homo-2DJ	
Leu	8.1	8.2	-148.3
Asn	7.8	-	-150.9
Phe	8.1	8.1	-148.2
Thr	7.4	-	-154.5
Trp_7	7.4	7.7	-154.5
Trp_8	8.1	8.1	-148.3

The experimental data collected was used as input into Biosym's Discover module for structure generation. The process of simulated annealing was followed more closely in the case of Lom-AKH-III than in the case of either of the two previous peptides. The absence of accurate distance constraints meant that before carrying out

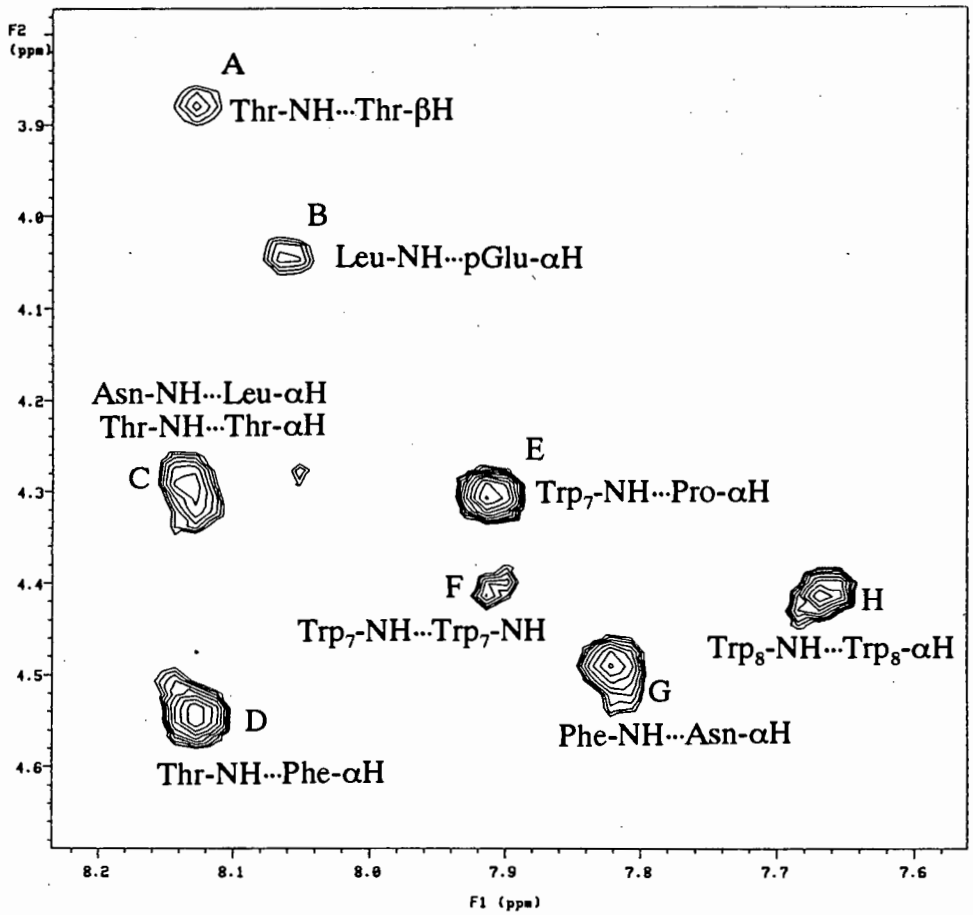


Figure 6.8: Fingerprint region of the NOESY spectrum recorded for Lom-AKH-III in d^6-DMSO.

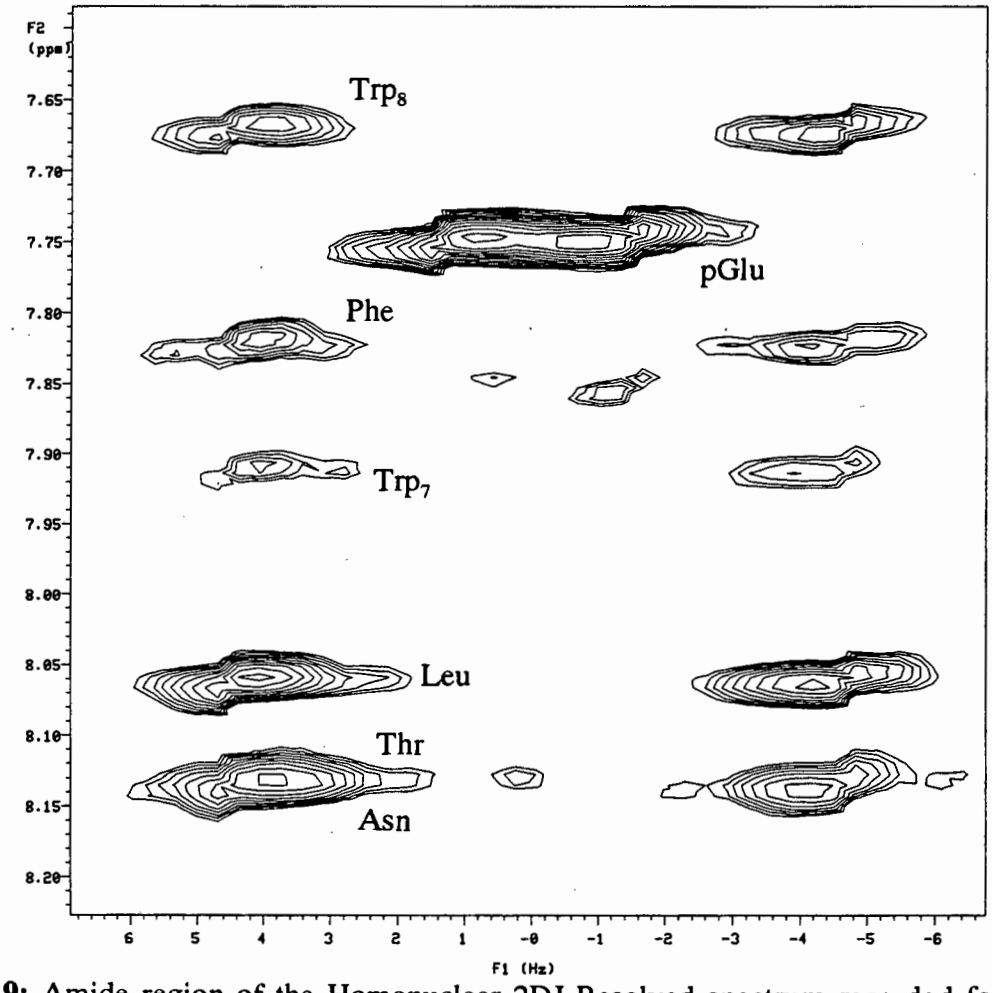


Figure 6.9: Amide region of the Homonuclear 2DJ-Resolved spectrum recorded for Lom-AKH-III in d^6-DMSO.

a full conformational space search, it was necessary to generate a set of constraints which were in satisfactory agreement with the experimental data i.e. until structural analysis by MORASS generated an *R*-factor of less than 30%. This was done by executing a series of short dynamics runs, where distances between the relevant proton pairs (those which generate peaks in the NOESY spectrum fingerprint region) were repeatedly manipulated until a feasible set of distance constraints was generated, which was then applied to all subsequent dynamics runs.

To ensure maximal coverage of conformational space, initial runs were performed at a temperature of 900K. This temperature proved ineffective in terms of producing acceptable structures. The high energy acquired by the molecule at this elevated temperature allowed the occurrence of absurdly short non-bonded distances. For this reason the temperature was lowered down to 600K, where the occurrence of such high energy interactions was significantly reduced. The one hundred structures generated were grouped into families of structures with a *RMS* deviation in aligned positions of backbone atoms of less than 0.3. These calculations required the generation of cluster and corresponding energy graphs (Figure 6.10(a)). The 20 families generated were studied, and any family in which all structures

Table 6.3: Integrated volumes recorded for Lom-AKH-III at different NOESY mixing times.

Peak	Integrated Volumes recorded for different mixing times (ms)		
	150	200	250
A	0.81	0.96	1.14
B	1.02	0.80	1.18
C	3.02	2.69	3.99
D	2.94	3.47	4.26
E	2.86	3.04	4.14
F	0.78	0.88	0.79
G	2.63	2.82	3.88
H	2.32	2.34	3.15

contained interactions violating the van der Waals minimum distance of approach rule were discarded. Three families, including the largest one, remained. The lowest energy structure of each of these families was subsequently submitted for a second dynamics run at 300K. Each of these runs generated another hundred structures, which were similarly analysed (Figure 6.10(b)). A total of eight families were produced, and the lowest energy structure of each family was submitted to MORASS for spin diffusion calculations. The scaling factor used to scale the theoretical data output by MORASS was calculated in the same manner as before for Lom-AKH-I, i.e. an overall scaling factor was generated by the quotient of the sums of the two data sets. *R*-factors were then calculated using the resultant scaled data. A complete set of results is listed below in Table 6.4.

Table 6.4: Comparative table of results allowing assessment of the extent of agreement between experimental and theoretical data for Lom-AKH-III.

Structure	R_1	R_2	Energy (kcal/ mol)	# NOESY Viol.	# H-bonds	D-A dist.	D-H-A angle
A	0.15	0.02(7)	311.03	0	1	3.04	155.0
B	0.15	0.02(7)	311.95	0	1	3.04	155.1
C	0.17	0.03(0)	310.19	1	1	3.04	155.0
D	0.18	0.03(2)	310.61	1	1	3.04	154.8
E	0.18	0.03(1)	310.73	1	1	3.03	155.1
F	0.18	0.03(2)	308.50	1	1	2.99	155.5
G	0.18	0.03(2)	308.75	1	1	3.02	154.8
H	0.18	0.03(1)	309.27	1	1	3.01	154.7

In the above D-A dist. refers to the Donor-Acceptor distance and the angle defined is the Donor-Hydrogen-Acceptor angle.

All further analysis of the structure of Lom-AKH-III will be based on structure A, which showed both the lowest *R*-factor and no NOESY violations (Figure 6.11).

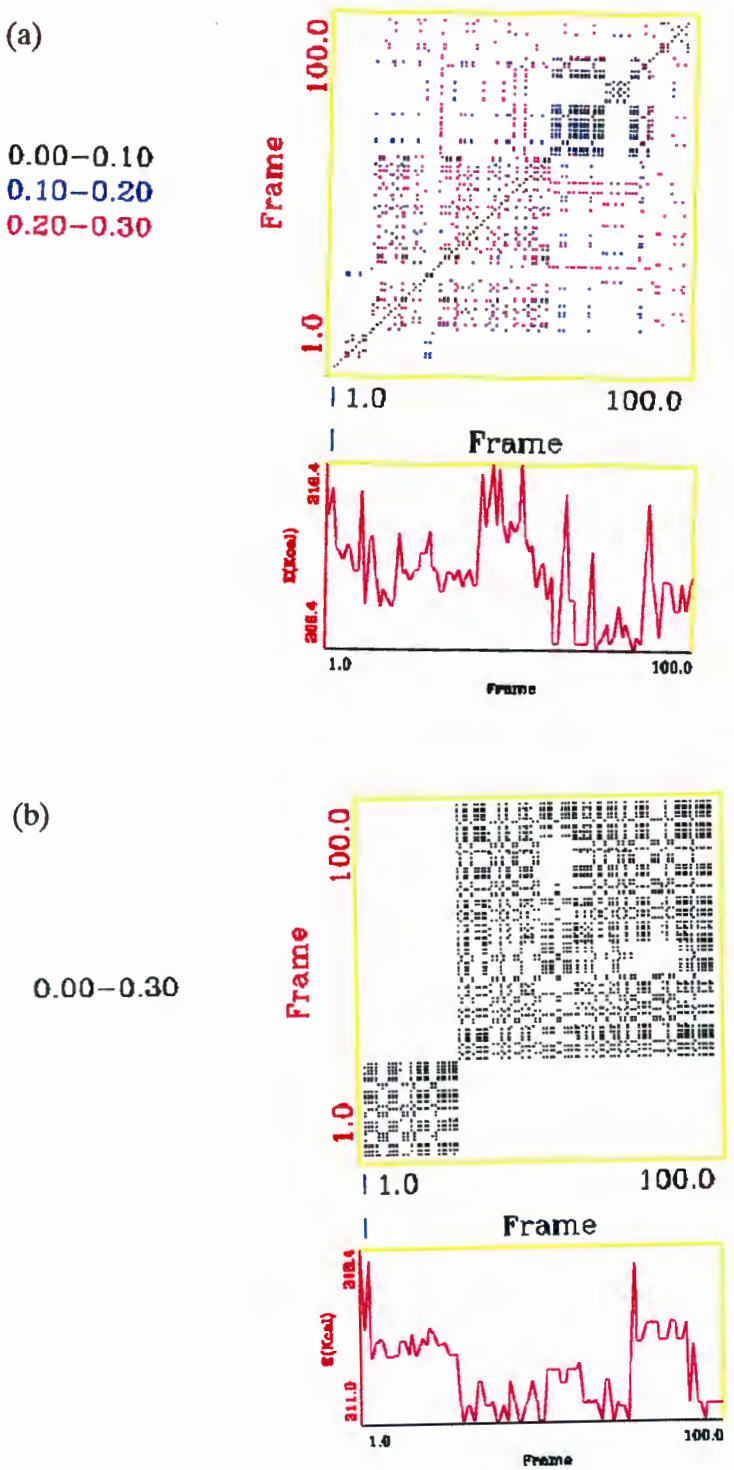


Figure 6.10: Cluster and energy vs frame graphs produced by (a) the first run performed at 600K and (b) one of the subsequent runs performed at 300K (the run which contained structure A). These graphs were required for generation of structure families.

Tables 6.5 and 6.6 illustrate the excellent agreement between the experimental and theoretical torsion angle and NOESY data respectively. Many of the possible intraresidue NH-αH interactions did not give peaks in the NOESY spectrum, implying that these protons are greater than 3.55Å apart. These constraints were also used in generating the final structures and when looking for NOESY violations. However, although the background count for the NOESY spectrum was calculated to be 0.00 ± 0.22, certain regions registered counts of up to 0.44. This would put nOe's resulting from distances of 3.55Å below the background count and therefore explain the absence of these peaks in the NOESY fingerprint region.

Table 6.5: A comparison between the theoretical and experimental torsion angle values.

Residue	Torsion Angle, ϕ	
	Exp.	Theor.
Leu	-148.3	-145.8
Asn	-150.9	-148.8
Phe	-148.2	-145.5
Thr	-154.5	-153.4
Trp	-154.5	-154.4
Trp	-148.3	-149.2

Lom-AKH-III formed a turn region which showed significant similarity to that observed in Lom-AKH-I - a loose open turn type involving only three residues, Thr, Pro and Trp_7, was again formed. The hydrogen bonding scheme observed was identical to that witnessed in Lom-AKH-I viz. $-X_5-O \cdots H-N-X_7-$ (Figure 6.12), and as before the turn could not be classified as any form of tight turn due to both angle and distance violations. This structure did demonstrate concavity, but to a far lesser extent than Lom-AKH-I - probably the result of its shorter primary sequence. There were two regions of high hydrophobicity - one in the turn region on the convex surface, where grouping of the Pro ring and two Trp side chains occurs, and one on the concave surface involving the Phe and Leu residue sidechains. The Asn sidechain created a hydrophillic region on the convex surface opposite to the hydrophobic cluster (Figure 6.13). Conservation of the turn within the family of structure A (Figure 6.14) and throughout the different runs was demonstrated by an overlay of the sample structures from each family which generated an *RMS* value of 1.26 for the backbone of the turn region -residue five to residue seven (Figure 6.15).

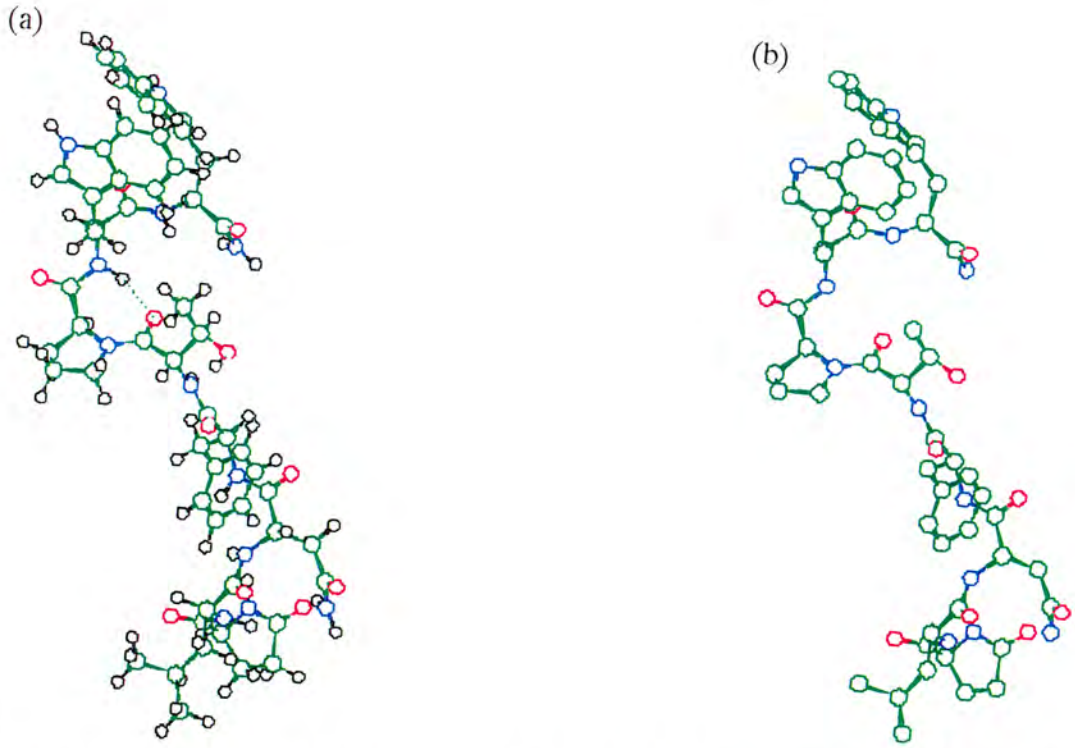


Figure 6.11: (a) Overview of structure A showing the hydrogen bonding scheme identical to Lom-AKH-I. The open turn structure is also clearly visible from this perspective and is emphasised in (b), the structure illustrated without the hydrogens.

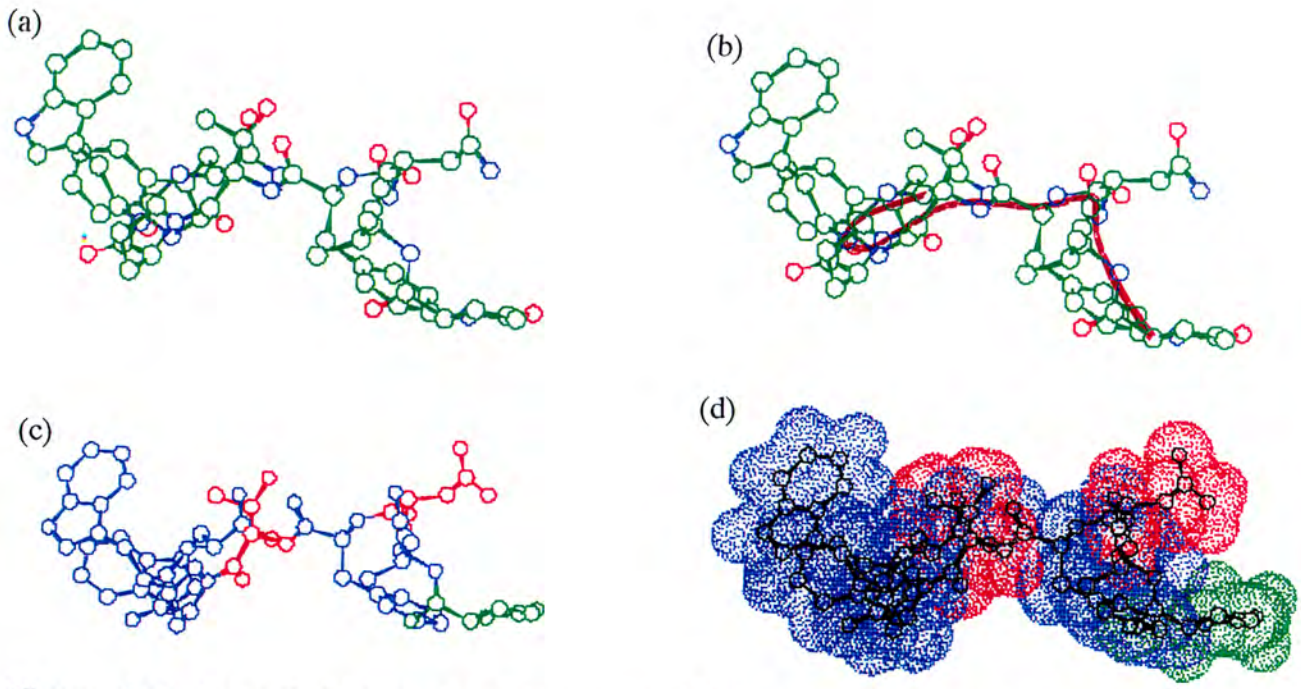


Figure 6.12: (a) and (b) Sideview of structure A illustrating the concave nature of the molecule - the latter has the addition of a brown backbone ribbon to emphasise the point, (c) the hydrophobic clusters on the convex surface in the region of the Pro residue, and on the concave surface involving the Leu and Phe sidechains. Also shown is the hydrophillic Asn sidechain on the convex surface opposite to the hydrophobic cluster, (d) van der Waals surface area of (c).

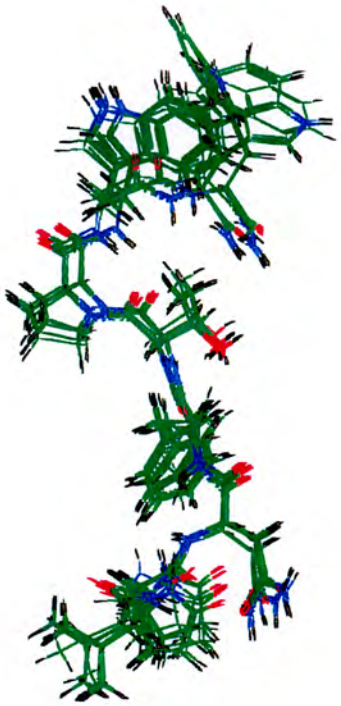


Figure 6.13: Family of structures having RMS differences of less 0.3 generated by the "family" module of Biosym which included structure A.

Figure 6.14: Overlay of all the structures analysed showing the good agreement between the different runs - (a) overview and (b) sideview of a backbone overlay, (c) and (d) overlay overview of a Pro overlay, without and with sidechains respectively, illustrating the good conservation of the turn region.

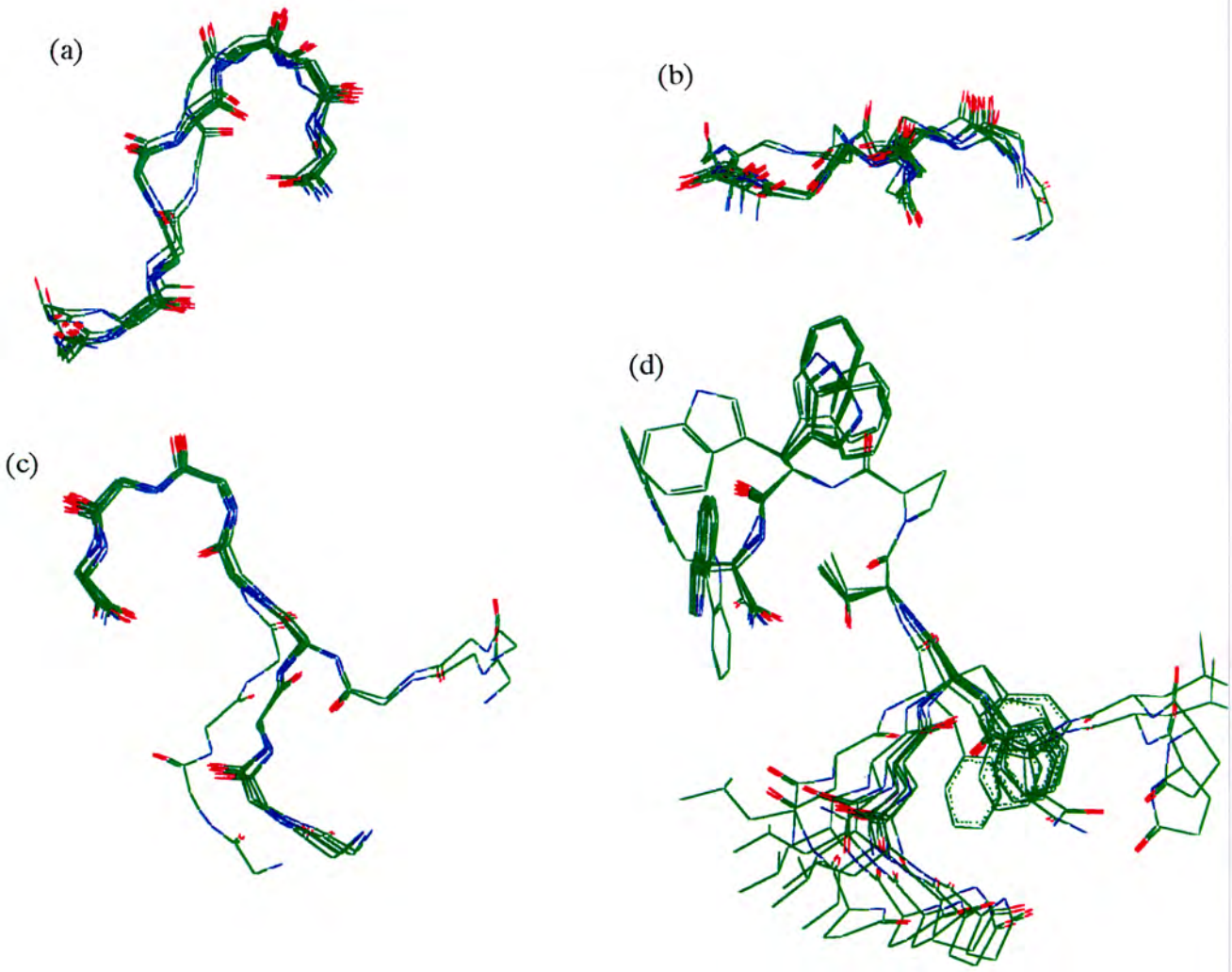


Table 6.6: Table showing the extent of agreement between the experimental and theoretical NOESY data.

Peak	Interaction	Distance	NOESY Integration		
			Exp.	Theor.	Diff.
A	Thr-NH···Thr-βH	3.05	0.96	1.18	0.22
B	Leu-NH···pGlu-αH	3.42	0.80	0.82	0.02
C	Asn-NH···Leu-αH	2.88	1.34	1.78	0.44
	Thr-NH···Thr-αH	3.00	1.34	1.22	0.12
D	Thr-NH···Phe-αH	2.58	3.47	3.72	0.25
E	Trp_7-NH···Pro-αH	2.66	3.04	3.04	0.00
F	Trp_7-NH···Trp_7-NH	3.15	0.88	1.19	0.30
G	Phe-NH···Asn-αH	2.85	2.82	2.28	0.54
H	Trp_8-NH···Trp_8-αH	2.83	2.34	1.76	0.58

When analysed with respect to *R*-factor calculations at different mixing times satisfactory values (R_1=0.20 and R_2=0.03(9) for the 150ms mixing time, and R_1=0.19 and R_2=0.03(8) for the 250ms mixing time) were obtained. These low values further demonstrated the viability of this structure as a possible secondary solution structure for Lom-AKH-III.

The structure postulated for Lom-AKH-I agrees with all the experimental data. This indicates that while the molecule is likely to be fluctional, there is a preferred conformation, which, by analogy with Lom-AKH-I, would be predicted to give Lom-AKH-III relatively high reactivity. Preliminary results have shown that, although this peptide has ED_{50} values higher than Lom-AKH-I and II for lipid mobilisation, at high doses it is capable of inducing a maximal response [3] - see later.

REFERENCES

1. Ouderjans, R.C.H.M.; Kooiman, F.P.; Heerma, W.; Versluis, C.; Slotboom, A.J.; Beenakkers, A.M.T. *Eur. J. Biochem.* **1991**, *195*, 351-359.

2. Heerma, W.; Versluis, C.; Lankhof, H.; Ouderjans, R.C.H.M.; Kooiman, F.P.; Beenakkers, A.M.T. *Analytica Chimica Acta*, **1991**, *248*, 553-561.

3. Ouderjans, R.C.H.M.; Kooiman, F.P.; Dijkhuizen, R.M; Beenakkers, A.M.T. *Proc. Exper. and Appl. Entomol.*, N.E.V. Amsterdam, **1992**, *3*, 165-166.

CHAPTER 7

CONCLUSION

The previously discussed results for Lom-AKH-I, II and III clearly indicated that Lom-AKH-I and III have definite structure in d^6-DMSO solution, whereas Lom-AKH-II was fluctional with no single structure satisfactorily fitting the experimental data. This information would seem to indicate that Lom-AKH-I and III have higher reactivity than Lom-AKH-II. Goldsworthy *et al.* previously proposed that a Pro residue at position six in the sequence was necessary for maximum activity [1]. This proposal was based on the idea that the β-bend generated by the Pro residue was essential for hormone-receptor binding [1,2]. This was later disproved when peptides other than Lom-AKH-II (which in certain cases shows comparatively low activities) which also contained Pro substitutions, showed 100% activity when tested in *L. migratoria*, for example Grb-AKH from the cricket and Scg-AKH-II from the desert locust [2]. Numerous studies have been carried out regarding the structure-activity relationships of each of these peptides [1-5]. It was shown by Stone *et al.* [3] that a minimum of eight residues was required if a peptide was to exhibit any sort of adipokinetic activity. Another essential requirement for maximal response are the blocked termini - pGlu (the L-enantiomer [3]) as the N-terminus and an amidated C-terminus [2]. Regarding Lom-AKH-I, the necessary existence of a Gly residue at position nine in the sequence has been demonstrated by the fact that the naturally occurring ...-Trp-Gly-ThrNH_2 sequence has significantly higher potency than a synthesised ...-Trp-Thr-GlyNH_2 sequence [3]. Trp at position eight is the only other invariant residue in the family of adipokinetic hormones [2]. The highly conserved genetic nature of these hormones suggests very specific hormone-receptor binding mechanisms. Mutations leading to amino acid substitutions which induce conformational changes and consequently destroy the binding capacity of a given hormone will not survive natural selection; thus the primary sequence of the hormone will to a large extent be conserved [2].

Although shown not be essential, it is still maintained that in peptides containing the Pro residue, the associated β-turn plays an important role in hormone-receptor binding for that particular molecule [2]. A type-I β-turn, involving the Thr_5, Pro, Asn_7 and Trp residues, was predicted for Lom-AKH-I [2,6]. A computational construction of this turn generated a hydrophobic cluster on one face of the molecule involving the Phe, Trp, and Leu residues [5]. This cluster has been suggested to play an integral role in the hormone-receptor binding process [5]. The results obtained for Lom-AKH-I and III both showed hydrophobic clusters on their convex surfaces. Lom-AKH-I was also shown to have a relatively hydrophillic pocket on the concave surface of the peptide.

CD results showed that although Lom-AKH-I appeared to adopt a random coil structure in an aqueous environment, as soon as SDS was added to the solution, thereby simulating membrane conditions, the peptide appeared to adopt a β-turn - predicted to be a type-I β-turn [6]. (Suzuki *et al.* have reported in protein crystal studies that more than 40% of Thr-Pro-XX type sequences adopt a type-I β-turn or one of the other closely related turn structures, σ- or τ-turn types [7,8].) An additional CD spectrum generated in water at 170K revealed that at this temperature, the peptide adopts some form of definite structure [6]. This structure could not be classified as an ordinary β-turn, but it has been suggested that the spectrum could be characteristic of a left-handed helical structure recently reported by Drake *et al.* [6,9]. The conclusion was therefore made that the spectrum recorded in water at 294K was the average between the two extreme spectra - that obtained on addition of SDS and that obtained at 170K [6]. This prediction of an averaged conformation state is consistent with the results obtained for Lom-AKH-I using NMR restrained molecular dynamics. Although a distinct turn region was identified, it was not any form of β-turn. The turn observed in both Lom-AKH-I and Lom-AKH-III only involved Thr-Pro-X, with no involvement of a fourth residue, i.e. the residue lying at position eight of the sequence. The geometrical arrangement of the Phe residue at position four in the sequence also contributed to irregularity in the turn region, further encouraging the formation of a very open, loose and spacious type of turn only stabilised by the presence of a single hydrogen bond between the carbonyl oxygen of residue i and the amide proton of residue (i+2) - where Thr is residue i in the turn, Pro, i+1, etc. A standard β-type-I turn would include a hydrogen bond between the carbonyl oxygen of residue i and the amide proton of residue (i+3) (Figure 7.1).

This irregular turn type was confirmed not only by the fact that there was no agreement between the experimental and theoretical ϕ torsion angle values expected for a type-I β-turn, but also by the absence of nOes between the Pro-αH and the amide proton of residue (i+2). (Although crosspeaks for this interaction are expected for both type-I and II β-turns, they are expected to be stronger in the latter turn type.) Thus, in conclusion, the NMR results obtained for Lom-AKH-I and Lom-AKH-III do show definite structure. In the case of Lom-AKH-I this seems to be in good agreement with the CD data previously collected and is consistent with the prediction that at room temperature a definite structure lying between that of a β-turn and a low temperature structure is observed. Attempts at producing NMR spectra for Lom-AKH-I in SDS were unsuccessful, and so it was not possible to verify the structure predicted under these conditions. The definite structure of Lom-AKH-I was proposed to be analogous to the intermediate structure obtained from the CD spectrum recorded in water at room temperature. This indicates that the use of d^6-DMSO, and not water, as a solvent does not necessarily adversely affect the results obtained for these peptides. The similarity between peptide NMR data recorded in d^6-DMSO and D_2O has

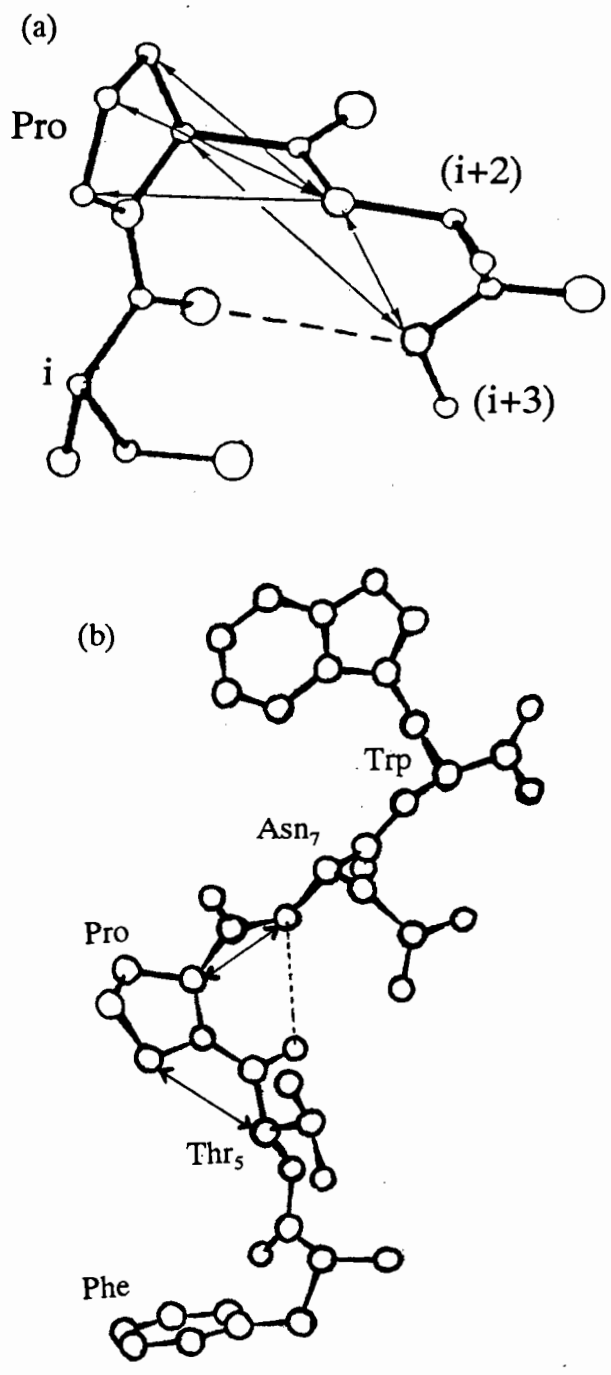


Figure 7.1: (a) A standard SPXX type-I β-turn. Solid lines indicate expected nOe connectivities; dotted lines represent possible hydrogen bonds. (Taken from Suzuki *et al.* [8]) (b) Turn type obtained for Lom-AKH-I. Solid lines represent the nOes obtained from the NOESY spectrum; dotted line represents the hydrogen bond present in this structure.

previouly been recorded [10,11].

Regarding the existence of β-sheet structures in the case of Lom-AKH-I and III, the sole existence of sequential intrastrand $d_{\alpha N}(i,i+1)$ interactions is not sufficient. To unambiguously deduce the existence of a β-sheet conformation, in addition to the above, there must also be evidence of interstrand $d_{NN}(i,j)$, $d_{\alpha N}(i,j)$ $d_{\alpha\alpha}(i,j)$ nOes [12]. Since none of these interstrand nOes were observed, the possibility of a β-sheet structure in terms of the peptide folding back on itself was discarded. The coupling constant values of between 6 and 9Hz further confirmed the absence a regular β-sheet structure - the range expected for β-structures is between 9 and 12Hz [13]. The experimental values obtained indicate that, although these two peptides do have definite conformations, they do not adopt any regular secondary structure. In addition to a loose turn around the Pro residue, both of these structures also displayed concavity when viewed side-on.

All molecular mechanics simulations were carried out in vacuum. It has been suggested that a dielectric should be added to the calculations as a final step in the structure generation process [14]. One reason for the acceptance of *in vacuo* conditions was that all the molecules were uncharged species and any solute-solvent, interactions would therefore only be the result of hydrophobic/hydrophillic interactions. Although the version of MacroModel used did have certain specific solvent modules [14], these did not include one for d^6-DMSO. The only solution to this problem was therefore simple inclusion of a distance dependent dielectric constant. This process was, however, not as well refined or as reliable as an actual solvent module. Bremer *et al.* [15], observed that the use of a variable dielectric in MM calculations applied to small charged molecules was unsatisfactory. The viabiliy of applying a macroscopic dielectric to microscopic domains has yet to be established [16]. Biosym's Discover module uses only the dielectric approach in an attempt to model solvent effects and thus, because of the uncertainty of this approach, together with the fact that further development in this field (development of specific models pertaining to solvent-solute interactions) is required before accurate determinations can be made, these calculations were excluded from the project [16].

One of the problems experienced with MD is that it invariably, if not always, provides a solution. However, this answer may not be realistic or bear any relationship to the real

structure. In this case though, the legitimacy of the use of NMR and MD as techniques for structure solution was demonstrated through the difference between the results obtained for Lom-AKH-I and III and those obtained for Lom-AKH-II. Whereas Lom-AKH-I and III showed definite structure, the same could not be said for Lom-AKH-II. This illustrated that the MD calculations did not generate impossible or unreasonable structures irrespective of the constraints applied. The fact that it was necessary to generate at least three structures in order to get satisfactory agreement between experimental and theoretical data for Lom-AKH-II, clearly illustrates that MD calculations in conjunction with the NMR constraints do indeed generate acceptable structures and that results are feasible.

The activity of peptides is generally measured through the generation of dose-response curves. This activity is usually considered in terms of ED_{50} values which correspond to the dose required to induce half the maximum response of a given peptide. Studies by Goldsworthy *et al.* [17] on the relative activities of Lom-AKH-I and II with respect to their effect on haemolymph lipid levels, glycogen phosphorylase activation and the increase in cAMP levels in the fat body (Figure 7.2), clearly show Lom-AKH-I to be the more potent of the two in terms of hyperlipaemic activity, formation of lipoprotein A^+ and glycogen phosphorylase activation. Lom-AKH-II has, however, been shown to be the more active peptide as far as raising the fat body cAMP levels is concerned [17][3]. Thus, although both peptides effect the same processes within the locust, they do so to different extents. Based on the fact that Lom-AKH-II is more effective in cAMP activation and that even in high doses (8pmol AKH-II:2pmol AKH-I) it fails to effectively compete with Lom-AKH-I in lipid mobilisation, Goldsworthy *et al.* [17] suggested the possibility of Lom-AKH-I and II binding to different receptors. It has, therefore, been proposed that the complete function of Lom-AKH-II has not, as yet, been fully elucidated. Loughton and Orchard [18] reported Lom-AKH-II to have some hyperglycaemic activity in locusts - this suggestion, however, contradicts the later findings in which Lom-AKH-I has been shown to be more effective than AKH-II in phosphorylase activation [2,17]. (There is one exception to this observation and that is by Ouderjans *et al.*

---

[3]According to Gäde, bioassays performed in *L. migratoria* showed Lom-AKH-I to quickly induce a 100% maximum response (ED_{50}<1pmol) in increasing the haemolymph lipid levels , while Lom-AKH-II plateaued at 70% (ED_{50}~5.8pmol) where it was maintained even at high dose levels [2].

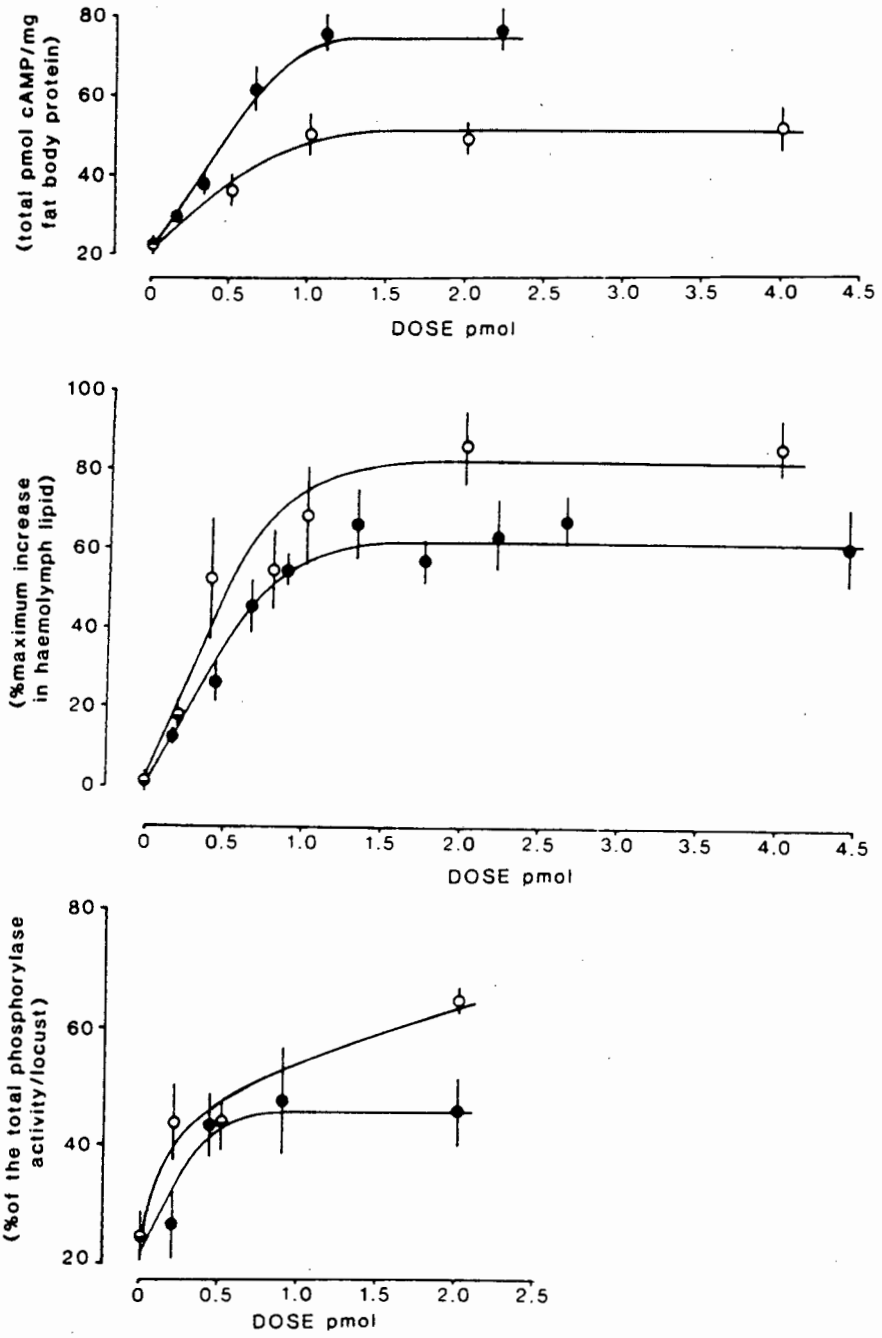


Figure 7.2: Dose response curves showing the comparative effects of natural AKH-I (open symbols) and II (solid symbols) on (a) haemolymph total lipid titre, (b) fat body cAMP, (c) glycogen phosphorylase activation and (d) lipoprotein A^+ formation (measured as increase in HEPP (heparin/EDTA) - precipitate material. ((a), (b) and (c) taken from Goldsworthy and Wheeler [19]; see also [17,20], (d) taken from Goldsworthy and Wheeler [17].)

who report not only that the ED_{50} values for phosphorylase activation by AKH-I and II are the same, but also that AKH-II seems to show a higher maximal response than AKH-I [21]). Because of the discrepancy between these results and those of previous experiments [17], the data recorded for Lom-AKH-III will be briefly mentioned, but not extensively discussed. Further research is therefore required to clarify the exact physiological role played by Lom-AKH-II in *L. migratoria* [17,20].

So far there has only been one publication comparing the relative activities of all three of the Lom peptides [21]. Interestingly, these results show that although Lom-AKH-III has a much higher ED_{50} value than either AKH-I or II at high doses (~3.5pmol), it reaches the same maximal response as AKH-I for lipid mobilisation. Lom-AKH-III showed low potency with regards to phosphorylase activation (Figure 7.3) [21]. Studies have also been done on mixtures of two or three of the hormones [21]. With regards to lipid mobilisation, these led to maximal responses lower than those for the individual hormones, with the exception of the Lom-AKH-I/III combination. However, all combinations had lower ED_{50} values indicating a higher potency in the lower dose range. The reverse was observed for phosphorylase activation where combinations generally led to higher maximal responses than those for the individual hormones. Although the AKH-I/II combination had an ED_{50} value greater than either hormone on its own, higher potencies (lower ED_{50} values) were reported for the AKH-I/II/III and AKH-I/III combinations [21].

To summarise, the CD results obtained for Lom-AKH-I suggested that peptide structure is directly affected by the presence of a membrane, and that this membrane could be essential for the formation of the correct binding structure [6]. The observation that certain small peptides have no structure until introduction of a membrane has been made before [22]. For this reason it was very difficult to make predictions regarding the activities of the three peptides studied. Lom-AKH-I and III both exhibited a definite structure in solution, i.e. a single structure could be generated which satisfied all the experimental data. This was in agreement with the CD data recorded for Lom-AKH-I. Both Lom-AKH-I and Lom-AKH-III had similar structure, adopting very similar turn types and having identical hydrogen bonding schemes. The fact that Lom-AKH-III is an octa- and not a decapeptide constitutes the greatest difference between the two peptides. This would account for the apparent incompleteness of

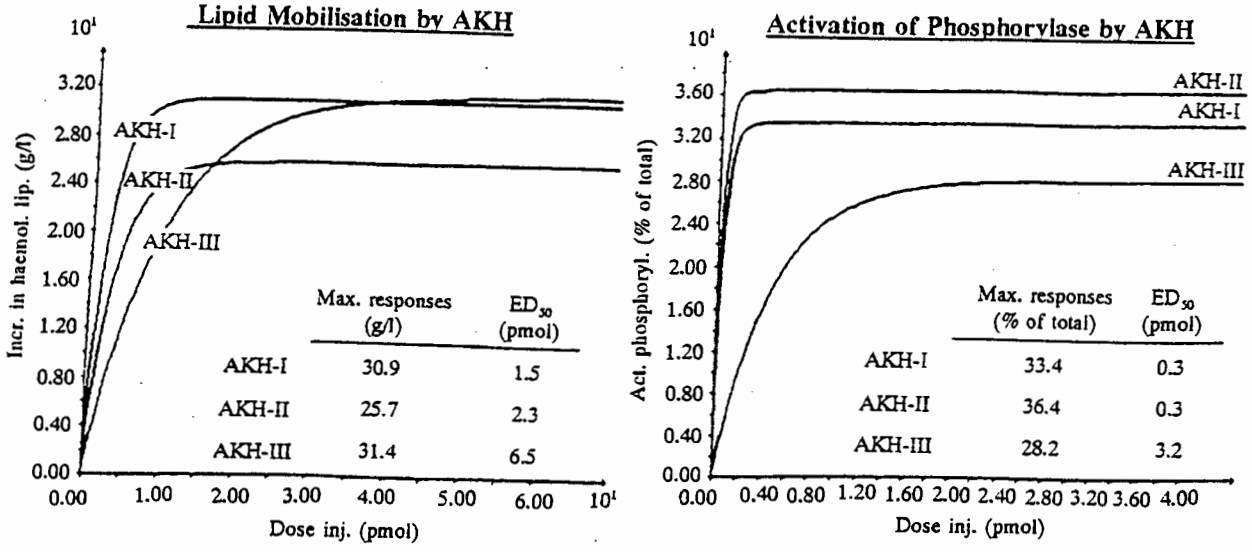


Figure 7.3: Dose response curves for haemolymph lipid mobilisation and glycogen phosphorylase activating effects of Lom-AKH-I, II and III in *Locusta migratoria*. (Taken from Ouderjans *et al.* [21])

the turn witnessed in Lom-AKH-III and could perhaps also have been the cause of its significantly higher ED_{50} values with regards to lipid mobilisation [21]. If the same trend were to be followed as for Lom-AKH-I, β-turn formation would be predicted in the presence of a membrane, but now the β-turn would occur at the peptide terminus, i.e. Trp_8, the last residue, would be included in the turn. Conversely, the structural similarities could account for the fact that Lom-AKH-III does eventually exhibit the same maximal response as Lom-AKH-I and also for the superior maximal response value generated by Lom-AKH-I/III combinations in lipid mobilisation. This combination also showed higher potency for phosphorylase activation.

Lom-AKH-II was indicated to be fluctional with no single dominant structure. These results are also in agreement with CD data obtained for this peptide, which showed no evidence of the formation of a β-turn either in the presence or absence of simulated membrane conditions (in fact no definite structural features were observed in any of the CD spectra recorded) [5]. The absence of a β-turn may be directly attributed to the lack of a Pro residue. However, as mentioned previously, this is not necessarily the cause of its higher ED_{50} and lower maximal

response values recorded for lipid mobilisation and in most cases phosphorylase activation. The fact that Lom-AKH-II was shown to exhibit greater reactivity with respect to increasing fat body cAMP levels indicates that in the presence of certain receptors, this peptide will be more conducive to adopting a structure having high affinity for the relevant site than Lom-AKH-I[4]. This adds to the previous suggestion that different receptors for Lom-AKH-I and II could exist [5]. Although there seems to be a significant amount of overlap with respect to function, there is a difference between the affinity of each of these peptides for a particular receptor, i.e. although able to bind to a variety of receptors, Lom-AKH-I will exhibit a far greater affinity for certain ones than others and vice versa for Lom-AKH-II, even though it shows a distinct lack of structure.

Ouderjans *et al.* [23] also carried out extensive studies regarding the biosynthesis and content of AKH-I, II and III in the CC during development from the larval through to the adult stages. Content is directly related to a combination of biosynthesis, release and degradation processes, and therefore gives an idea of the relevance of each these peptides at the different stages of development [23]. It was shown that there was virtually rectilinear increase in the content of AKH-I and II in the CC, while that of Lom-AKH-III was shown to be exponential (Figure 7.4), thus giving a constant AKH-I/II ratio and an irregular AKH-I/III ratio [23] (Figure 7.5). Biosynthesis of AKHs was measured in terms of the incorporation of radioactive Phe, a residue common to all three peptides. For Lom-AKH-I and II biosynthesis increased during larval stages of development, but levelled off and became constant during the adult stage. AKH-III, however, showed a continuous, although not linear, increase [23]. (Figure 7.6) Once again the AKH-I/II ratio was virtually constant, while that observed for AKH-I/III showed the same irregularity as before [23] (Figure 7.7). A study was also done to determine the DNA content of the CC during the different development stages. The results showed the same trend as those obtained for the biosynthesis of AKH-I and II, indicating a direct relationship between hormone biosynthesis and hormone gland growth [23]. This is consistent with the fact that the AKH-I and II genes are linked, and both of these hormones are stored together in the secretory granules [23]. The AKHs are present in much greater quantities than are necessary for physiological response. The fact that the content of AKH-I and II increases

[4]no data is as yet available for Lom-AKH-III regarding its effect on cAMP levels.

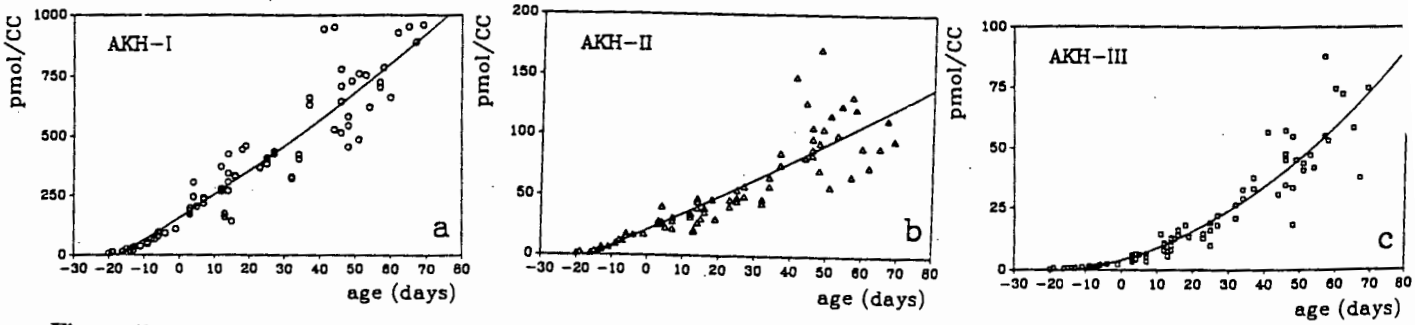


Figure 7.4: Content (pmol/CC) of (a) AKH-I, (b) AKH-II and (c) AKH-III in the CC of the female *Locusta migratoria* during larval and adult development. Day 0: imaginal ecdysis. (Taken from Ouderjans *et al.* [20].)

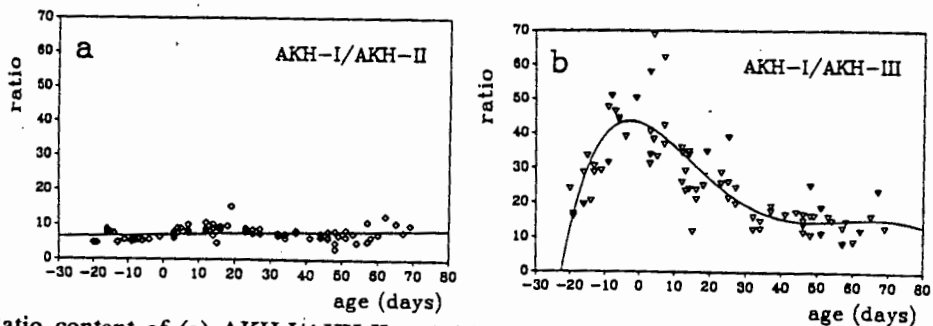


Figure 7.5: Ratio content of (a) AKH-I/AKH-II and (b) AKH-I/AKH-III of the CC of the female *Locusta migratoria* during larval and adult development. (Taken from Ouderjans *et al.* [23].)

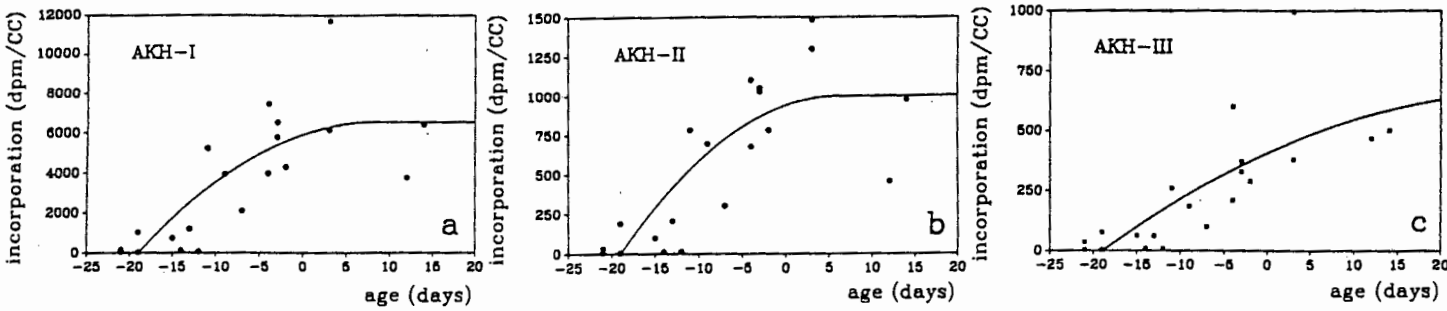


Figure 7.6: Incorporation of labelled Phe into (a) AKH-I, (b) AKH-II and (c) AKH-III of female *Locusta migratoria* during larval and adult development. (Taken from Ouderjans *et al.* [23].)

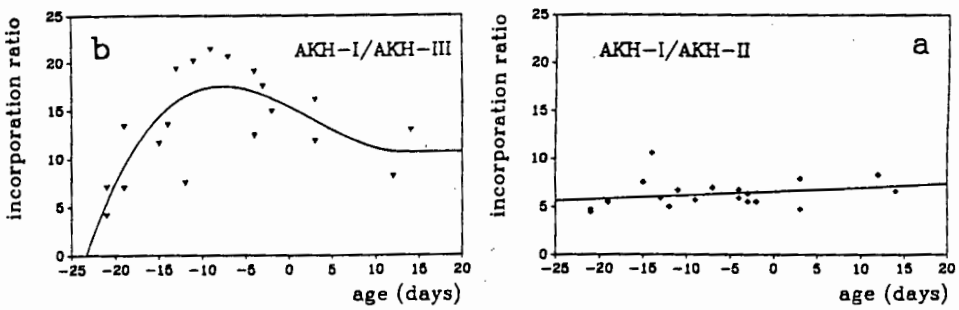


Figure 7.7: Ratio of incorporation of labelled Phe into (a) AKH-I/AKH-II and (b) AKH-I/AKH-III of female *Locusta migratoria* during larval and adult development. (Taken from Ouderjans *et al.* [23].)

steadily with age while biosynthesis levels off in the adult could thus be the result of increased biosynthesis having little effect on the AKH content in the larval stages due to only small amounts of AKH being produced [23]. The different trend observed for Lom-AKH-III could be attributed either to the continued increase in biosynthesis in the adult stage or, more importantly, to it playing a significant, but as yet unknown, role in larval development, thus leading to its reduced content in the larvae even though biosynthesis is high [23].

Although it was originally thought that peptides consisting of about twenty amino acid residues or less were too short to exhibit any sort of defined secondary structure, particularly the formation of β-turns and helices, numerous recent studies have shown the opposite, and stable structures are now being reported for many short peptides dissolved in aqueous solutions at room temperature [24-28]. Although still fluctuating between a number of different conformations, a certain particular conformation was proved to be more stable than the others and therefore more highly populated [24]. The results obtained in this thesis for Lom-AKH-I, II and III illustrated both scenarios (both confirmed by the relevant CD data). Whereas Lom-AKH-I and III were shown to have a single predominant structure in solution, Lom-AKH-II was shown to be highly fluctional existing as at least three different conformers. Thus it may be concluded that secondary structure determination of these three insect neuropeptides, Lom-AKH-I, II and III, making use of NMR and molecular mechanics calculations, was effectively carried out. These structures were, however, determined in d^6-DMSO solution. In order to be able to make conclusive remarks regarding the final structures, it would be necessary to carry out studies in solutions more synonymous to the natural environments in which these peptides occur. More work is also required in the determination of other functions of these peptides (particularly with regards to AKH-II and III), as well as in the study and identification of the AKH receptors and the hormone-receptor binding processes. This work is essential if the absolute role played by these neuropeptides in the metabolic flight pathways is to be completely elucidated in the future. Another area for future research would be to extend the structure/activity correlations to modified analogues of these three naturally occurring peptides.

REFERENCES

1. Goldsworthy, G.J.; Mallison, K.; Wheeler, C.H.; Gäde, G. *J. Insect Physiol.* **1986**, *32(5)*, 433-438.

2. Gäde, G. *J. Insect Physiol.* **1990**, *36(1)*, 1-12.

3. Stone, J.V.; Mordue, W.; Broomfield, C.E.; Hardy, P.M. *Eur. J. Biochem.* **1978**, *89*, 195-202.

4. Poulos, C.; Karagiannis, K.; Lee, M.; Goldsworthy, G.J. *Int. J. Peptide Protein Res.* **1994**, *44*, 589-593.

5. Goldsworthy, G.J.; Mordue, W. *Biol. Bull.* **1989**, *177*, 218-224.

6. Goldsworthy, G.J.; Drake, A.F.; Wilmot, C.M.; Wheeler, C.H.; Gäde, G.; Thornton, J.M., unpublished observations.

7. Suzuki, M.; Yagi, N. *Proc. R. Soc. Lond.* **1991**, *246*, 231-235.

8. Suzuki, M.; Gerstein, M.; Johnson, T. *Protein Engineering* **1993**, *6(6)*, 565-574.

9. Drake, A.F.; Silligardi, G.; Fibbons, W.A. *Biophys. Chem.* **1988**, *31*, 143-146.

10. Charey, K.V.R.; Srivastava, S.; Hoshur, R.V.; Roy, K.B.; Govil, G. *Eur. J. Biochem.* **1986**, *158*, 323-332.

11. Wessels, P.L.; Feeney, J.; Gregory, H.; Gormley, J.J. *J.C.S. Perkin II* 1973, 1691-1698.

12. Dyson, H.J.; Wright, P.E. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 519-538.

13. Dyson, H.J.; Wright, P.E. *Two Dimensional NMR Spectroscopy, Applications for Chemists and Biochemists* (Croasmun, W.R.; Carlson, R.M.; eds.), VCH Publishers, Inc., 1994, 655-698.

14. *Macromodel Handbook*, Department of Chemistry, Columbia University, New York, 1990

15. Bremer, J.; Mendz, G.L. *Aust. J. Chem.* **1988**, *41*, 1841-1856.

16. *Discover User Guide*, Biosym Technologies, San Diego, 1993, Part 1, 1/2-1 - 1/2-119.

17. Goldsworthy, G.J.; Mallison, K.; Wheeler, C.H. *J. Insect Physiol.* **1986**, *32(1)*, 92-101.

18. Loughton, B.G.; Orchard, I. *J. Insect Physiol.* **1981**, *27*, 383-385.

19. Goldsworthy, G.J.; Wheeler, C.H. *Insect Flight*, (Danthanarayana, W.; ed.), Springer-Verlag, Berlin, 1986, 49-59.

20. Raabe, M. *Recent Developments in Insect Neurohormones*, Plenum Press, New York, 1989.

21. Oudejans, R.C.H.M.; Dijkhuizen, R.M.; Kooiman, F.P.; Beenakkers, A.M.T. *Proc. Exper. and Appl. Entomol.* **1992**, *3*, 165-166.

22. Personal Commun.

23. Oudejans, R.C.H.M.; Mes, T.H.M.; Kooiman, F.P.; Van Der Horst, D.J. *Peptides*, **1993**, *14*, 877-881.

24. Hinds, M.G.; Welsh, J.H.; Brennand, D.M.; Fisher, J; Glennie, M.J.; Richards, N.G.J.; Turner, D.L.; Robinson, J.A. *J. Med. Chem.* **1991**, *34*, 1777-1789.

25. Dyson, H.J.; Rance, M.; Houghten, R.A.; Lerner, R.A.; Wright, P.E. *J. Mol. Biol.* **1988**, *201*, 161.

26. Jelicks, L.A.; Naider, F.R.; Shenbagamurthi, P.; Becker, J.M.; Broido, M.S. *Biopolymers* **1988**, *27*, 431.

27. Haas, E.; Montelione, G.T.; McWherter, C.A.; Scheraga, H.A. *Biochemistry* **1987**, *26*, 1672.

28. Reed, J.; Lieth, W.E.; Kubler, D.; Suhai, S.; Kinzel, V. *Eur. J. Biochem.* **1988**, *178*, 141.

CHAPTER 8

EXPERIMENTAL

8.1 Sample Preparation

All three peptides were obtained from the Zoology Department, University of Cape Town. Purification of the samples, necessary for successful use of NMR techniques, was carried out using high performance liquid chromatography (HPLC) on a Gilson (GG) HPLC instrument using a nucleosil C8 semi-preparative column. The eluting solvent consisted of a 60% acetonitrile solution (solution A) and distilled water (solution B). In order to force these peptides off the column, the percentage of A in the eluting solvent was increased by 10% over a twenty minute run. The actual percentage of A required for the eluting process was dependent on the hydrophobicity of the peptide. Once collected off the column, the sample was freeze dried and weighed.

8.2 NMR Spectroscopy

All NMR spectra were recorded on a Varian Unity 400MHz spectrometer using a 5mm inverse detection probe. The hydrophobic nature of all three peptides made them insoluble in water, and thus an alternative relatively non-polar solvent was required - in this case d^6-DMSO was the solvent of choice. For each peptide, an initial temperature arrayed proton experiment was carried out in order to establish the temperature at which maximum spectral resolution occurred. Subsequent to this, COSY [1], TOCSY [2,3], Homonuclear 2DJ-Resolved [4,5] and NOESY (150, 200, 250 and 300ms) [6] spectra were recorded. The relatively long mixing times used in the NOESY experiments were necessary for the sufficient build-up of NOE required for generation of adequate signal to noise ratios. In many of these experiments solvent suppression, using presaturation methods, was necessary to diminish a dominant H_2O

peak in the resultant spectra and induce significant improvements in the corresponding signal to noise ratios. Presaturation was carried out using either the transmitter or the decoupler (see Chapter 2). The choice of method was dictated by the experimental requirements, and their use is recorded in the parameter sets attached to each spectrum in Appendix 2.

The chemical shift reference was, in all cases, set at the centre of the H_2O solvent peak. This setting was confirmed by checking that the corresponding chemical shift at the centre of the d^6-DMSO solvent peak was in agreement with the literature value (δ=2.5). In processing of the 2D spectral data, a sine bell, shifted sine bell and Gaussian window function was applied in both dimensions [1,7]. Full details for the processing are supplied with each spectrum. A baseline correction was also applied to all NOESY spectra.

Estimations of the αH-NH coupling constants were made either from the proton or Homonuclear 2DJ-Resolved spectra,while information regarding dipolar coupling was obtained from the NOESY experiments. Computerised peak integration using the Varian software was carried out as a means of generating accurate intensity measurements.

In an effort to simulate membrane conditions, several attempts were made to record spectra for the peptides dissolved in aqueous SDS solution. However, none of these attempts were successful and the study was discontinued.

8.3 Structure Calculations

Both the MacroModel [8] and Biosym (Discover module) [9] software packages were used for molecular mechanics calculations. MacroModel calculations were carried out on a VAX/VMS system, while the Biosym was run on a Silicon Graphics machine using an Indigo^6 R4000 processor.

8.3.1 MacroModel

MacroModel, version 3.0 (released February 1990), and its associated batch-mode molecular modelling facility - Batchmin, version 3.1 (released April 1990), were used to carry out

energy minimisations and extensive Monte Carlo searches. The united atom AMBER force field developed by S.J. Weiner specifically for application to nucleic acids and proteins [10] was the force field of choice for all MacroModel molecular mechanics calculations. A global use-directed search mode with an energetic window set to 50kJ/mol was employed for all the MC searches. The variable torsion angles were predominantly assigned to torsions associated with α-carbon - carbonyl carbon atom pairs, although certain β-carbon - γ-carbon atom pairs of residues with long side chains were also selected. The PR conjugate gradient method was employed for structure minimisation using derivative convergence and a convergence limit of 0.01kJ/Å-mol. In order to reduce the number of structures stored, the backbone of any structure generated was compared to that of previously stored structures. Structures were considered to be duplicates unless a least squares superimposition of the backbone atoms located one or more pairs of equivalent atoms separated by more than 0.25Å. A chirality check of all the α-carbon atoms was also necessary at the end of each MC step to ensure that their chirality was maintained throughout the search. Due to time constraints, it was necessary to terminate the search after 3150 structures had been minimised. The results saved through creation of a STOP file. MC searches were typically allowed to run for between one and four days of CPU time.

8.3.2 Biosym

The builder module of Biosym was used to construct the straight chain peptide, which was optimised to reduce the immediate occurrence of high-energy steric interactions. These structures were used in subsequent restrained molecular dynamics calculations.

The extensive overlap in the NOESY fingerprint region prevented direct generation of a reasonable set of distance constraints. For this reason it was necessary to carry out an initial series of dynamics runs in order to allow constraint manipulation until reasonable agreement between experimental and theoretical data was achieved. The final distance constraint set was used in subsequent simulated annealing calculations. Two types of distance constraint were implemented - those forcing protons closer together in agreement with the existing nOes, and those forcing protons further apart in agreement with absent nOes. The torsion angle constraints were applied to all calculations.

All molecular dynamics calculations were carried using the Discover module, version 2.9.0/3.1.0 (released January 1993) in conjunction with the Consistent Valence Forcefield (CVFF) developed specifically for amino acids [11]. Each dynamics run followed the same general procedure: the peptide was initially minimised using the steepest descent and conjugate gradient methods until maximum derivatives of less than 10 and 0.1kcal/Å-mol respectively were achieved. Thereafter the system was equilibrated at constant temperature for 10ps to allow for uniform distribution of kinetic energy throughout the system. Dynamics was then resumed at the same temperature for 1000ps, during which time structures were regularly saved at 10ps intervals. The total energy of the system was kept constant over this period, thereby allowing the different bond types to take on more realistic energies. Because it was the total energy that was being kept constant, small variations in the temperature now occurred. On completion of this run, the archive file containing the one hundred saved structures was accessed and each structure minimised in the same way as before, with the exception of the maximum derivative for the conjugate gradient minimisation, which was now reduced to 0.001 kcal/Å-mol. The hundred structures produced at the end of this first run were divided into families using the Analysis module. This was done by superimposing the structures, and those differing by less than a specified minimum *RMS* value were grouped together to form a family. The minimum *RMS* values were calculated as follows [12]:

$$RMS = \left[\sum_{i=1}^{N} \frac{(x_i - x_i')^2 + (y_i - y_i')^2 + (z_i - z_i')^2}{N} \right]^{1/2} \qquad \textbf{(8.1)}$$

where x, y and z correspond to atomic coordinates, and N is the number of atoms in each molecule. The actual annealing process involved performing the first run at a temperature of 900K and generating the appropriate structure families. A sample structure from each family was subsequently subjected to a second dynamics run, this time at a reduced temperature of 600K. The process was again repeated at a temperature of 300K. This annealing process was essential for the location of low energy minima, since dynamics runs performed at high temperatures of 900K tend to locate higher energy minima and therefore high energy conformers [13]. The reason for the high temperature run is to maximise coverage/sampling of conformational space. This can be explained using Arrhenius' model for reaction rates, which is given by:

$$k = Aexp[-\Delta E/RT] \tag{8.2}$$

where k is the reaction rate, A, the pre-exponential function, T, the temperature, ΔE, the activation energy and R, the universal gas constant. From this equation it may be seen that a decrease in temperature is accompanied by a substantial decreases in collision rate and therefore also in the chances of overcoming high energy barriers [13]. However, because these high temperatures often lead to structures containing interatomic distances violating van der Waals minimum distance of approach rule, the starting temperature is often reduced. Although the above was the intended theoretical simulated annealing protocol, it was often varied in accordance with problems encountered with each individual peptide. Therefore the actual procedure followed in each case is described separately for each peptide in the relevant results and discussion section. The actual programs used in each case, including the force constants applied in constraining distances and torsions, are supplied in Appendix 3.

8.3.3 Structure Analysis

In order to assess the extent of agreement between experimental and theoretical data, structures were submitted to MORASS, version 2.1 (released May 1994) [14]. This program served to generate a theoretical NOESY spectrum corresponding to an input structure which could then be compared to the experimental data. MORASS was essential both for final structure identification, and for the initial process of data manipulation required to generate a reasonable set of distance constraints. The extent of agreement between the two data sets was estimated through the determination of R- and $R^{1/6}$ factors, R_1 and R_2 respectively, as follows:

$$R_1 = \frac{\sum |V_{theor} - V_{exp}|}{\sum V_{exp}} \tag{8.3}$$

$$R_2 = \frac{\sum |(V_{theor})^{1/6} - (V_{exp})^{1/6}|}{\sum (V_{exp})^{1/6}} \tag{8.4}$$

R_1 values of less than 30% were considered feasible, and the corresponding structures were accepted and saved.

REFERENCES

1. *Guide to NMR experiments*, VNMR version 4.3, Varian Associates Inc., 1993.cosy

2. Levitt, M.; Freeman, R.; Frenkiel, T. *J. Magn. Res* **1982**, *47*, 328.

3. Bax, A.; Davis, D. *J. Magn. Res* **1985**, *65*, 355.

4. Aue, W.; Karhan, J.; Ernst, R. *J. Chem. Phys.* **1976**, *64*, 4226.

5. Nagayama, K.; Backmann, O.; Wüthrich, K.; Ernst, R. *J. Magn. Res* **1978**, *31*, 133.

6. States, D.J.; Haberkorn, R.A.; Ruben, D.J. *J. Magn. Res* **1982**, *48*, 286-292.

7. Derome, A.E. *Modern NMR Techniques for Chemistry Research*, Pergamon Press, Oxford, 1987.

8. *MacroModel Handbook*, Department of Chemistry, Columbia University, New York, 1990

9. *Discover User Guide, Version 2.9.0/3.1.0*, Biosym Technologies, San Diego, 1993, Parts 1-3.

10. Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765-784.

11. *Discover User Guide, Version 2.9.0/3.1.0*, Biosym Technologies, San Diego, 1993, Part 1, 1/3-1 - 1/3-34.

12. *Insight II User Guide, Version 2.3.0*, Biosym Technologies, San Diego, 1993, Part 2, 7-1 - 7-52.

13. *Discover User Guide, Version 2.9.0/3.1.0*, Biosym Technologies, San Diego, 1993, Part 1, 1/2-1 - 1/2-119.

14. Meadows, R.P.; Post, C.B.; Gorenstein, D.G.; Luxon, B.A. *MORASS User Guide*, Purdue University, West Lafayette, USA, 1994.

APPENDIX 1

Preliminary results obtained for Lom-AKH-II using the alternative NOESY assignments.

Table A1.1: Alternative chemical shift assignments for Lom-AKH-II in d^6-DMSO.

Residue	Amide	$C_\alpha H$	$C_\beta H$	$C_\beta H'$	Other
pGlu	7.757	4.021	2.208	1.872	2.050 γH 2.110 γH'
Leu	8.034	4.275	1.357	1.357	1.562 γH 0.858 δCH_3 0.822 δCH_3'
Asn	8.115	4.498	2.519	2.370	
Phe	7.984	4.488	3.059	2.800	7.12-7.26 φH's
Ser	8.068	4.290	3.620	3.596	4.98 OH
Ala	7.960	4.239			1.213 CH_3
Gly	8.020	3.562			3.710 $C_\alpha H'$
Trp	7.894	4.429	3.126	2.932	10.880 φNH 7.120 C2H 7.580 C4H 6.965 C5H 7.045 C6H 7.310 C7H

The four protons of the two NH_2 groups were assignable at 7.38;7.06 and 7.44;7.00, respectively. It was impossible to assign these two groups to any specific residue.

Table A1.2: Comparative table of results allowing assessment of the extent of agreement between experimental and theoretical data for the alternative structure of Lom-AKH-II.

Structure	R_1	R_2	Energy (kcal/ mol)	# NOESY Viol.	# H-bonds	D-A dist.	D-H-A angle
A1	0.18(6)	0.04(8)	282.10	1	0	-	-
B1	0.19(3)	0.04(9)	279.22	1	0	-	-
C1	0.22(0)	0.05(4)	280.97	1	1	2.77[1]	132.3
D1	0.20(6)	0.05(0)	282.93	1	1	3.19[2]	125.6
E1	0.21(2)	0.05(4)	280.47	1	0	-	-

In the above D-A dist. refers to the Donor-Acceptor distance and the angle defined is the Donor-Hydrogen-Acceptor angle.
[1]Ser-O-H···O=Ser hydrogen bond
[2]Phe-N-H···O_{sc}=Asn (Asn side chain carbonyl) hydrogen bond.

The above results set was obtained from a restrained MD run carried out at 600K in which eighty structures were collected at 10ps intervals. The family module generated twnty one families in which structures had an *RMS* deviation in aligned positions of less than 0.5. Only five of the families contained structures which obeyed the van der Waals minimum distance of approach rule and therefore only these families were accepted. The sample structures taken from each acceptable family was one in which all non-bonded distances were acceptable. Structure A1, which generated the lowest *R*-factor value, was therefore the basis for all further structure analysis.

Table A1.3: Table illustrating the extent of agreement between the experimental and theoretical data for structure A1.

Peak	Interaction	Distance	NOESY integration		
			Exp.	Theor.	Diff.
A	Gly-NH ··· Gly-αH_1	3.16	0.062	0.108	-0.046
B	Ser-NH ··· Ser-βH_1	4.10	0.048	0.032	0.016
C	Gly-NH ··· Ser-βH_2	3.62	0.075	0.040	0.035
D	Gly-NH ··· Gly-αH_2	3.17	0.100	0.108	-0.008
E	Trp-NH ··· Gly-αH_2	3.33	0.141	0.080	0.061
F	Leu-NH ··· pGlu-αH	3.19	0.168	0.126	0.042
G	Asn-NH ··· Leu-αH	2.60	0.333	0.326	0.007
H	Gly-NH ··· Ala-αH	3.16	0.138	0.101	0.037
	Leu-NH ··· Leu-αH	2.88	0.184	0.176	0.008
I	Ala-NH ··· Ser-αH	2.59	0.285	0.334	0.049
J	Asn-NH ··· Asn-αH	3.08	0.096	0.118	-0.022
	Ser-NH ··· Phe-αH	2.67	0.296	0.294	0.002
K	Ala-NH ··· Phe-αH	2.66	0.301	0.300	0.001
	Phe-NH ··· Asn-αH	3.44	0.101	0.086	0.015
L	Trp-NH ··· Trp-αH	3.01	0.040	0.133	-0.093

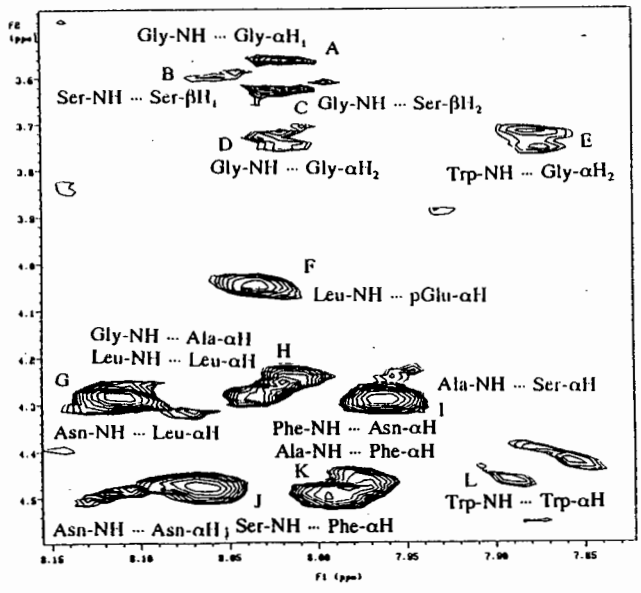


Figure A1.1: Fingerprint region of the NOESY spectrum illustrating the alternative peak assignments for Lom-AKH-II.

Table A1.4: A comparison between the theoretical and experimental torsion angle values for structure A1.

Residue	Torsion Angle, ϕ	
	Structure A1	Exp.
Leu	-149.9	-150.1
Asn	-147.2	-151.3
Phe	-145.1	-147.2
Ser	-137.6	-150.1
Ala	-138.9	-150.1
Gly	177.0	88.4/-168.6
Trp	-147.1	-150.1

Although Table A1.3 indicates good agreement between the thoeretical and experimental NOESY data, the same cannot be said for the ϕ torsion angle data, shown in Table A1.4, associated with the Gly, Ser and Ala residues. To explain this poor agreement, further investigation is required.

The figures given below show the alternative backbone structure of Lom-AKH-II to be relatively linear with no distinct features. There were no consistent hydrogen bonds and thus this structure could also be predicted to be more flexible than either Lom-AKH-I and III. These results would again be consistent with the corresponding activity data discussed in Chapter 7.

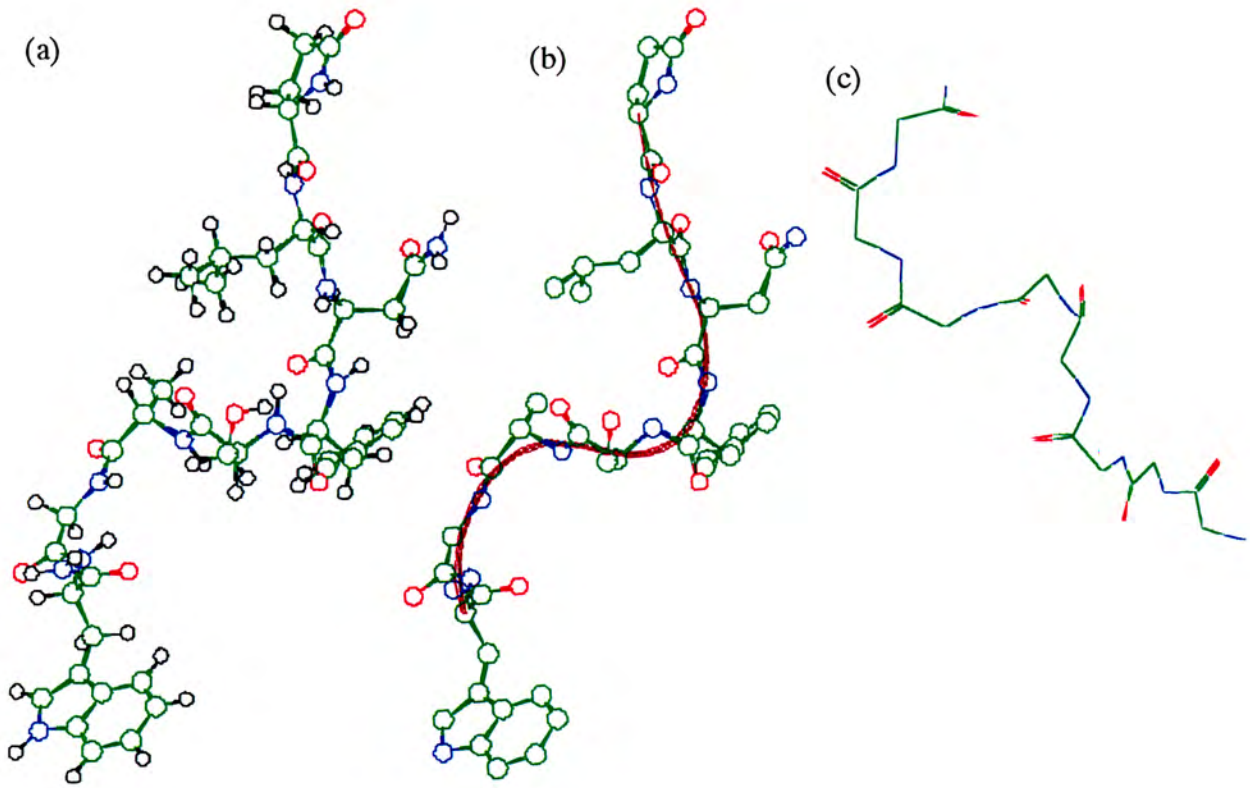


Figure A1.2: Structure A1 (a) with all atoms visible, (b) without hydrogens (brown ribbon is representative of the backbone) and (c) showing the linearity of the backbone structure.

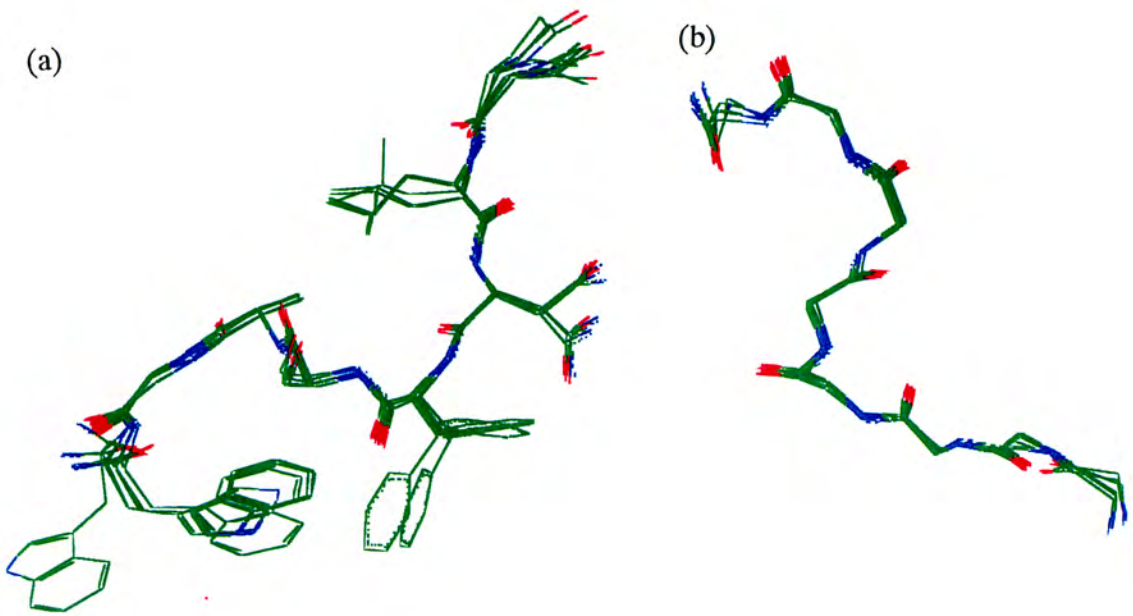
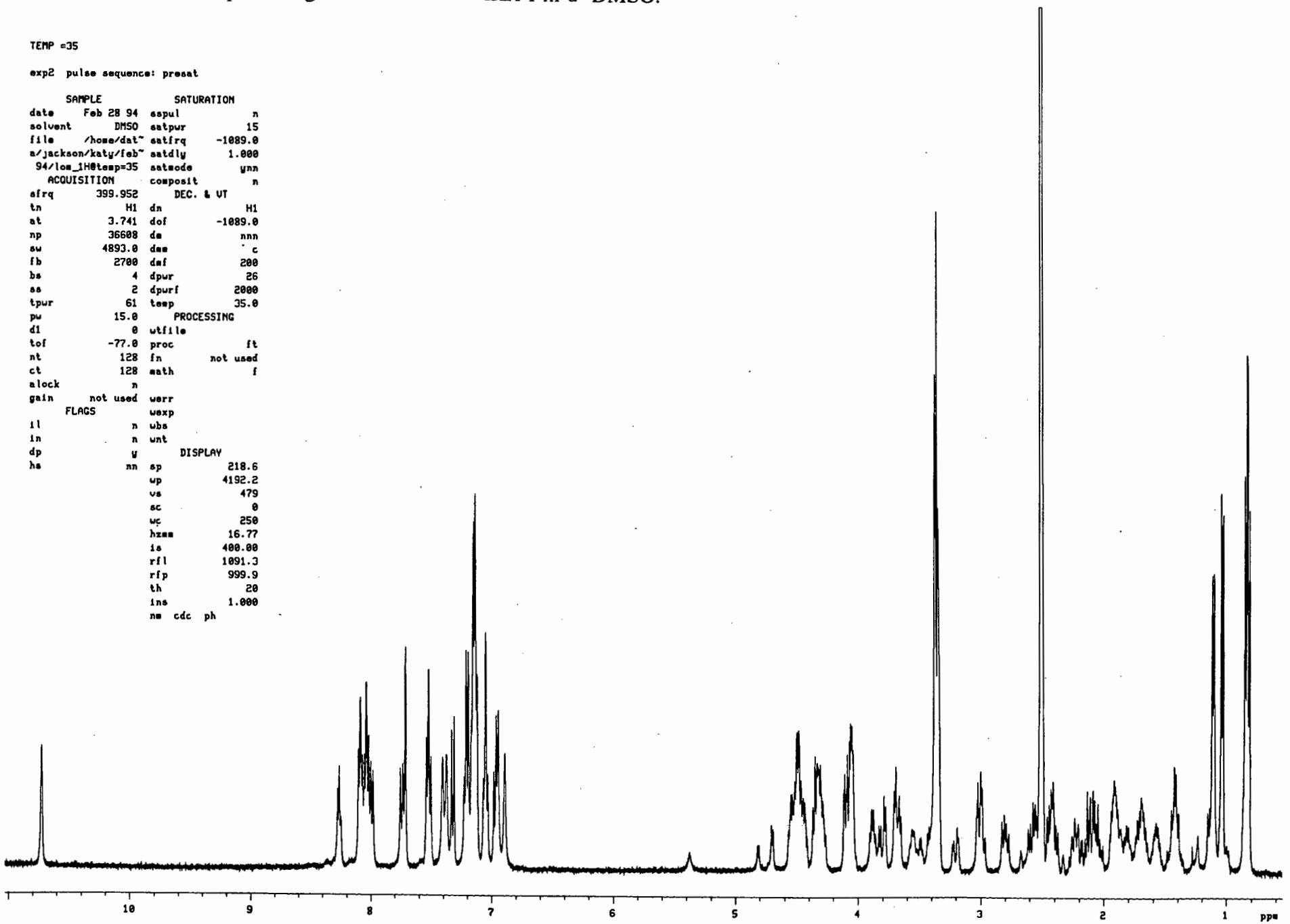


Figure A1.3: Overlay of the family associated with structure A1 - (a) full structures without hydrogens and (b) backbones.

APPENDIX 2

Figure A2.1: Full ^{1}H spectrum generated for Lom-AKH-I in d^{6}-DMSO.



Chemical Shift vs Temperature graph generated for Lom-AKH-I

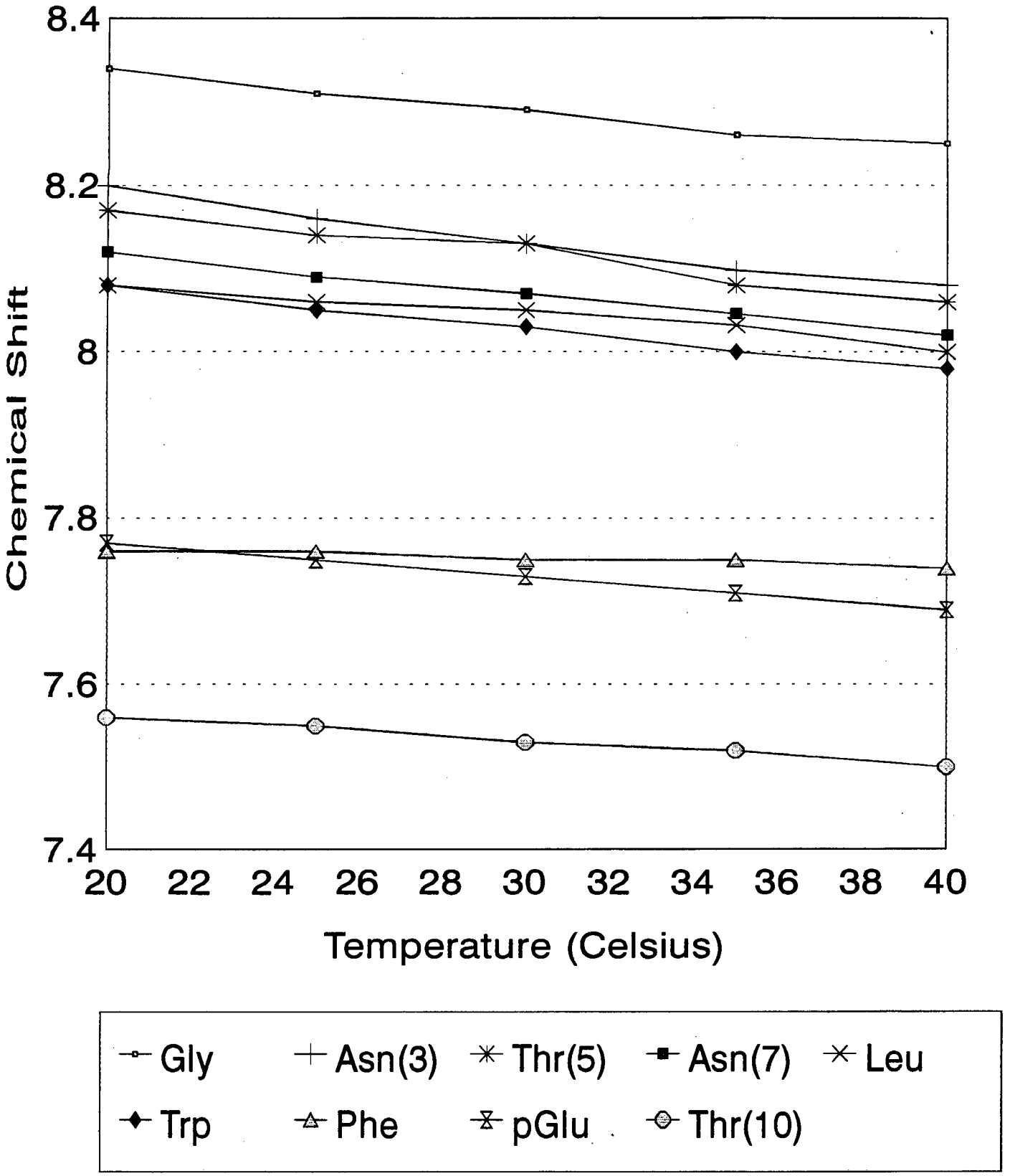


Figure A2.2: Full COSY spectrum generated for Lom-AKH-I in d^6-DMSO.

TEMP =35

expl pulse sequence: relayh

SAMPLE		DEC. & VT	
date	Apr 7 94	dn	H1
solvent	DMSO	dof	-1089.0
file	/home/dat~	dm	yyn
a/jackson/katy/feb~		dmm	c
94/lom/lom_cosy_07~		dmf	200
	-04-94	dpwr	26
ACQUISITION		temp	35.0
sfrq	399.952	PROCESSING	
tn	H1	sb	0.082
at	0.163	sbs	not used
np	2048	wtfile	
sw	6281.4	proc	ft
fb	3500	fn	2048
bs	4	math	f
ss	4		
tpwr	63	werr	
pw	14.5	wexp	
p1	14.5	wbs	
d1	2.000	wnt	
phase	0	2D PROCESSING	
tof	-1089.0	sb1	0.021
nt	16	sbs1	not used
ct	16	wtfile1	
alock	n	proc1	ft
gain	30	fn1	2048
FLAGS		DISPLAY	
il	n	sp	154.3
in	n	wp	4236.7
dp	y	vs	6500
hs	nn	sc	0
2D ACQUISITION		wc	120
sw1	6281.4	hzmm	35.31
ni	256	is	400.00
2D DISPLAY		rf1	2798.2
sp1	25.3	rfp	999.9
wp1	4427.1	th	5
sc2	0	ins	1.000
wc2	120	ai cdc av	
rf11	2798.2		
rfp1	999.9		

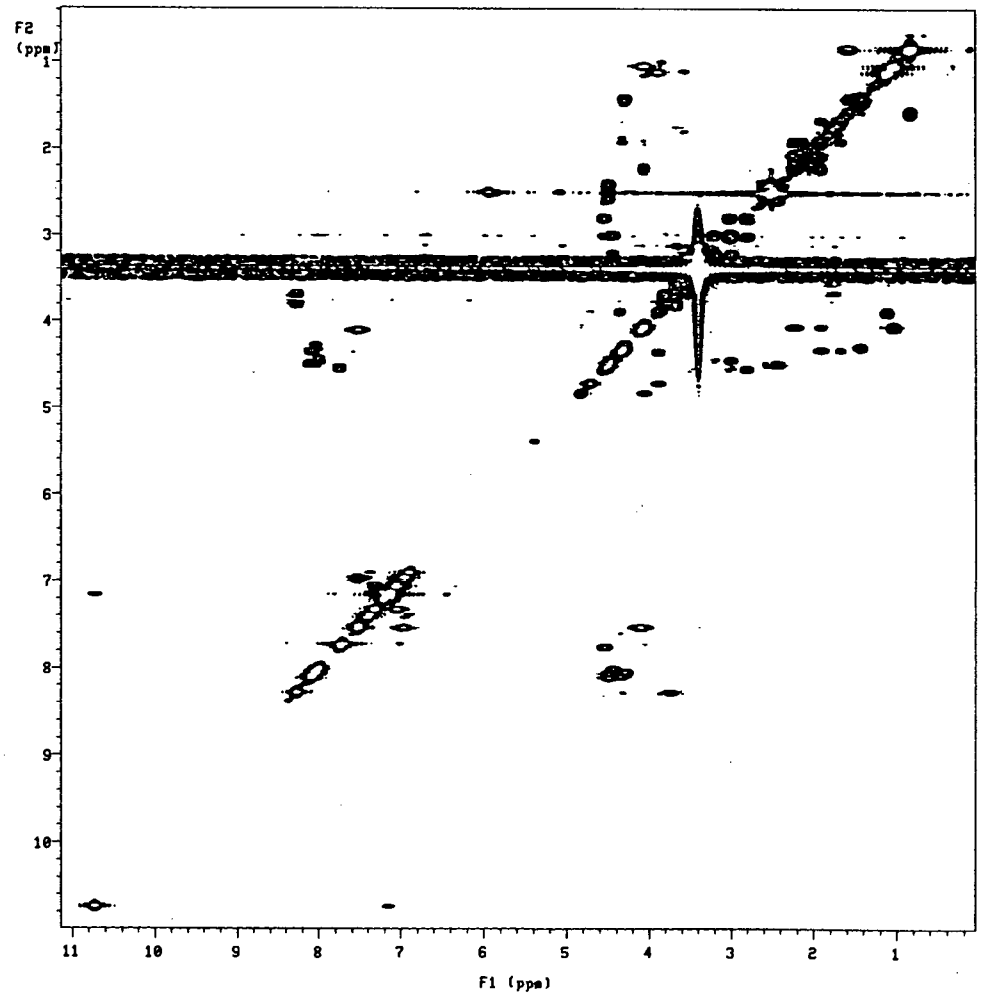


Figure A2.3: Full TOCSY spectrum generated for Lom-AKH-I in d^6-DMSO.

TEMP =35

exp2 pulse sequence: tntocsy

SAMPLE		DEC. & VT		ACQUISITION ARRAYS	
date	Feb 28 94	dn	H1	array	phase
solvent	DMSO	dof	-1089.0	arraydim	512
file	/home/dat~	dm	nnn		
a/jackson/katy/feb~		dmm	c	i	phase
94/lom/lom_tntocsy~		dmf	200	1	1
	_80ms	dpwr	26	2	2
ACQUISITION		temp	35.0		
sfrq	399.952	PROCESSING			
tn	H1	sb	-0.209		
at	0.209	sba	-0.209		
np	2048	gf	0.097		
sw	4893.0	gfa	not used		
fb	2700	wtfile			
bs	4	proc	ft		
ss	16	fn	2048		
pllvl	61	math	f		
pl	17.0				
d1	0	werr			
tof	-1089.0	wexp			
nt	16	wbs			
ct	16	wnt			
alock	n	2D PROCESSING			
gain	48	sb1	0.027		
FLAGS		sbs1	-0.017		
il	n	gf1	0.025		
in	n	gfs1	not used		
dp	y	wtfile1			
hs	n	proc1	ft		
PRESATURATION		fn1	1024		
aspul	n	DISPLAY			
tnsat	undefined	sp	23.8		
dpwr	26	wp	3563.3		
presat	undefined	vs	1447		
SPIN LOCK		sc	6		
tpwr	49	wc	120		
pw	68.0	hzmm	29.69		
window	149.6	is	400.00		
trim	0.002	rfl	2109.6		
mix	0.080	rfp	999.9		
2D ACQUISITION		th	5		
sw1	4893.0	ins	1.000		
ni	256	ai ph			
phase	arrayed				
2D DISPLAY					
sp1	131.8				
wp1	3456.7				
sc2	0				
wc2	120				
rfl1	2132.0				
rfp1	999.9				

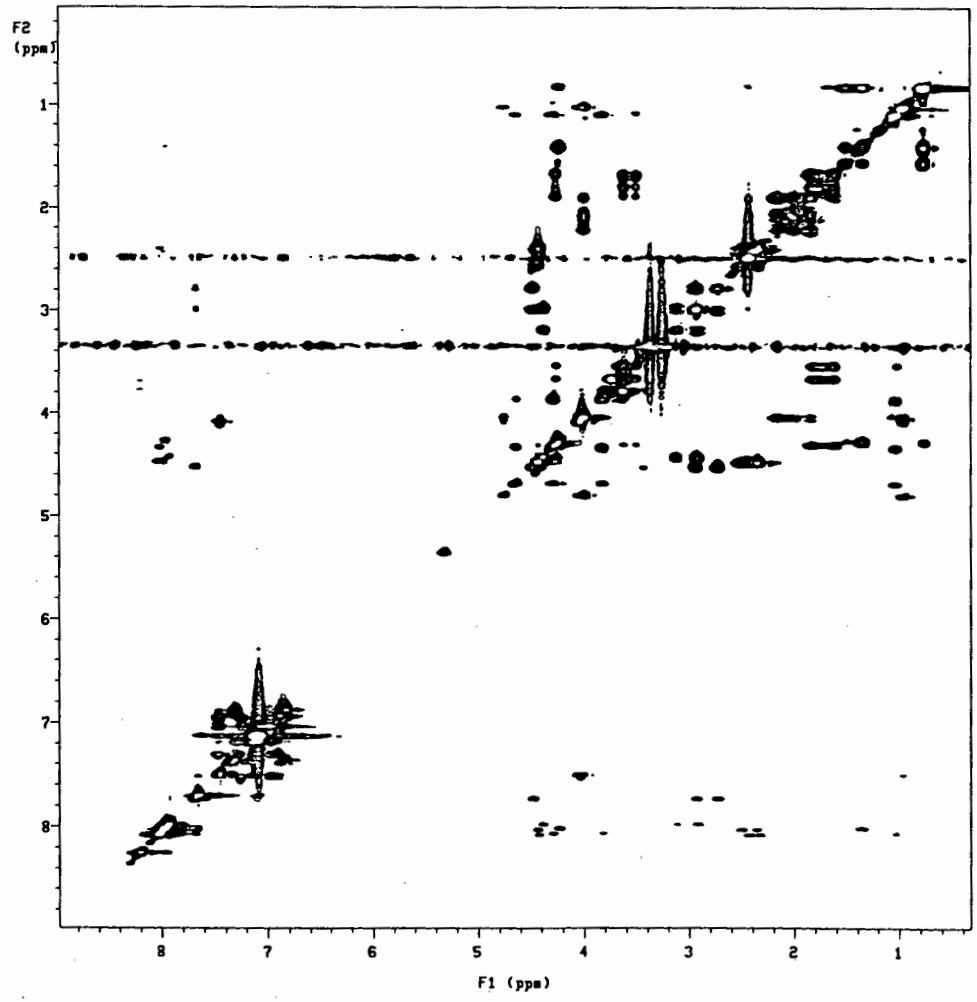


Figure A2.4: Full NOESY spectrum (mix=150ms) generated for Lom-AKH-I in d^6-DMSO.

TEMP =35

expl pulse sequence: tnnoesy

SAMPLE		DEC. & VT		ACQUISITION ARRAYS	
date	Mar 2 94	dn	H1	array	phase
solvent	DMSO	dof	-1089.0	arraydim	512
file	/home/dat~	da	nnn		
a/jackson/katy/feb~		dmm	c	i	phase
94/lomsproc_29_08_~		dmf	200	1	1
	94	dpur	26	2	2
ACQUISITION		temp	35.0		
sfrq	399.952	PROCESSING			
tn	H1	sb	-0.163		
at	0.163	sbs	-0.163		
np	2048	gf	0.075		
sw	6281.4	gfs	not used		
fb	3500	wtfile			
bs	4	proc	ft		
ss	16	fn	2048		
pw	17.0	math	f		
tpwr	61				
d1	1.000	werr			
mix	0.150	wexp			
tof	-1089.0	wbs			
nt	16	wnt			
ct	16	2D PROCESSING			
alock	n	sb1	0.020		
gain	57	sbs1	-0.015		
FLAGS		gf1	0.032		
il	n	gfs1	not used		
in	n	wtfile1			
dp	y	proc1	ft		
hs	nn	fn1	1024		
SATURATION		DISPLAY			
satpul	n	sp	-103.6		
satmode	wwyn	wp	4556.0		
satdly	1.000	vs	150		
satpwr	15	sc	6		
2D ACQUISITION		wc	120		
sw1	6281.4	hzmm	4.66		
ni	256	is	400.00		
phase	arrayed	rf1	2798.2		
2D DISPLAY		rfp	999.9		
sp1	-163.4	th	1		
wp1	4634.2	ins	1.000		
sc2	0	ai ph			
wc2	120				
rf11	2798.2				
rfp1	999.9				

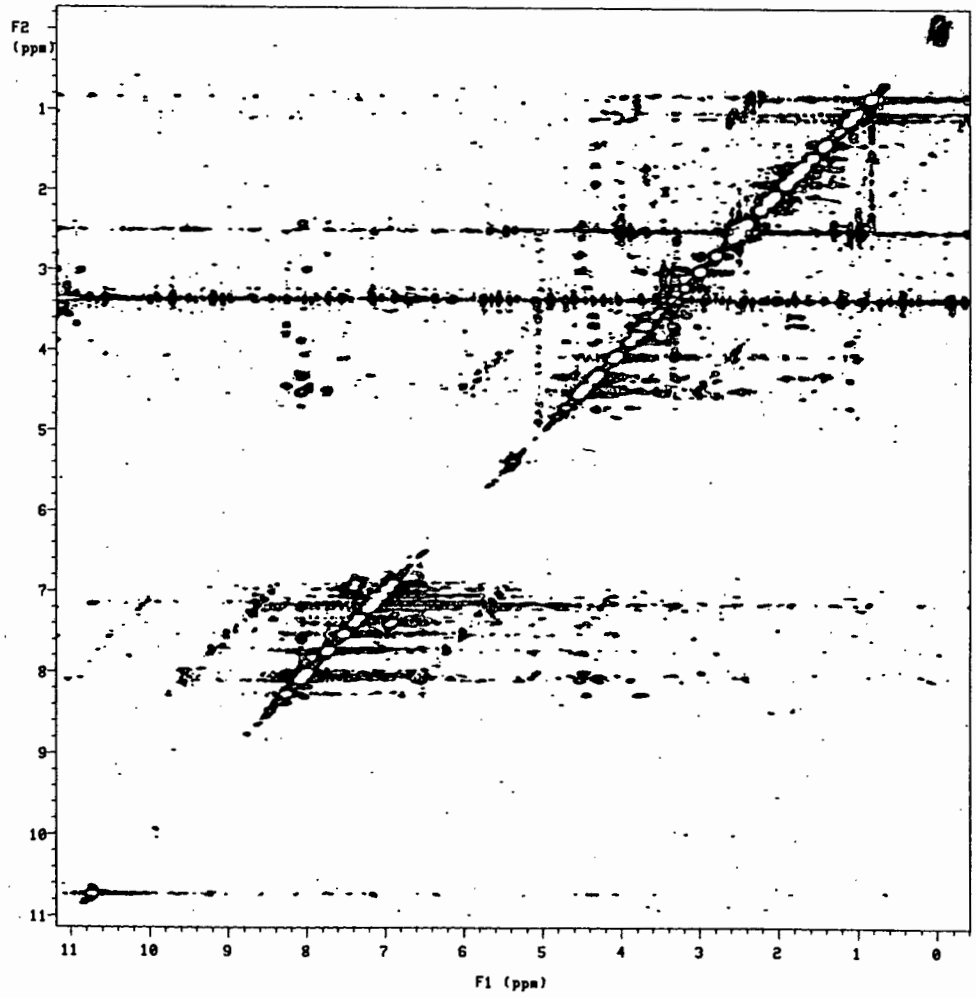


Figure A2.5: Full NOESY spectrum (mix=200ms) generated for Lom-AKH-I in d^6-DMSO.

TEMP =35

exp2 pulse sequence: tnnoesy

SAMPLE		DEC. & VT		ACQUISITION ARRAYS	
date	Mar 2 94	dn	H1	array	phase
solvent	DMSO	dof	-1089.0	arraydim	512
file	/home/dat~	dm	nnn		
a/jackson/katy/feb~		dmm	c	i	phase
94/mix200_proc		dmf	200	1	1
ACQUISITION		dpwr	26	2	2
sfrq	399.952	temp	35.0		
tn	H1	PROCESSING			
at	0.163	sb	-0.163		
np	2048	sbs	-0.163		
sw	6281.4	gf	0.075		
fb	3500	gfs	not used		
bs	4	wtfile			
ss	16	proc	ft		
pw	17.0	fn	2048		
tpwr	61	math	f		
d1	1.000				
mix	0.200	werr			
tof	-1089.0	wexp			
nt	16	wbs			
ct	16	wnt			
alock	n	2D PROCESSING			
gain	57	sb1	-0.021		
FLAGS		sbs1	-0.008		
il	n	wtfile1			
in	n	proc1	ft		
dp	y	fn1	1024		
hs	nn	DISPLAY			
SATURATION		sp	-146.6		
sspul	n	wp	4599.0		
satmode	yyyn	vs	300		
satdly	1.000	sc	0		
satpwr	15	wc	120		
2D ACQUISITION		hzmm	5.01		
sw1	6281.4	is	400.00		
ni	256	rf1	2798.2		
phase	arrayed	rfp	999.9		
2D DISPLAY		th	3		
sp1	-249.5	ins	1.000		
wp1	4671.1	ai ph			
sc2	0				
wc2	120				
rf11	2798.2				
rfp1	999.9				

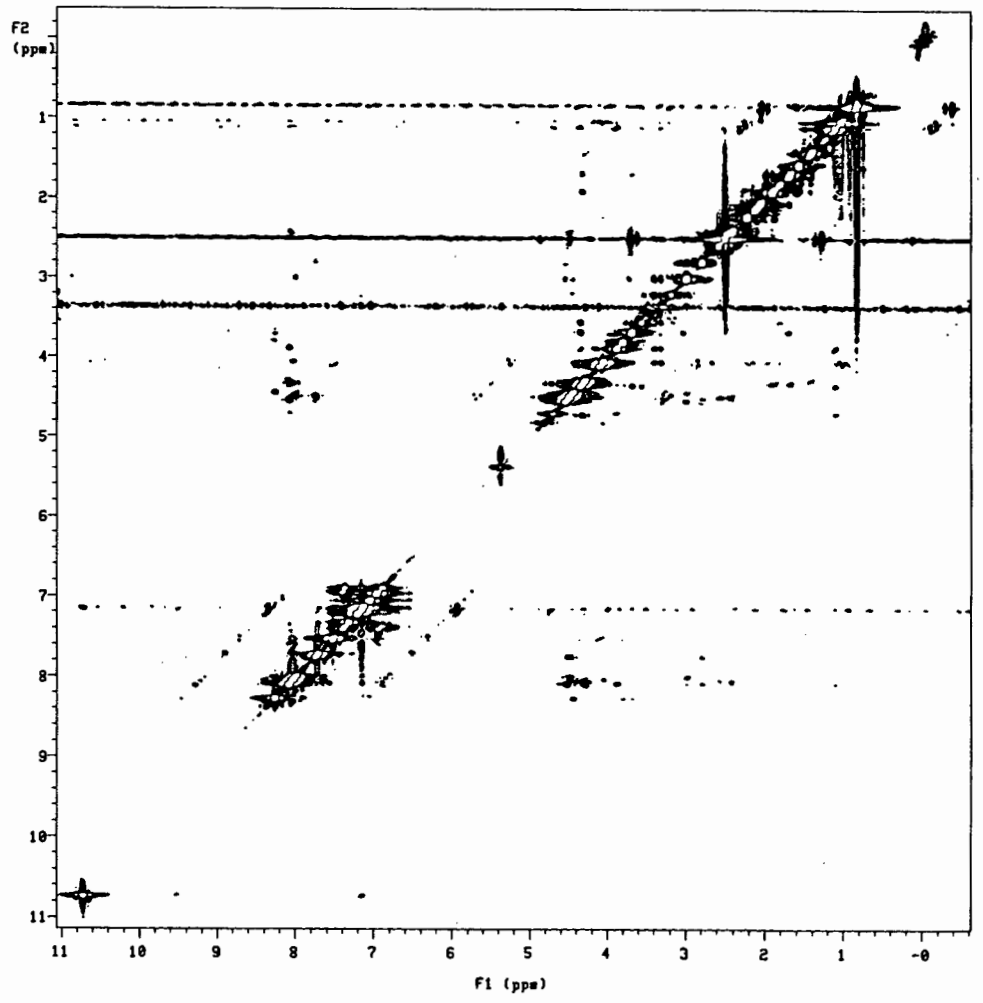


Figure A2.6: Full NOESY spectrum (mix=250ms) generated for Lom-AKH-I in d^6-DMSO.

TEMP =35

exp2 pulse sequence: tnnoesy

SAMPLE		DEC. & VT		ACQUISITION ARRAYS	
date	Mar 3 94	dn	H1	array	phase
solvent	DMSO	dof	-1089.0	arraydim	512
file	/home/dat~	dm	nnn		
a/jackson/katy/feb~		dmm	c	1	phase
	94/proc250	dmf	200	1	1
ACQUISITION		dpwr	26	2	2
sfrq	399.952	temp	35.0		
tn	H1	PROCESSING			
at	0.163	sb	-0.163		
np	2048	sbs	-0.163		
sw	6281.4	gf	0.075		
fb	3500	gfs	not used		
bs	4	wtfile			
ss	16	proc	ft		
pw	17.0	fn	2048		
tpwr	61	math	f		
d1	1.000				
mix	0.250	werr			
tof	-1089.0	wexp			
nt	16	wbs			
ct	16	wnt			
alock	n	2D PROCESSING			
gain	57	sb1	0.021		
FLAGS		sbs1	-0.007		
il	n	gf1	0.040		
in	n	gfs1	not used		
dp	y	wtfile1			
hs	nn	proc1	ft		
SATURATION		fn1	2048		
aspul	n	DISPLAY			
satmode	yyyn	sp	-214.2		
satdly	1.000	wp	4666.5		
satpwr	15	vs	2500		
2D ACQUISITION		sc	0		
sw1	6281.4	wc	120		
ni	256	hzmm	3.99		
phase	arrayed	is	573.11		
2D DISPLAY		rfl	2798.2		
sp1	-379.9	rfp	999.9		
wp1	4863.0	th	6		
sc2	0	ins	1.000		
wc2	120	ai ph			
rfl1	2798.2				
rfp1	999.9				

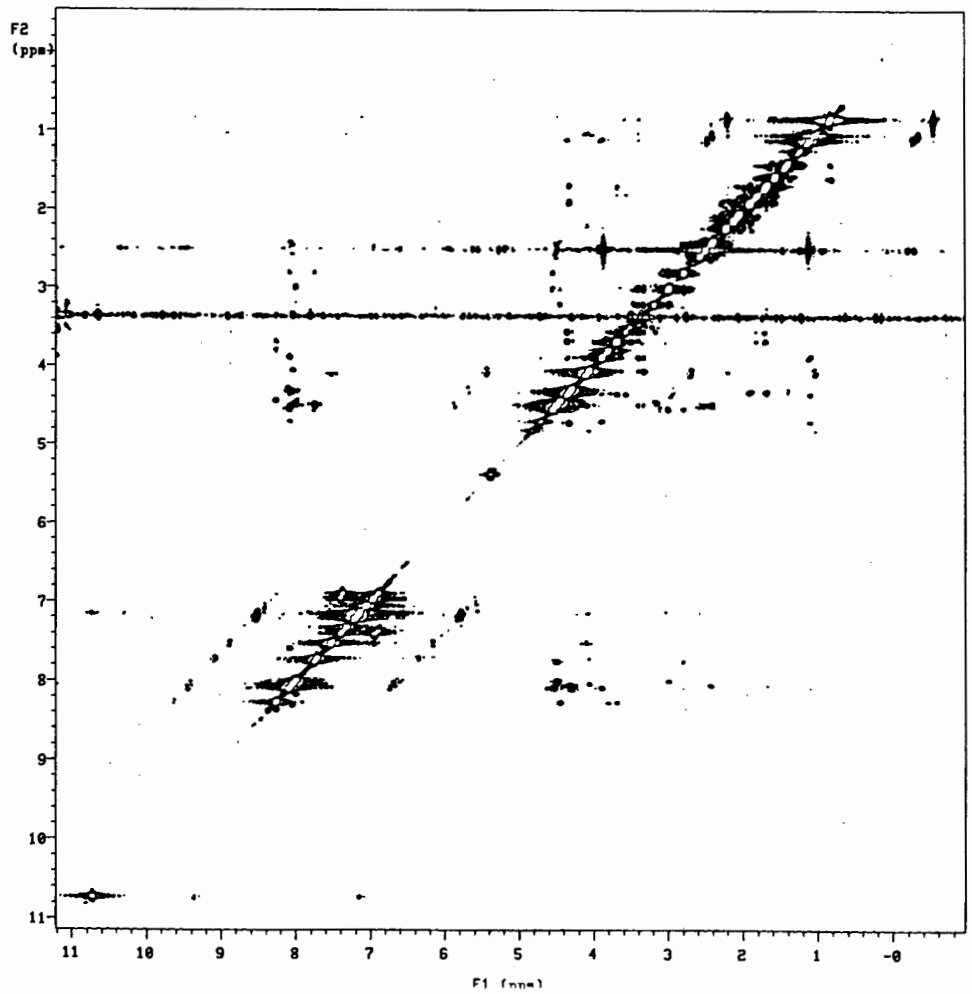


Figure A2.7: Full NOESY spectrum (mix=300ms) generated for Lom-AKH-I in d^6-DMSO.

TEMP =35

expl pulse sequence: tnnoesy

SAMPLE		DEC. & VT		ACQUISITION ARRAYS	
date	Mar 3 94	dn	H1	array	phase
solvent	DMSO	dof	-1089.0	arraydim	512
file	/home/dat~	dm	nnn		
a/jackson/katy/feb~		dmm	c	i	phase
94/mix300_proc		dmf	200	1	1
ACQUISITION		dpwr	26	2	2
sfrq	399.952	temp	35.0		
tn	H1	PROCESSING			
at	0.163	sb	-0.163		
np	2048	sbs	-0.163		
sw	6281.4	gf	0.075		
fb	3500	gfs	not used		
bs	4	wtfile			
ss	16	proc	ft		
pw	17.0	fn	2048		
tpwr	61	math	f		
d1	1.000				
mix	0.300	werr			
tof	-1089.0	wexp			
nt	16	wbs			
ct	16	wnt			
alock	n	2D PROCESSING			
gain	57	sb1	0.020		
FLAGS		sbs1	-0.016		
il	n	wtfile1			
in	n	proc1	ft		
dp	y	fn1	2048		
hs	nn	DISPLAY			
SATURATION		sp	-85.2		
sspul	n	wp	4549.9		
satmode	yyyn	vs	1000		
satdly	1.000	sc	6		
satpwr	15	wc	120		
2D ACQUISITION		hzmm	37.92		
sw1	6281.4	is	777.75		
ni	256	rf1	2798.2		
phase	arrayed	rfp	999.9		
2D DISPLAY		th	5		
sp1	-36.1	ins	1.000		
wp1	4513.0	ai	ph		
sc2	0				
wc2	120				
rf11	2798.2				
rfp1	999.9				

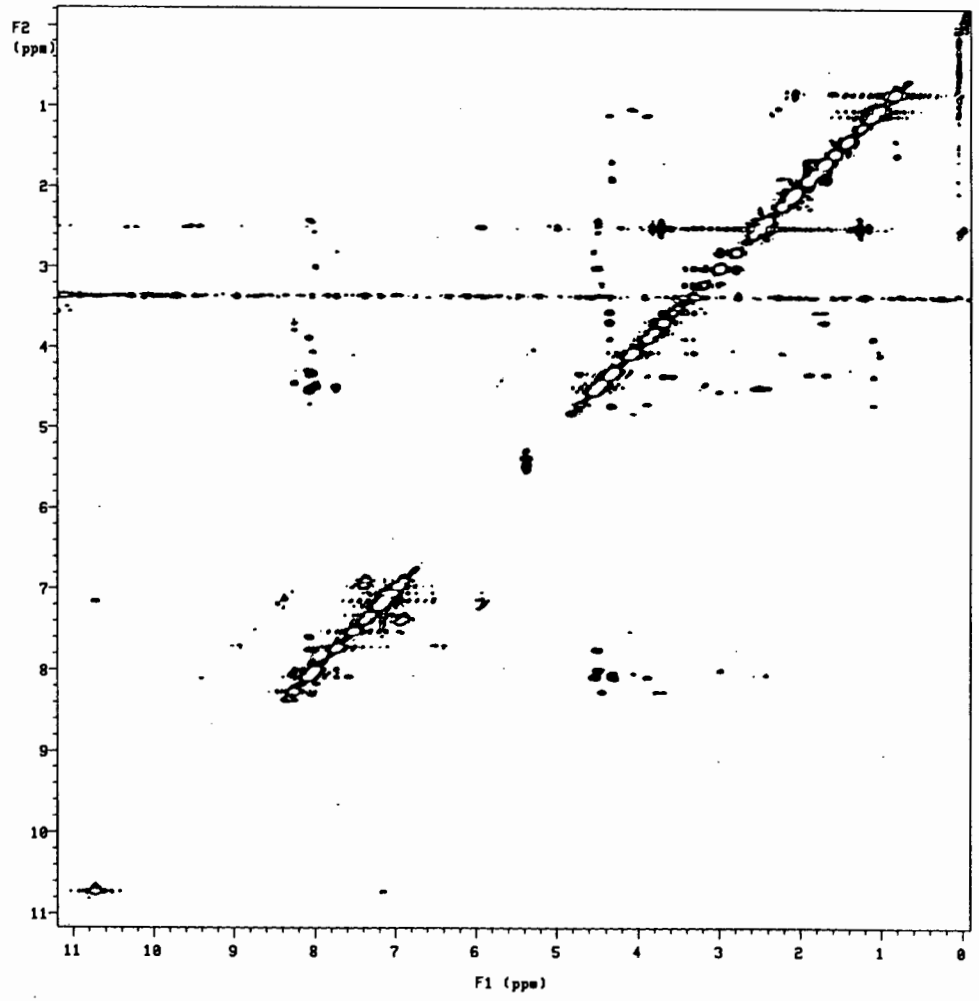


Figure A2.8: Full Homonuclear 2DJ-Resolved spectrum generated for Lom-AKH-I in d^6- DMSO.

exp1 pulse sequence: hom2dj

SAMPLE		DEC. & UT	
date	Jun 1 94	dn	H1
solvent	DMSO	dof	-1076.1
file	/home/dat~	dm	ynn
a/jackson/katy/feb~		dmm	c
94/lom_hom2dj_02-0~		dmf	200
	6-94	dpwr	25
ACQUISITION		temp	35.0
sfrq	399.952	PROCESSING	
tn	H1	sb	0.084
at	0.172	sbs	not used
np	2048	wtfile	
sw	5946.2	proc	ft
fb	3300	fn	8192
ss	16	math	f
tpwr	60		
pw	19.5	werr	
p1	39.0	wexp	
d1	1.000	wbs	
tof	8.4	wnt	
nt	16	2D PROCESSING	
ct	16	sb1	0.496
alock	n	sbs1	-0.054
gain	24	wtfile1	
FLAGS		proc1	ft
il	n	fn1	256
in	n	DISPLAY	
dp	y	sp	30.0
hs	nn	wp	4492.7
2D ACQUISITION		vs	100000
sw1	50.0	sc	5
ni	512	wc	120
2D DISPLAY		hzmm	37.44
sp1	-20.3	is	400.00
wp1	40.6	rf1	3837.4
sc2	0	rfp	3304.0
wc2	120	th	7
rfl1	25.0	ins	1.000
rfp1	0	ai av	

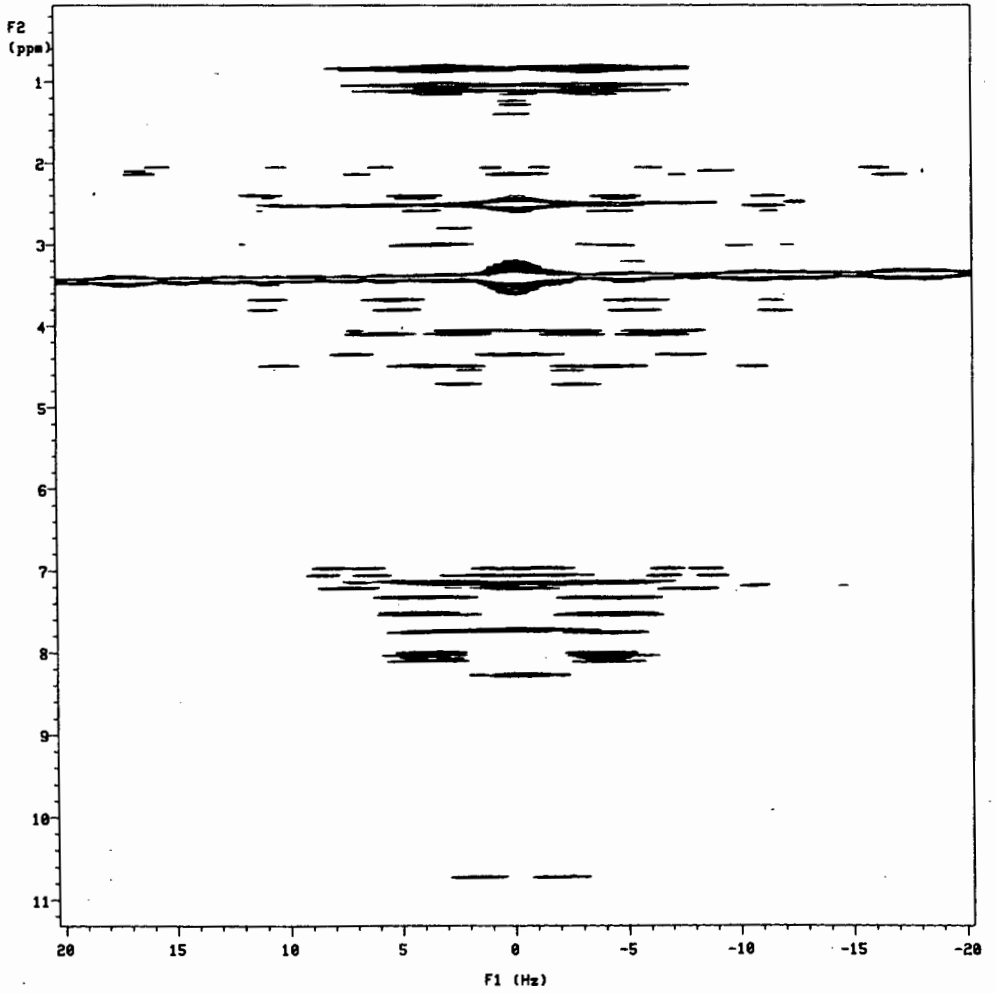


Figure A2.9: Full ^{1}H spectrum generated for Lom-AKH-II in d^{6}-DMSO.

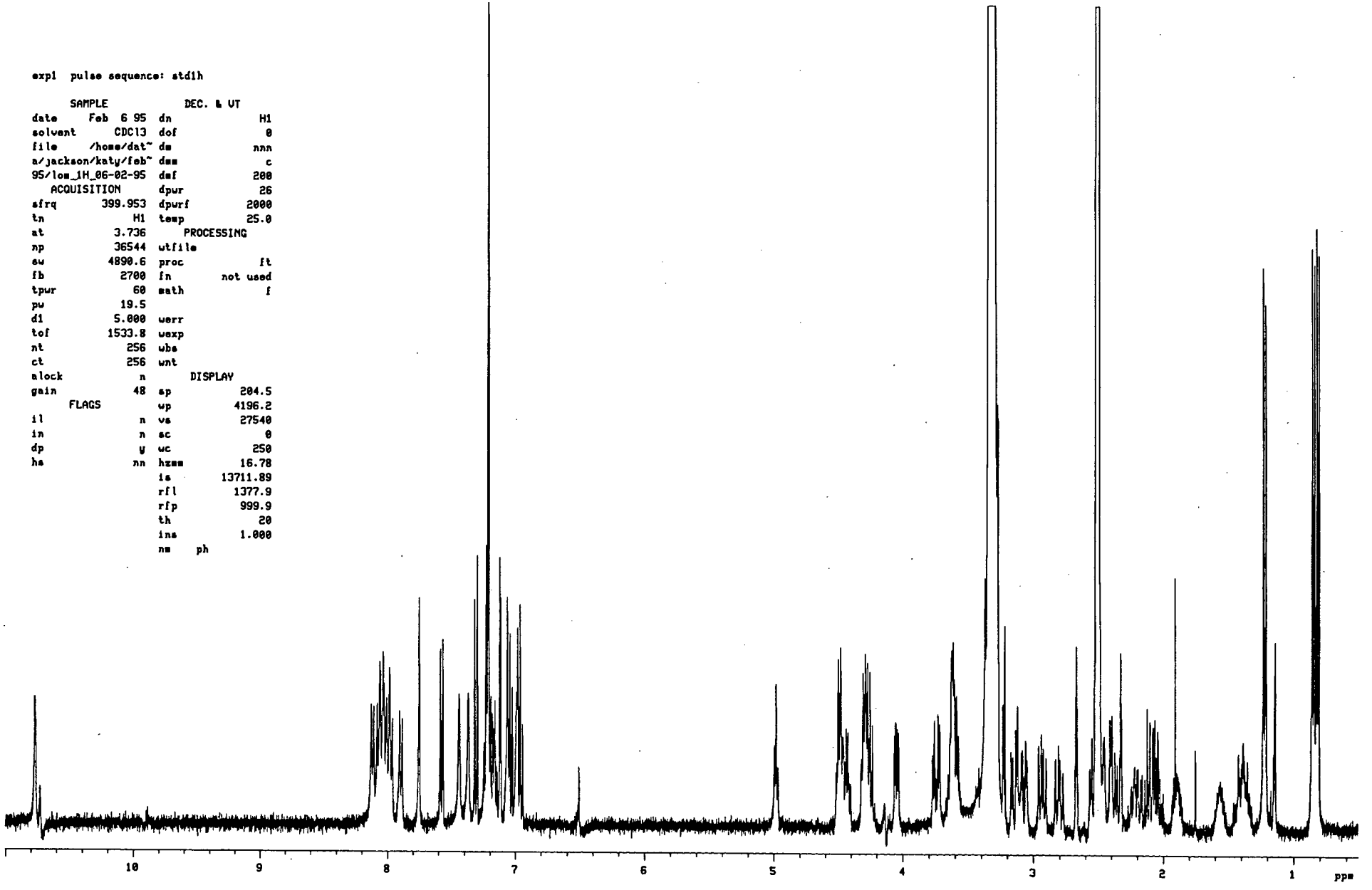


Figure A2.10: Temperature arrayed ^{1}H spectra generated for Lom-AKH-II in d^{6}-DMSO.

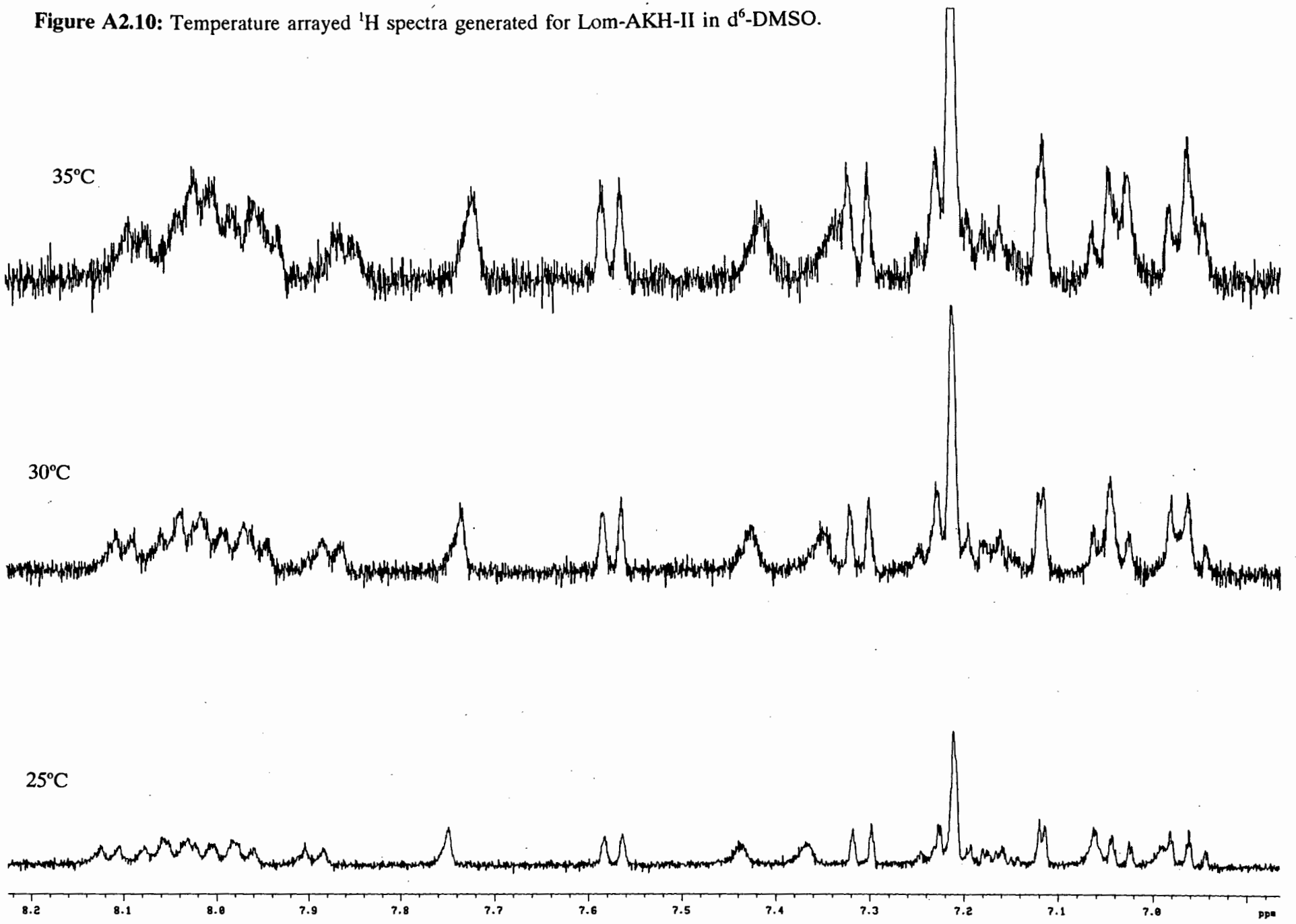


Figure A2.11: Full COSY spectrum generated for Lom-AKH-II in d^6-DMSO.

```
expl  pulse sequence: relayh

      SAMPLE              DEC. & VT
date      Feb  1 95  dfrq       not used
solvent       CDCl3  dn              H1
file      /home/dat~ dpur            26
a/jackson/katy/feb~  dof             0
95/lom_cosy_02_02-~  dm            nnn
                 95  dmm             c
   ACQUISITION       dmf           200
sfrq        399.953  dseq      undefined
tn               H1  dres      undefined
at            0.209  homo            y
np             2048  temp         25.0
sw           4890.6      PROCESSING
fb             2700  sb           0.105
ss                2  sba        not used
tpwr             60  wtfile
pw             19.5  proc            ft
p1             19.5  fn            2048
d1            1.000  math             f
tof          1533.8
nt               32  werr
ct               32  wexp
tau               0  wbs
relay             0  wnt
alock             n    2D PROCESSING
gain             45  sb1          0.052
      FLAGS          sbs1       not used
il                n  wtfile1
in                n  proc1           ft
dp                y  fn1           1024
hs               nn
  2D ACQUISITION
sw1          4890.6
ni1             512
phase             0
      DISPLAY
sp             195.6
wp            4297.8
vs            47210
sc               66
wc              120
hzcm          35.81
is           400.00
rfl          1377.9
rfp           999.9
th                9
ins           1.000
ai  cdc  av
    2D DISPLAY
sp1            157.9
wp1           4345.0
sc2               0
wc2             120
rfl1         1377.9
rfp1          999.9
```

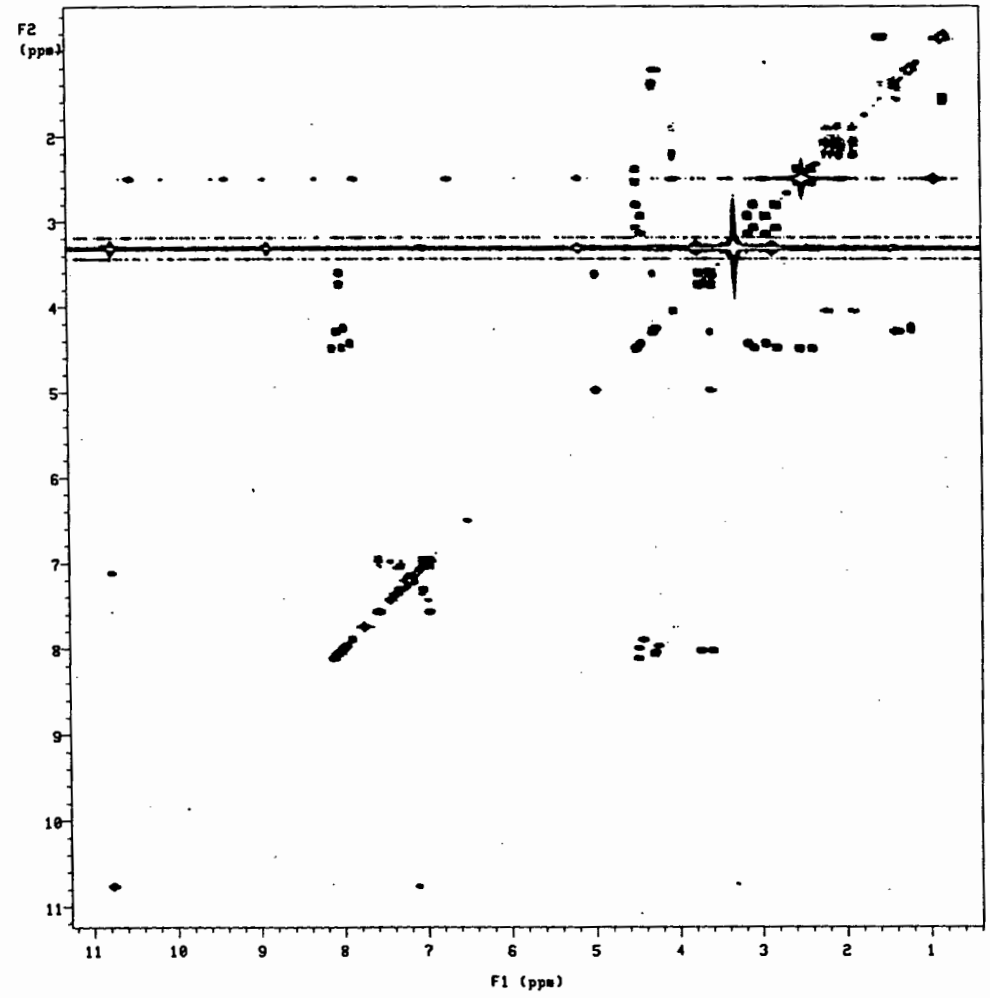


Figure A2.12: Full TOCSY spectrum generated for Lom-AKH-II in d^6-DMSO.

lom2 tocsy

expl pulse sequence: tntocsy

SAMPLE		DEC. & VT		ACQUISITION ARRAYS	
date	Feb 3 95	dn	H1	array	phase
solvent	CDCl3	dof	793.3	arraydim	1024
file	/home/dat~	dm	nnn		
a/jackson/katy/feb~		dmm	c	i	phase
95/protoc_02_06		dmf	200	1	1
ACQUISITION		dpur	5	2	2
sfrq	399.952	temp	25.0		
tn	H1	PROCESSING			
at	0.210	gf	0.097		
np	2048	gfs	not used		
sw	4865.6	wtfile			
fb	2700	proc	ft		
bs	16	fn	2048		
ss	128	math	f		
pllvl	60				
pl	19.5	werr			
dl	0	wexp			
tof	793.3	wba			
nt	32	wnt			
ct	32	2D PROCESSING			
alock	n	sb1	0.018		
gain	60	sbs1	0.002		
FLAGS		gf1	0.038		
il	y	gfs1	not used		
in	n	wtfile1			
dp	y	proc1	ft		
hs	n	fn1	1024		
PRESATURATION		DISPLAY			
sspul	n	sp	203.0		
tnsat	undefined	wp	3262.8		
dpur	5	vs	300		
presat	undefined	sc	6		
SPIN LOCK		wc	110		
tpwr	49	hzmm	44.23		
pw	68.0	is	400.00		
windou	149.6	rfl	4352.5		
trim	0.002	rfp	3247.6		
mix	0.080	th	3		
2D ACQUISITION		ins	1.000		
sw1	4865.6	ai	ph		
ni	512				
phase	arrayed				
2D DISPLAY					
sp1	181.1				
wp1	3218.3				
sc2	0				
wc2	110				
rfl1	4351.9				
rfp1	3247.6				

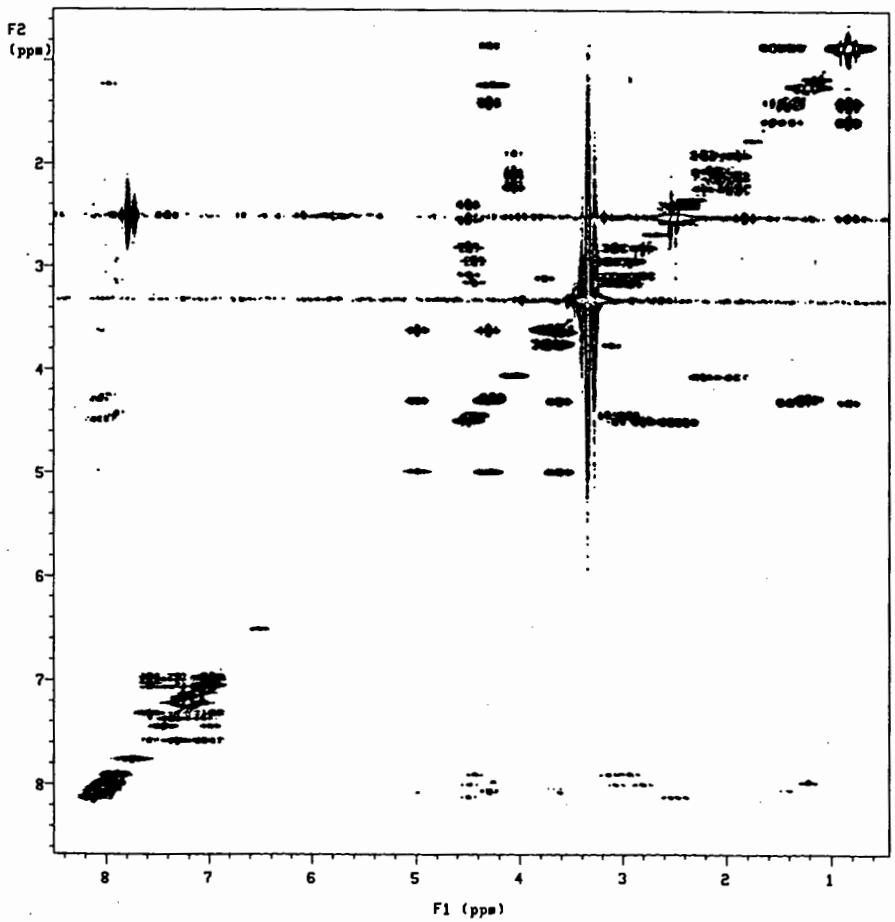


Figure A2.13: Full TOCSY spectrum (positive peaks only) generated for Lom-AKH-II in d^6-DMSO.

lom2 tocsy

expl pulse sequence: tntocsy

SAMPLE		DEC. & VT		ACQUISITION ARRAYS	
date	Feb 3 95	dn	H1	array	phase
solvent	CDCl3	dof	793.3	arraydim	1024
file	/home/dat~	dm	nnn		
a/jackson/katy/feb~		dmm	c	1	phase
95/protoc_02_06		dmf	200	1	1
ACQUISITION		dpur	5	2	2
sfrq	399.952	temp	25.0		
tn	H1	PROCESSING			
at	0.210	gf	0.097		
np	2048	gfs	not used		
sw	4865.6	wtfile			
fb	2700	proc	ft		
bs	16	fn	2048		
ss	128	math	f		
pllvl	60				
pl	19.5	werr			
d1	0	wexp			
tof	793.3	wbs			
nt	32	wnt			
ct	32	2D PROCESSING			
alock	n	sb1	0.018		
gain	60	sbs1	0.002		
FLAGS		gf1	0.038		
il	y	gfs1	not used		
in	n	wtfile1			
dp	y	proc1	ft		
hs	n	fn1	1024		
PRESATURATION		DISPLAY			
sspul	n	sp	193.5		
tnsat	undefined	wp	3186.7		
dpur	5	vs	258		
presat	undefined	sc	66		
SPIN LOCK		wc	110		
tpwr	49	hzmm	28.97		
pw	68.0	is	400.00		
window	149.6	rfl	4352.5		
trim	0.002	rfp	3247.6		
mix	0.080	th	3		
2D ACQUISITION		ins	1.000		
sw1	4865.6	ai	ph		
ni	512				
phase	arrayed				
2D DISPLAY					
sp1	289.7				
wp1	3180.2				
sc2	0				
wc2	110				
rfl1	4351.9				
rfp1	3247.6				

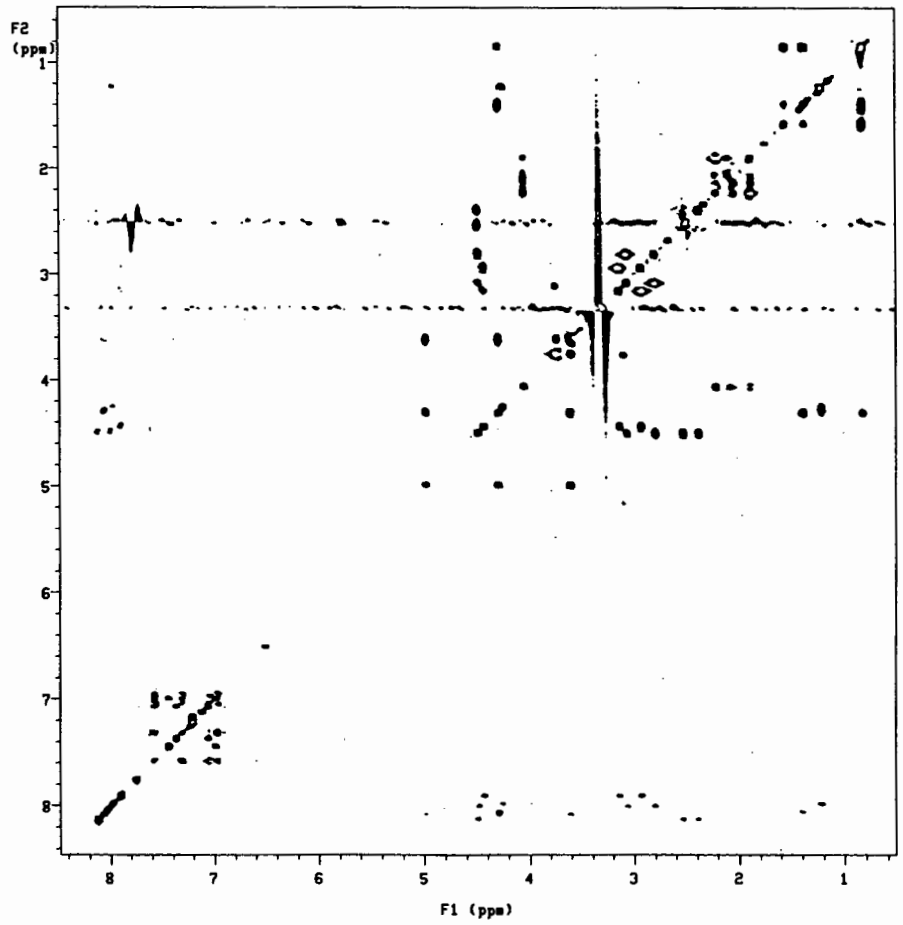


Figure A2.14: Full NOESY spectrum (mix=150ms) generated for Lom-AKH-II in d^6-DMSO.

noesy 150ms lom2

exp2 pulse sequence: tnnoesy

SAMPLE		DEC. & VT		ACQUISITION ARRAYS	
date	Feb 3 95	dn	H1	array	phase
solvent	CDCl3	dof	793.3	arraydim	1024
file	/home/dat~	dm	nnn		
a/jackson/katy/pro~		dmm	c	i	phase
	c95/lomii150	dmf	200	1	1
ACQUISITION		dpwr	5	2	2
sfrq	399.952	temp	25.0		
tn	H1	PROCESSING			
at	0.210	sb	-0.210		
np	2048	sbs	-0.210		
sw	4865.6	gf	0.097		
fb	2700	gfs	not used		
bs	16	wtfile			
ss	128	proc	ft		
pl	19.5	fn	2048		
pw	19.5	math	f		
tpwr	60				
d1	0	werr			
mix	0.150	wexp			
tof	793.3	wbs			
nt	32	wnt			
ct	32	2D PROCESSING			
alock	n	sb1	0.052		
gain	60	sbs1	-0.019		
FLAGS		gf1	0.066		
il	y	gfs1	not used		
in	n	wtfile1			
dp	y	proc1	ft		
hs	n	fn1	1024		
SATURATION		DISPLAY			
sspul	n	sp	-1101.6		
satmode	yyyn	wp	4865.6		
satdly	1.000	vs	1700		
satpwr	0	sc	0		
2D ACQUISITION		wc	120		
sw1	4865.6	hzmm	4.64		
ni	512	is	400.00		
phase	arrayed	rf1	4205.1		
2D DISPLAY		rfp	3103.4		
sp1	-1104.7	th	5		
wp1	4865.6	ins	1.000		
sc2	0	ai cdc ph			
wc2	120				
rf11	4208.1				
rfp1	3103.4				

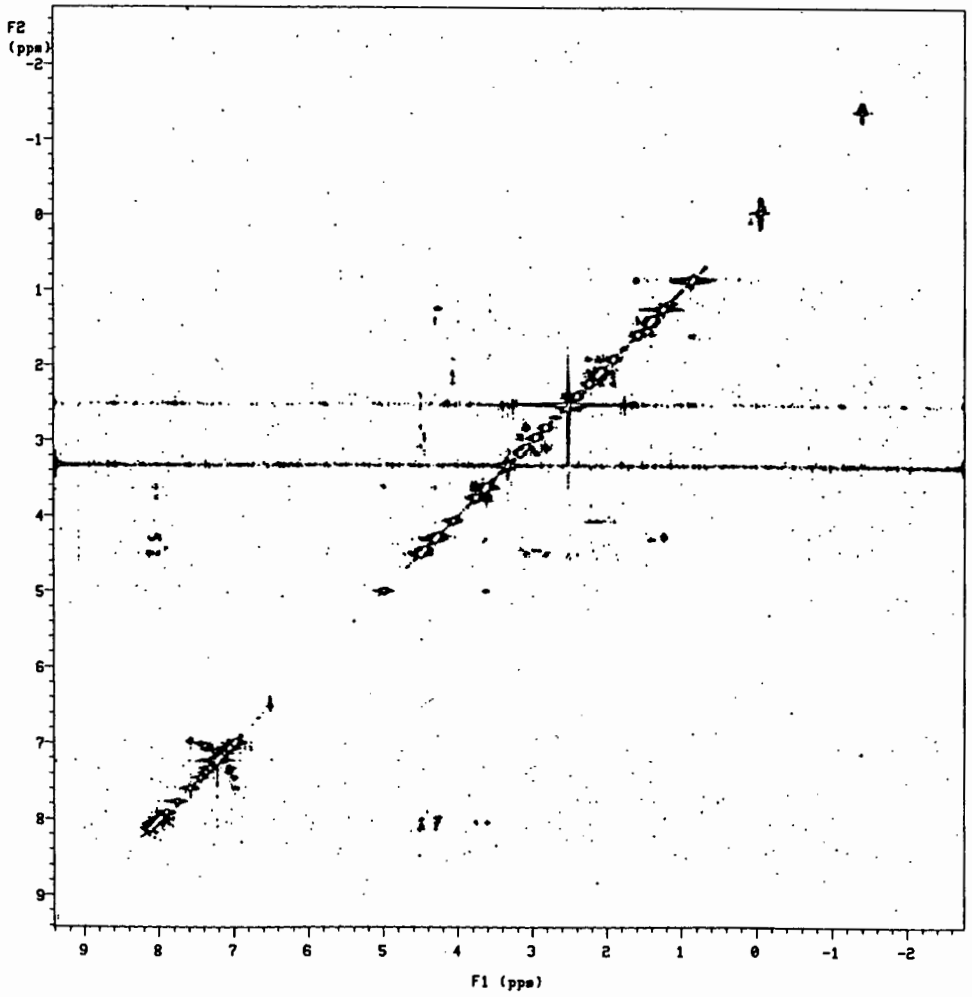


Figure A2.15: Full NOESY spectrum (mix=200ms) generated for Lom-AKH-II in d^6-DMSO.

lom2 noesy 250ms

exp2 pulse sequence: tnnoesy

SAMPLE		DEC. & VT		ACQUISITION ARRAYS	
date	Feb 17 95	dn	H1	array	phase
solvent	CDCl3	dof	793.3	arraydim	1024
file	/home/dat~	dm	nnn		
a/jackson/katy/feb~		dmm	c	i	phase
95/noesy200_26_4_2		dmf	200	1	1
ACQUISITION		dpwr	5	2	2
sfrq	399.952	temp	25.0		
tn	H1	PROCESSING			
at	0.210	sb	-0.210		
np	2048	sbs	-0.210		
sw	4865.6	gf	0.097		
fb	2700	gfs	not used		
ss	128	wtfile			
pl	19.5	proc	ft		
pw	19.5	fn	2048		
tpwr	60	math	f		
d1	0				
mix	0.200	werr			
tof	793.3	wexp			
nt	64	wbs			
ct	64	wnt			
alock	n	2D PROCESSING			
gain	60	sb1	0.053		
FLAGS		sbs1	-0.010		
il	y	gf1	0.049		
in	n	gfs1	not used		
dp	y	wtfile1			
hs	n	proc1	ft		
SATURATION		fn1	2048		
sspul	n	DISPLAY			
satmode	yyyn	sp	94.5		
satdly	1.000	wp	3324.6		
satpwr	0	vs	4500		
2D ACQUISITION		sc	6		
sw1	4865.6	wc	120		
ni	512	hzmm	3.80		
phase	arrayed	is	400.00		
2D DISPLAY		rf1	4349.6		
sp1	108.2	rfp	3245.6		
wp1	3324.6	th	6		
sc2	0	ins	1.000		
wc2	110	ai ph			
rfl1	4350.3				
rfp1	3245.6				

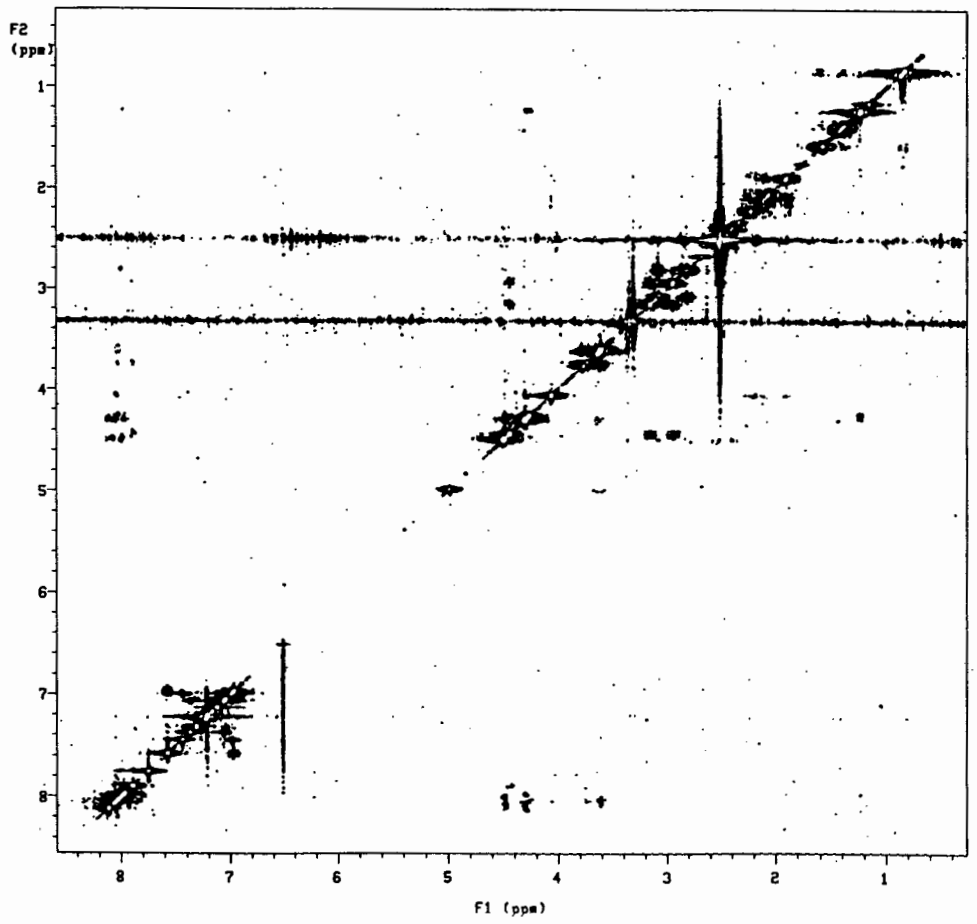


Figure A2.16: Full NOESY spectrum (mix=250ms) generated for Lom-AKH-II in d^6-DMSO.

lom2 noesy 250ms

exp2 pulse sequence: tnnoesy

SAMPLE		DEC. & UT		ACQUISITION ARRAYS	
date	Feb 3 95	dn	H1	array	phase
solvent	CDCl3	dof	793.3	arraydim	1024
file	/home/dat~	dm	nnn		
a/jackson/katy/feb~		dmm	c	i	phase
	95/lom1i250	dmf	200	1	1
ACQUISITION		dpwr	5	2	2
sfrq	399.952	tesp	25.0		
tn	H1	PROCESSING			
at	0.210	sb	-0.210		
np	2048	sbs	-0.210		
sw	4865.6	gf	0.097		
fb	2700	gfs	not used		
bs	16	wtfile			
ss	128	proc	ft		
p1	19.5	fn	2048		
pw	19.5	math	f		
tpwr	60				
d1	0	werr			
mix	0.250	wexp			
tof	793.3	wbs			
nt	32	wnt			
ct	32	2D PROCESSING			
alock	n	sb1	0.009		
gain	60	sbs1	-0.002		
FLAGS		gf1	0.013		
il	y	gfs1	not used		
in	n	wtfile1			
dp	y	proc1	ft		
hs	n	fn1	1024		
SATURATION		DISPLAY			
satpul	n	sp	-569.0		
satmode	yyyn	wp	4865.6		
satdly	1.000	vs	550		
satpwr	0	sc	0		
2D ACQUISITION		wc	120		
sw1	4865.6	hzmm	5.75		
ni	512	is	400.00		
phase	arrayed	rf1	2432.8		
2D DISPLAY		rfp	1863.8		
sp1	-569.0	th	2		
wp1	4865.6	ins	1.000		
sc2	0	ai cdc ph			
wc2	120				
rf11	2432.8				
rfp1	1863.8				

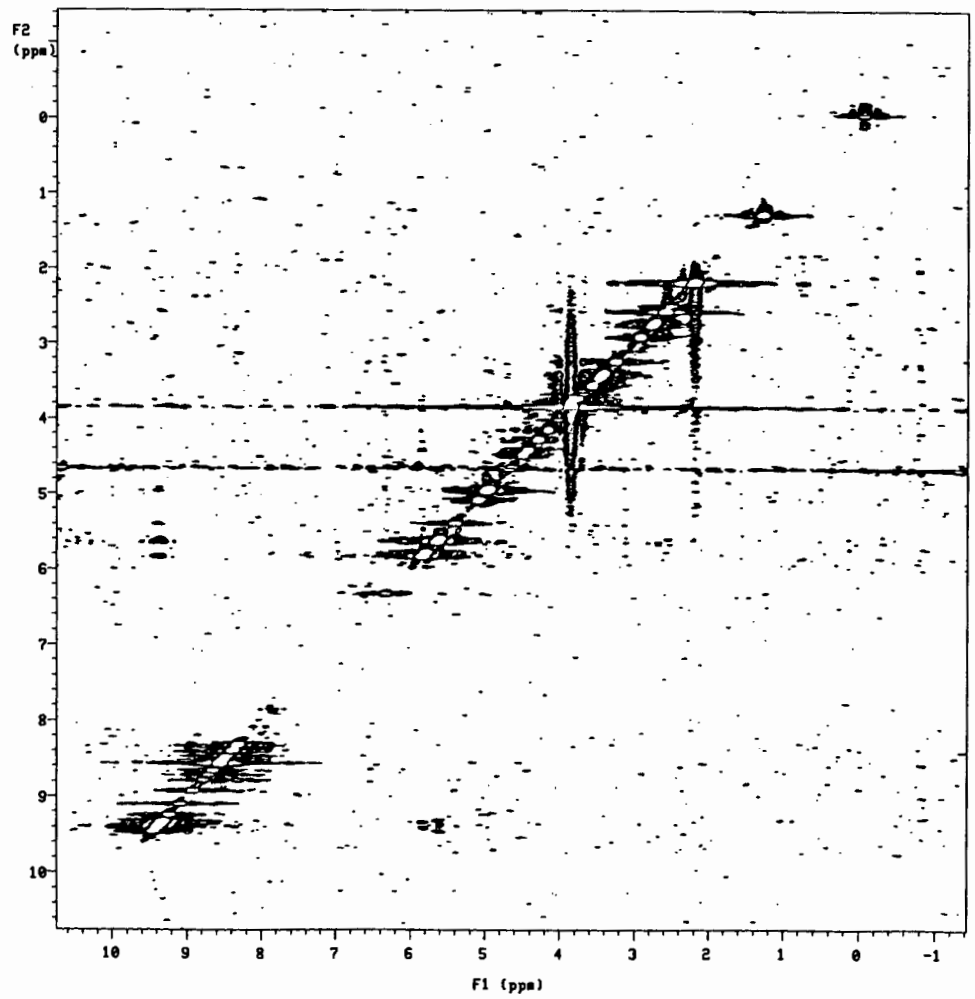


Figure A2.17: Full Homonuclear 2DJ-Resolved spectrum generated for Lom-AKH-II in d^6- DMSO.

expl pulse sequence: hom2dj

	SAMPLE		DEC. & VT
date	Feb 17 95	dfrq	not used
solvent	CDCl3	dn	H1
file	/home/dat~	dpwr	26
a/jackson/katy/feb~		dof	793.7
	95/lom2_hom2dj	dm	yyn
	ACQUISITION	dmm	c
sfrq	399.953	dmf	200
tn	H1	dseq	undefined
at	0.209	dres	undefined
np	2048	homo	y
sw	4890.6	temp	25.0
fb	2700		PROCESSING
ss	32	sb	0.105
tpwr	60	sbs	not used
pw	19.5	wtfile	
p1	39.0	proc	ft
d1	1.000	fn	4096
tof	1533.8	math	f
nt	64		
ct	64	werr	
alock	n	wexp	
gain	48	wbs	
	FLAGS	wnt	
il	n		2D PROCESSING
in	n	sb1	1.263
dp	y	sbs1	not used
hs	nn	wtfile1	
	2D ACQUISITION	proc1	ft
sw1	50.0	fn1	128
ni	256		
phase	0		
	DISPLAY		
sp	159.5		
wp	4281.3		
vs	3200		
sc	0		
wc	120		
hzmm	35.68		
is	36.58		
rfl	1377.9		
rfp	999.9		
th	3		
ins	1.000		
ai	aw		
	2D DISPLAY		
sp1	-19.0		
wp1	34.9		
sc2	0		
wc2	120		
rfl1	24.6		
rfp1	0		

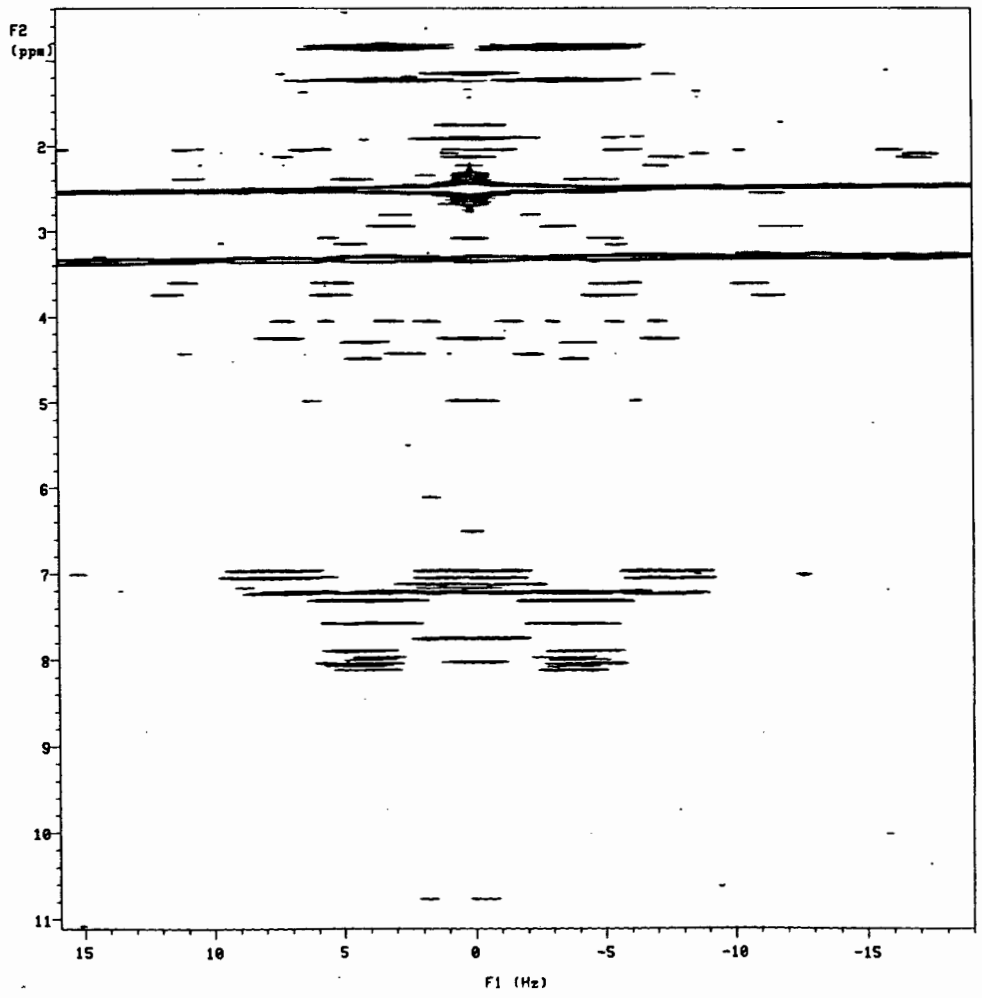


Figure A2.18: Full ^{1}H spectrum generated for Lom-AKH-III in d^{6}-DMSO.

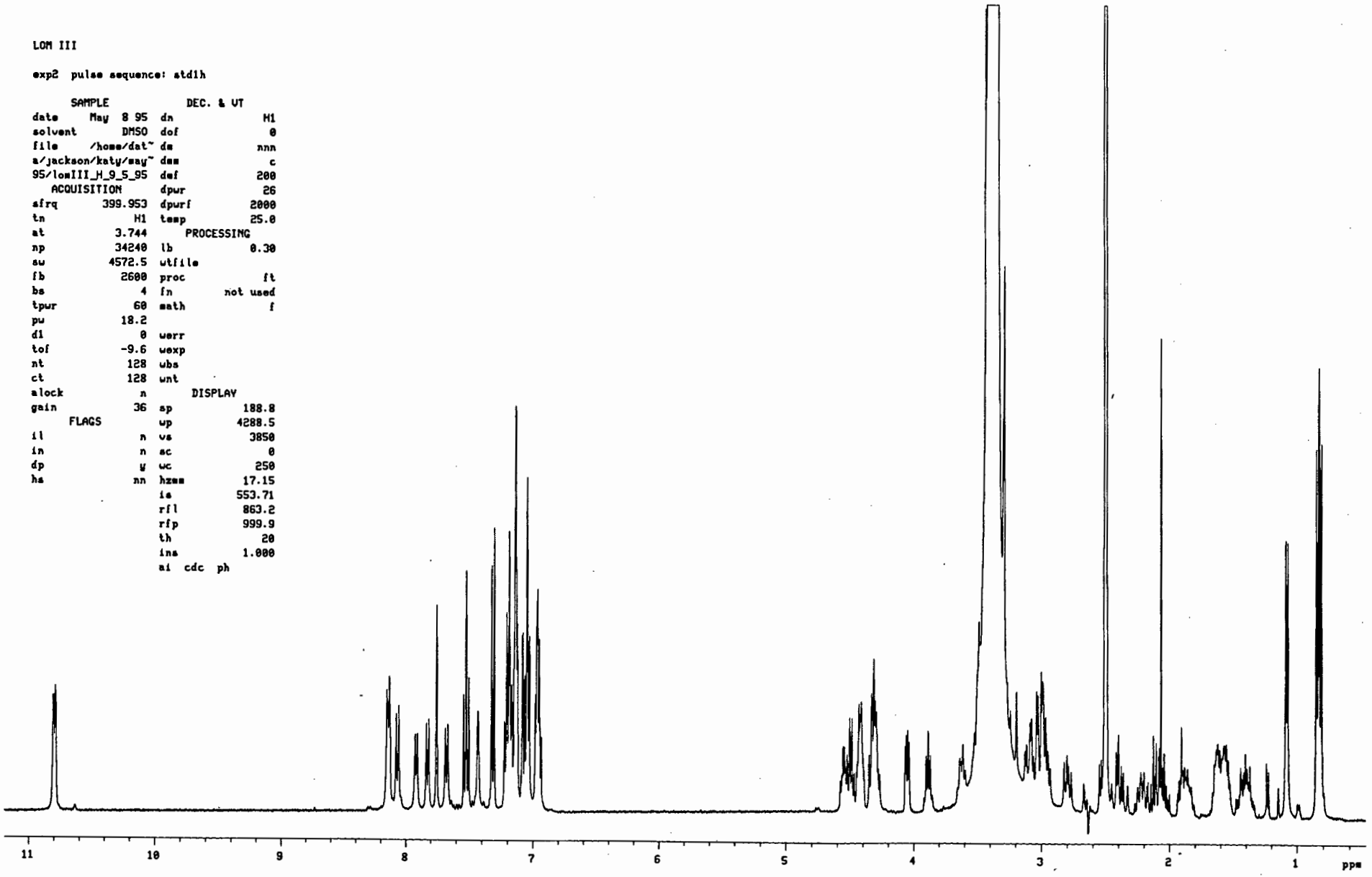
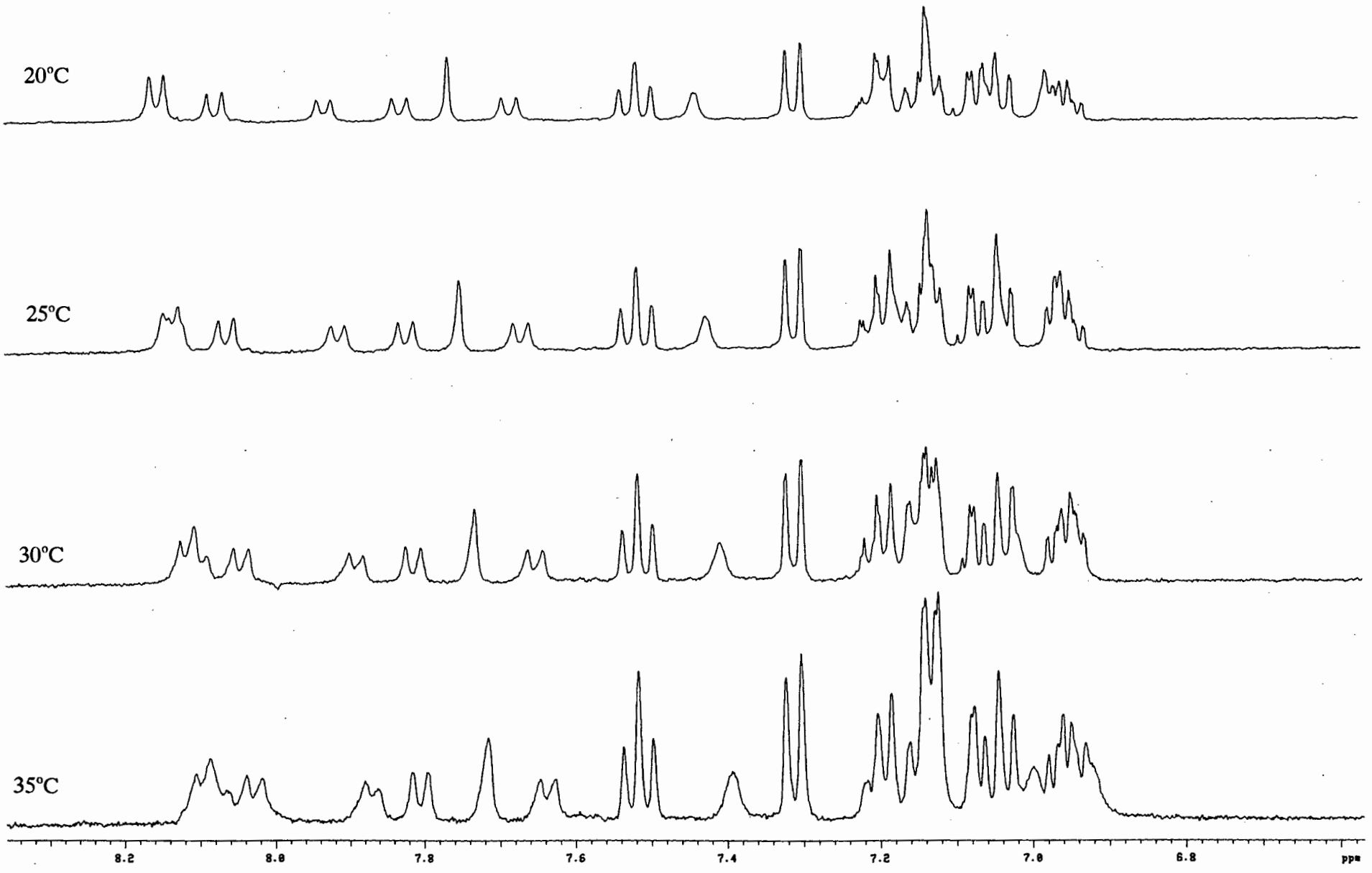


Figure A2.19: Temperature arrayed ^{1}H spectra generated for Lom-AKH-III in d^{6}-DMSO.



Chemical Shift vs Temperature graph generated for Lom-AKH-III

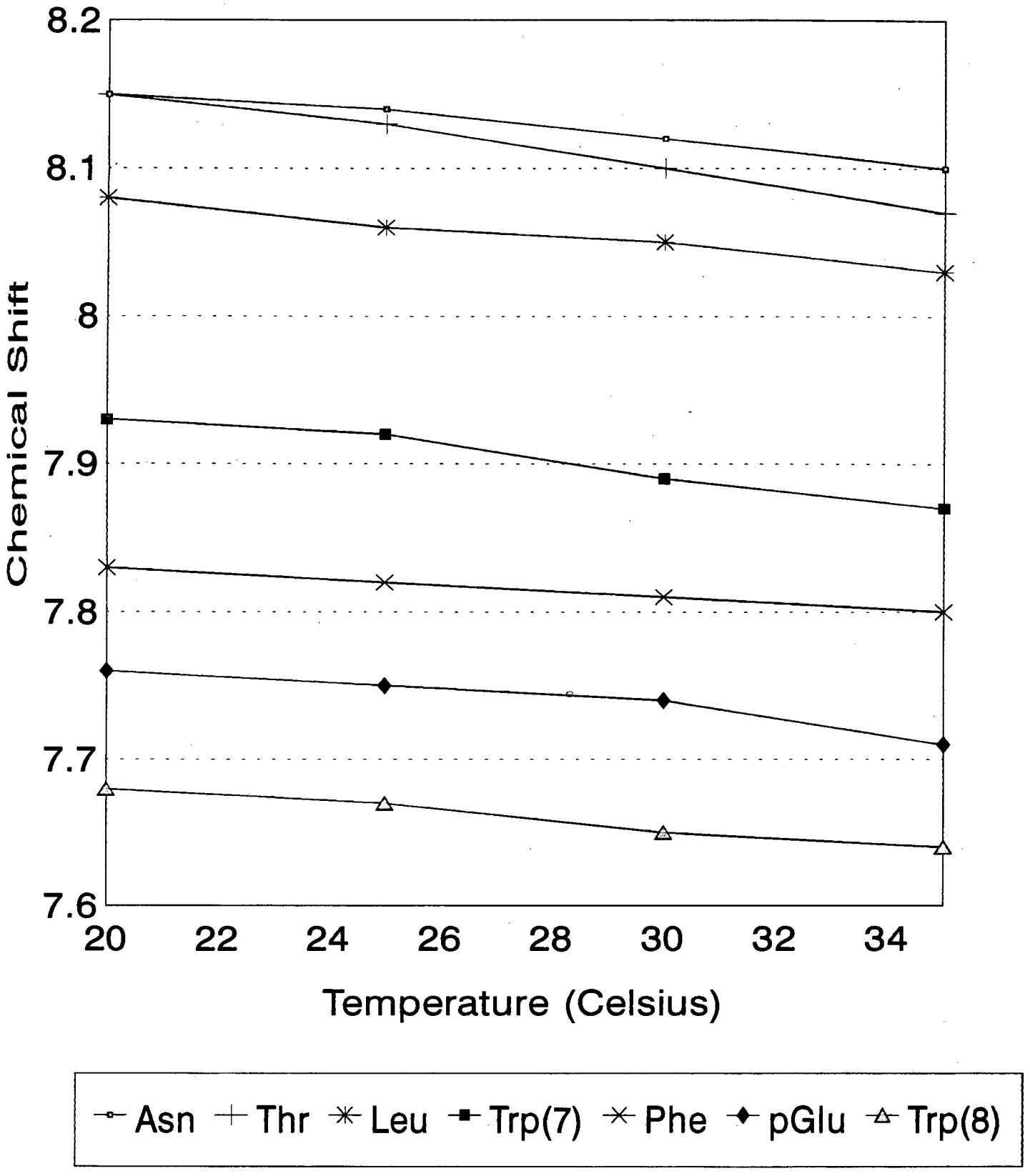


Figure A2.20: Full COSY spectrum generated for Lom-AKH-III in d^6-DMSO.

```
LOM III

expl  pulse sequence: relayh

      SAMPLE           2D SPECTRUM
date      May  3 95  sw1          3123.3
temp         25.0  ni              320
solvent      DMSO      PROCESSING
file     /home/dat~ fn            1024
a/jackson/katy/may~ sb           0.082
95/lomIII_cosy_4_5~ sbs       not used
               _95    2D PROCESSING
  PULSE SEQUENCE    fn1           1024
d1              1  sb1           0.042
p1           18.0  sbs1      not used
relay           0       DISPLAY
tau             0  sp           237.2
     OBSERVE       wp          3123.3
tn             H1  vs             600
sfrq       399.952 sc               0
tof         -633.4 wc             120
pw90         18.0  axis            pp
tpwr           60  rfl           762.7
     DECOUPLE      rfp           999.9
dn             H1  lp              3.4
dfrq      not used rp            139.4
dof             0  ai  cdc  av
dm            nnn      2D DISPLAY
dmm             c  sp1          237.2
dmf           200  wp1          3123.3
dpwr           26  th               2
homo            y  sc2              0
     SPECTRUM      wc2            120
sw         3123.3  rfl1          762.7
fb           1800  rfp1          999.9
at          0.164  lp1              0
np           1024  rp1              0
ct            128
ss              2
gain           56
```

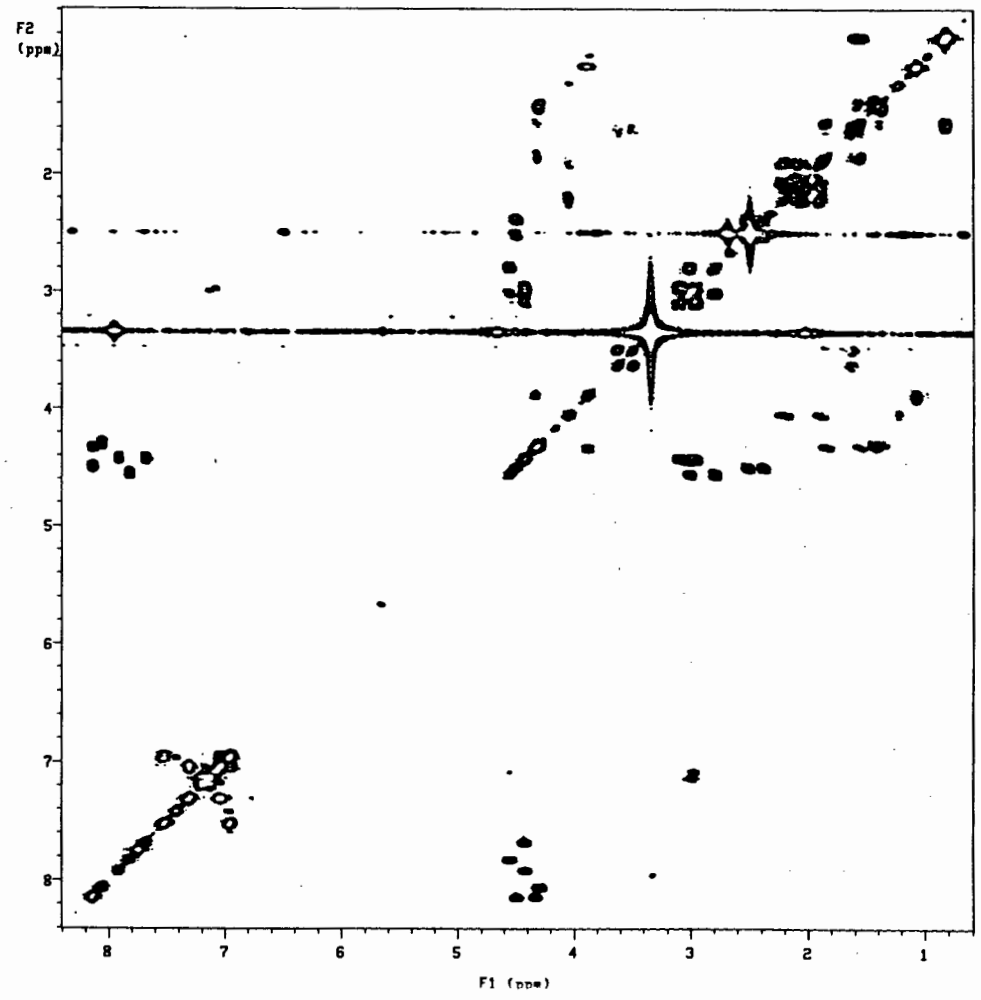


Figure A2.21: Full TOCSY spectrum generated for Lom-AKH-III in d^6-DMSO.

LOM III

expl pulse sequence: tocsy

SAMPLE		DEC. & VT		ACQUISITION ARRAYS	
date	May 11 95	dfrq	not used	array	phase
solvent	DMSO	dn	H1	arraydim	512
file	/home/dat~	dpur	8		
a/jackson/katy/may~		dof	-1079.0	1	phase
95/lomIII_TOCSY_de~		dm	yyn	1	1
coupler_sat		dmm	c	2	2
ACQUISITION		dmf	200		
sfrq	399.953	dseq	undefined		
tn	H1	dres	undefined		
at	0.236	homo	y		
np	2048	temp	25.0		
sw	4338.9	PROCESSING			
fb	2400	sb	-0.236		
bs	4	sbs	-0.236		
ss	32	wtfile			
tpwr	58	proc	ft		
pw	23.0	fn	2048		
pl	18.2	math	f		
dl	1.000				
tof	-159.4	werr			
nt	64	wexp			
ct	64	wbs			
alock	n	wnt			
gain	52	2D PROCESSING			
FLAGS		sb1	0.030		
il	n	sbs1	-0.022		
in	n	gf1	0.026		
dp	y	gfs1	not used		
hs	yn	wtfile1			
2D ACQUISITION		proc1	ft		
sw1	4338.9	fn1	1024		
ni	256				
phase	arrayed				
DISPLAY					
sp	107.1				
wp	4338.9				
vs	150				
sc	6				
wc	110				
hzmm	3.89				
is	56841.51				
rfl	3148.5				
rfp	3255.6				
th	1				
ins	1.000				
ai ph					
2D DISPLAY					
sp1	105.9				
wp1	4338.9				
sc2	0				
wc2	110				
rfl1	3149.8				
rfp1	3255.6				

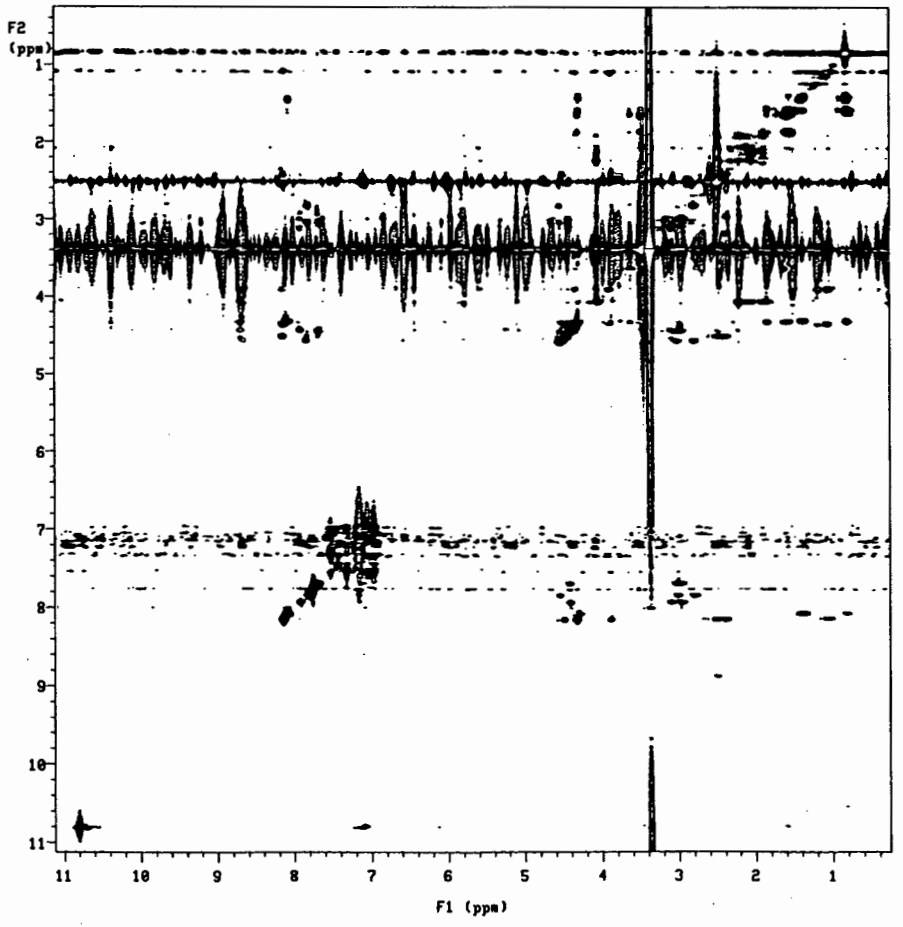


Figure A2.22: Full NOESY spectrum (mix=150ms) generated for Lom-AKH-III in d^6-DMSO.

exp2 pulse sequence: noesy

SAMPLE		DEC. & VT		ACQUISITION ARRAYS	
date	May 12 95	dfrq	not used	array	phase
solvent	DMSO	dn	H1	arraydim	512
file	/home/dat~	dpwr	8		
a/jackson/katy/pro~		dof	0	i	phase
c95/lomiiinoesy_15~		dm	nnn	1	1
	0	dmm	c	2	2
ACQUISITION		dmf	200		
sfrq	399.952	dseq	undefined		
tn	H1	dres	undefined		
at	0.164	homo	y		
np	1024	temp	25.0		
sw	3123.3	PROCESSING			
fb	1800	sb	-0.164		
ss	16	sbs	-0.164		
tpwr	60	gf	0.076		
pw	18.2	gfs	not used		
d1	1.000	wtfile			
presat	0	proc	ft		
mix	0.150	fn	1024		
tof	-633.4	math	f		
nt	64				
ct	64	werr			
alock	n	wexp			
gain	38	wbs			
FLAGS		wnt			
il	y	2D PROCESSING			
in	n	sb1	0.041		
dp	y	sbs1	-0.014		
hs	yn	gf1	0.106		
sspul	n	gfs1	not used		
2D ACQUISITION		wtfile1			
sw1	3123.3	proc1	ft		
ni	256	fn1	1024		
phase	arrayed				
DISPLAY					
sp	237.2				
wp	3123.3				
vs	450				
sc	0				
wc	120				
hzmm	3.36				
is	3.47				
rfl	762.7				
rfp	999.9				
th	0				
ins	1.000				
ai cdc ph					
2D DISPLAY					
sp1	237.2				
wp1	3123.3				
sc2	0				
wc2	120				
rfl1	762.7				
rfp1	999.9				

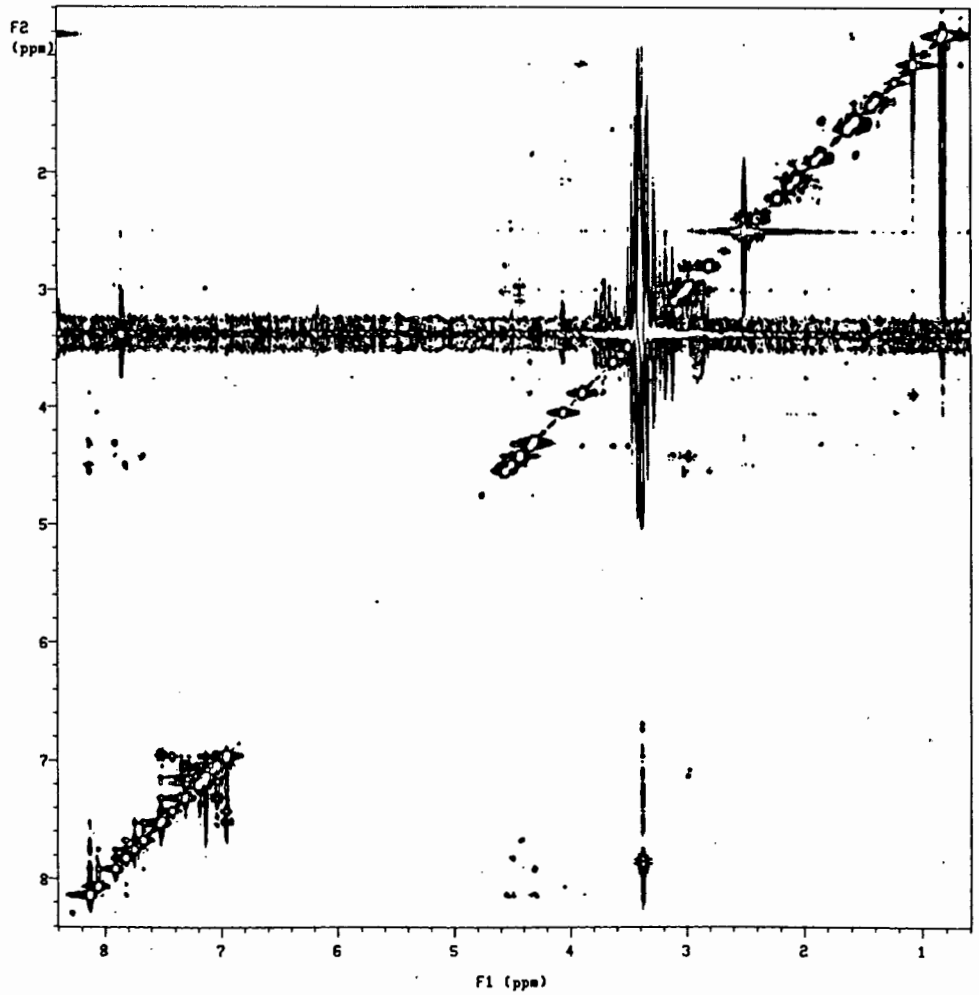


Figure A2.23: Full NOESY spectrum (mix=200ms) generated for Lom-AKH-III in d^6-DMSO.

expl pulse sequence: noesy

SAMPLE		DEC. & VT		ACQUISITION ARRAYS	
date	May 14 95	dfrq	not used	array	phase
solvent	DMSO	dn	H1	arraydim	512
file	/home/dat~	dpwr	8		
a/jackson/katy/pro~		dof	0	1	phase
c95/lomiiinoesy_20~		dm	nnn	1	1
	0	dmm	c	2	2
ACQUISITION		dmf	200		
sfrq	399.952	dseq	undefined		
tn	H1	dres	undefined		
at	0.164	homo	y		
np	1024	temp	25.0		
sw	3123.3	PROCESSING			
fb	1800	sb	-0.164		
ss	16	sbs	-0.164		
tpwr	60	gf	0.076		
pw	18.2	gfs	not used		
d1	1.000	wtfile			
presat	0	proc	ft		
mix	0.200	fn	1024		
tof	-633.4	math	f		
nt	64				
ct	64	werr			
alock	n	wexp			
gain	38	wbs			
FLAGS		wnt			
il	y	2D PROCESSING			
in	n	sb1	0.041		
dp	y	sbs1	-0.006		
hs	yn	gf1	0.048		
aspul	n	gfs1	not used		
2D ACQUISITION		wtfile1			
sw1	3123.3	proc1	ft		
ni	256	fn1	1024		
phase	arrayed				
DISPLAY					
sp	237.2				
wp	3123.3				
vs	35000				
sc	0				
wc	120				
hzmm	3.06				
is	705.05				
rfl	762.7				
rfp	999.9				
th	6				
ins	1.000				
ai ph					
2D DISPLAY					
sp1	237.2				
wp1	3123.3				
sc2	0				
wc2	120				
rfl1	762.7				
rfp1	999.9				

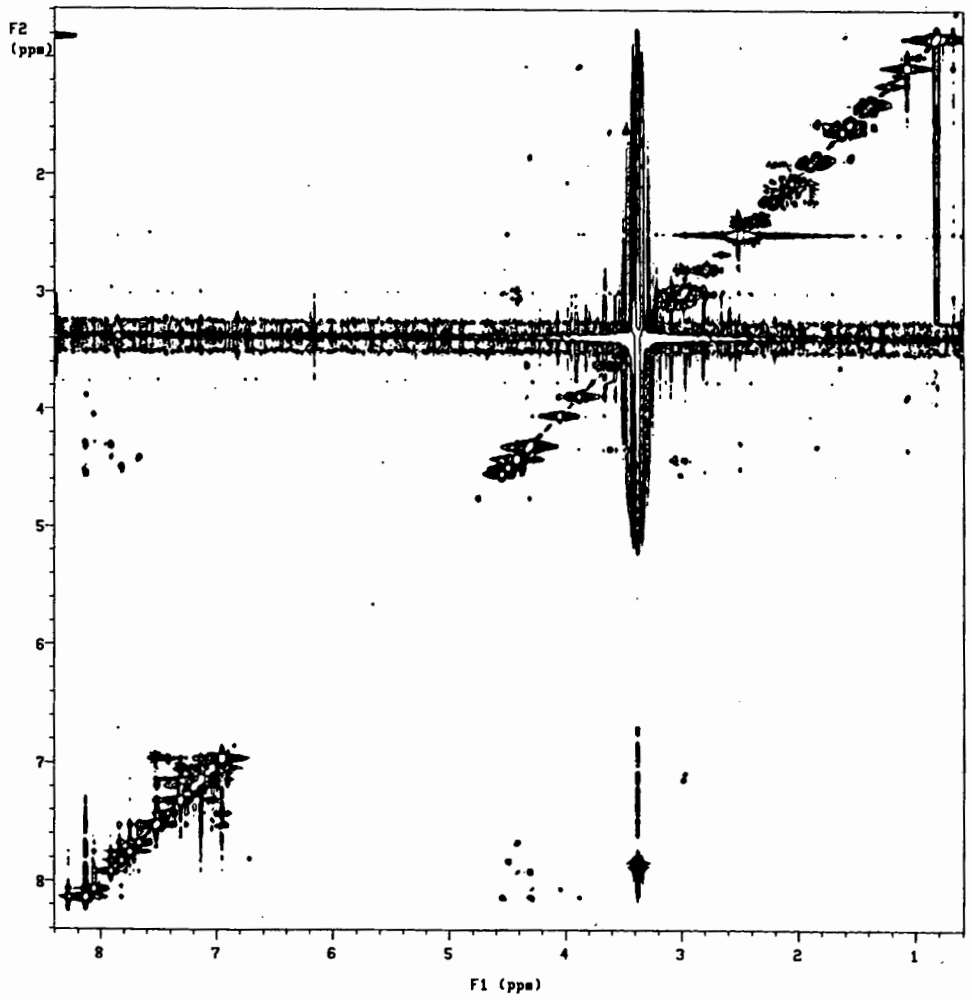


Figure A2.24: Full NOESY spectrum (mix=250ms) generated for Lom-AKH-III in d^6-DMSO.

```
expl  pulse sequence: noesy

      SAMPLE            DEC. & VT       ACQUISITION ARRAYS
date      May 14 95  dfrq      not used  array        phase
solvent       DMSO   dn             H1  arraydim       S12
file      /home/dat~ dpur           8
s/jackson/katy/pro~  dof            0   1           phase
c95/lomiiinoesy_25~  dm           nnn   1             1
                  0  dmm            c   2             2
   ACQUISITION       dmf          200
sfrq        399.952  dseq     undefined
tn               H1  dres     undefined
at            0.164  homo           y
np             1024  temp        25.0
sw           3123.3       PROCESSING
fb             1800  sb         -0.164
ss               16  sbs        -0.164
tpwr             60  gf          0.076
pw             18.2  gfs      not used
d1            1.000  wtfile
presat           0   proc           ft
mix           0.250  fn           1024
tof          -633.4  math           f
nt               64
ct               64  werr
alock             n  wexp
gain             38  wbs
      FLAGS          wnt
il                y    2D PROCESSING
in                n  sb1        0.030
dp                y  sbs1       -0.013
hs               yn  gf1         0.102
aspul             n  gfs1     not used
  2D ACQUISITION     wtfile1
sw1          3123.3  proc1          ft
ni            256    fn1          1024
phase      arrayed
     DISPLAY
sp           237.2
wp           3123.3
vs             548
sc               6
wc             120
hzmm         26.03
is            0.00
rfl           762.7
rfp           999.9
th               0
ins           1.000
ai  cdc  ph
    2D DISPLAY
sp1          238.7
wp1          3123.3
sc2              0
wc2            120
rfl1         3012.9
rfp1         3251.6
```

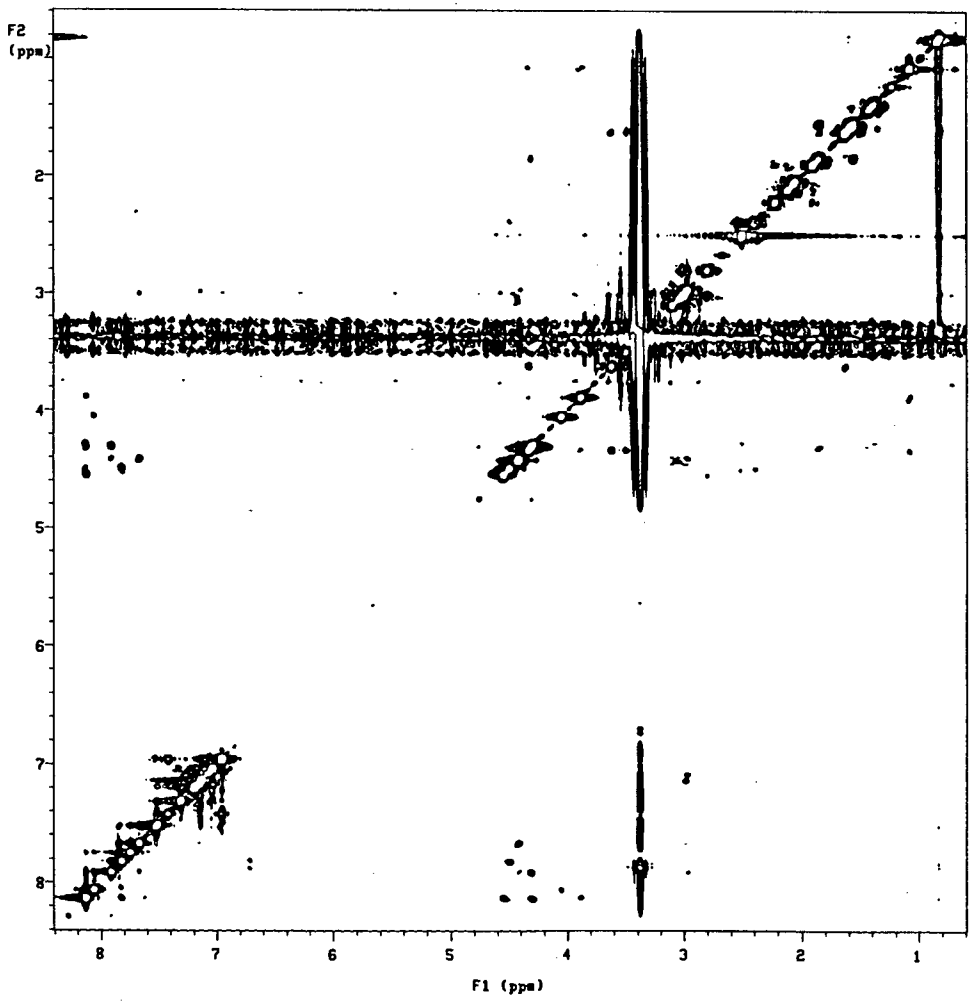
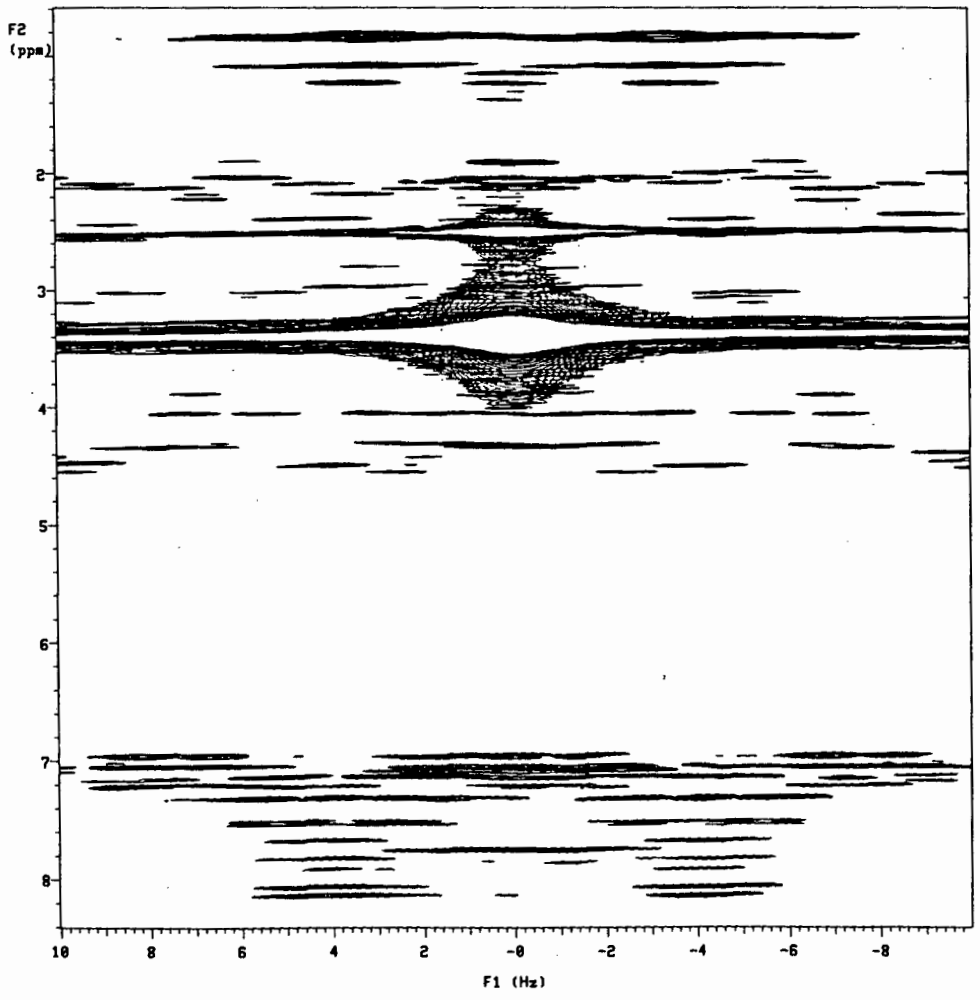


Figure A2.25: Full Homonuclear 2DJ-Resolved spectrum generated for Lom-AKH-III in d^6- DMSO.

hom2dj

exp2 pulse sequence: hom2dj

	SAMPLE		2D SPECTRUM
date	May 12 95	sw1	20.0
temp	25.0	ni	128
solvent	DMSO		PROCESSING
file	/home/dat~	fn	2048
a/jackson/katy/may~		sb	0.082
95/hom2dj_14_5		sbs	not used
	PULSE SEQUENCE		2D PROCESSING
d1	1	fn1	256
p1	36.5	sb1	0.640
	OBSERVE	sbs1	not used
tn	H1		DISPLAY
sfrq	399.952	sp	237.4
tof	-633.4	wp	3123.3
pw90	18.2	vs	7000
tpwr	60	sc	0
	DECOUPLE	wc	120
dn	H1	axis	ph
dfrq	not used	rfl	2863.5
dof	0	rfp	3100.8
dpwr	26	lp	3.4
homo	y	rp	140.6
	SPECTRUM	ai	cdc av
sw	3123.3		2D DISPLAY
fb	1800	sp1	-10.0
at	0.164	wp1	20.0
np	1024	th	2
ct	64	sc2	0
ss	16	wc2	120
gain	38	rfl1	10.0
		rfp1	0
		lp1	0
		rp1	0



APPENDIX 3

Input File A3.1: Macromodel MC search.

```
$ SET DEFAULT CEM:[SNTKAT01.DGEOM.LOMM]
$ RBATMIN
BatchMin V3.1c                Starting Time 29-MAR-95  16:36:53
 LOM227.dat
 lom228
 FFLD      3
 DEMX      0    100                 100.0000    50.0000
 MCMM    5000      0           0
 MCOP      50
 MCSS      2      0                 50.0000
 MCNV      2      7      0      0
 TORS      3      5     48     50     0.0000   180.0000
 TORS     15     17     34     36     0.0000   180.0000
 TORS     52     55     68     70     0.0000   180.0000
 TORS     81     83     96     98     0.0000   180.0000
 TORS     80     81    110    112     0.0000   180.0000
 TORS    132    134    134    137     0.0000   180.0000
 TORS    141    153      1      3     0.0000   180.0000
 TORS    114    117     19     22     0.0000   180.0000
 FXTA      5     13     15     17           -146.6700    3.0000
 FXTA     17     32     34     36           -149.8400    3.0000
 FXTA     36     46     48     50           -148.3400    3.0000
 FXTA     50     66     68     70           -155.3500    3.0000
 FXTA     83     94     96     98           -148.8100    3.0000
 FXTA     98    108    110    112           -153.8200    3.0000
 FXTA    112    132    134    137           -168.5200    3.0000
 FXTA    137    139    141    153           -143.1300    3.0000
 FXTA     71     70     80     91            180.0000    0.0000
 FXDI     92     69                          2.5000    0.3000
 FXDI    109     95                          5.0000    2.5000
 FXDI    109     97                          2.8000    0.4000
 FXDI     93     69                          2.3500    0.1000
 FXDI     75     67                          3.8000    0.2000
 FXDI     73     67                          2.5000    0.2000
 FXDI    111     95                          2.5000    0.2000
 FXDI    133    140                          5.0000    2.5000
 FXDI     33    142                          5.0000    2.5000
 FXDI      2     14                          5.0000    2.5000
 FXDI     82     95                          2.4000    0.2000
 FXDI     35     47                          2.3000    0.1000
 FXDI     82     67                          2.8000    0.2000
 FXDI    109    133                          5.0000    2.5000
 FXDI    111    133                          2.8000    0.1000
 FXDI      2    135                          9.0000    5.0000
 FXDI      2     82                          9.0000    5.0000
 FXDI      2    111                          9.0000    5.0000
```

```
FXDI      2     73                    9.0000    5.0000
FXDI      2    136                    9.0000    5.0000
FXDI      2    142                    9.0000    5.0000
FXDI     47    135                    9.0000    5.0000
FXDI     47    136                    9.0000    5.0000
FXDI     47     75                    9.0000    5.0000
FXDI      4    133                    9.0000    5.0000
FXDI     73    133                    9.0000    5.0000
FXDI     16    133                    9.0000    4.5000
FXDI      9    133                    9.0000    4.5000
FXDI    102    109                    9.0000    6.8000
COMP      1      3      5     13
COMP     15     17     32     34
COMP     36     46     48     50
COMP     66     68     70     80
COMP     81     83     94     96
COMP    108    110    112    132
COMP    134    137    139    141
BGIN
READ
MINI      1      0   1000
CHIG      3     15     34     48
CHIG     68     81     96    110
CHIG    141
CONV      2      2
END
```

Input File A3.2: First restrained MD run for Lom-AKH-I.

```
      BEGIN SIMULATION
    reduce output
    constrain using "lom.rstrnt"
!

      assign torsion name tor1
*     residue PGLU 1 atom C
*     residue LEU 1B atom N, atom CA atom C
      assign torsion name tor2
*     residue LEU 1B atom C
*     residue ASN 1C atom N, atom CA atom C
      assign torsion name tor3
*     residue ASN 1C atom C
*     residue PHE 1D atom N, atom CA atom C
      assign torsion name tor4
*     residue PHE 1D atom C
*     residue THR 1E atom N, atom CA atom C
      assign torsion name tor5
*     residue PRO 1F atom C
*     residue ASN 1G atom N, atom CA atom C
      assign torsion name tor6
*     residue ASN 1G atom C
*     residue TRP 1H atom N, atom CA atom C
     assign torsion name tor7
*     residue TRP 1H atom C
*     residue GLY 1I atom N, atom CA atom C
     assign torsion name tor8
*     residue GLY 1I atom C
*     residue THR 1J atom N, atom CA atom C
     assign torsion name tor9
*     residue THR 1E atom CA, atom C
*     residue PRO 1F atom N, atom CA
      print torsions
     force tor1 to -146.67 using 60.0 kcal/rad**2 in molecule 1
*     residue LEU 1B
     force tor2 to -149.84 using 60.0 kcal/rad**2 in molecule 1
*     residue ASN 1C
     force tor3 to -148.34 using 60.0 kcal/rad**2 in molecule 1
*     residue PHE 1D
     force tor4 to -155.35 using 60.0 kcal/rad**2 in molecule 1
*     residue THR 1E
     force tor5 to -148.81 using 100 kcal/rad**2 in molecule 1
*     residue ASN 1G
     force tor6 to -153.82 using 60.0 kcal/rad**2 in molecule 1
*     residue TRP 1H
     force tor7 to -168.52 using 60.0 kcal/rad**2 in molecule 1
```

```
*    residue GLY 1I
     force tor8 to -143.13 using 60.0 kcal/rad**2 in molecule 1
*    residue THR 1J
     force omega to trans in residues PGLU 1 to THR 1J
*    using 60.0 kcal/mol

     pull atom1 in molecule 1 residue THR 1E atom HA
*    and atom2 in molecule 1 residue PRO 1F atom HD1 to 2.4
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue THR 1E atom HA
*    and atom2 in molecule 1 residue PRO 1F atom HD2 to 2.8
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue THR 1E atom HN
*    and atom2 in molecule 1 residue THR 1E atom HB  to 2.54
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue THR 1E atom HN
*    and atom2 in molecule 1 residue THR 1E atom HG1 to 3.708
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue ASN 1C atom HN
*    and atom2 in molecule 1 residue LEU 1B atom HA to 2.500
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue PRO 1F atom HA
*    and atom2 in molecule 1 residue ASN 1G atom HN to 2.500
*    with a force = 40 kcal/mol
     pull atom1 in molecule 1 residue GLY 1I atom HN
*    and atom2 in molecule 1 residue TRP 1H atom HA to 3.000
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue PHE 1D atom HN
*    and atom2 in molecule 1 residue ASN 1C atom HA to 2.600
*    with a force = 60 kcal/mol

!


     MINIMIZE USING STEEPEST DESCENT FOR 1000 CYCLES
*    USING NO CROSS, NO MORSE, NO CHARGES
*    UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN 1000.0  KCAL/A
     MINIMIZE USING CONJUGATE GRADIENTS FOR 1000 CYCLES
*    USING NO CROSS, NO MORSE, NO CHARGES
*    UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN   10.0  KCAL/A

     INITIALIZE DYNAMICS AT 450 K FOR 10000 STEPS OF 1 FS
*    WITH NO CROSS AND NO MORSE AND NO CHARGES
*    WRITE HISTORY  EVERY 1000 STEPS
*    WRITE AVERAGES EVERY 1000 STEPS

     RESUME DYNAMICS AT 450 K FOR 10000 STEPS OF 1 FS
```

```
    *    WITH NO CROSS AND NO MORSE AND NO CHARGES
    *    WRITE HISTORY  EVERY 1000 STEPS
    *    WRITE AVERAGES EVERY 1000 STEPS
!
      nstructs=100

      istruct=0
100    istruct=istruct+1
       RESUME DYNAMICS AT 300 K FOR  5000 STEPS OF 1 FS
    *    WITH NO CROSS AND NO MORSE AND NO CHARGES
       ARCHIVE AS FILE NUMBER istruct
      IF istruct .LT. nstructs THEN 100

      istruct=0
200    istruct=istruct+1
       RETRIEVE FILE NUMBER istruct

         MINIMIZE USING STEEPEST DESCENT FOR 10000 CYCLES
    *      USING NO CROSS, NO MORSE, NO CHARGES
    *      UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN 1000.0  KCAL/A
         MINIMIZE USING CONJUGATE GRADIENTS FOR 10000 CYCLES
    *      USING NO CROSS, NO MORSE, NO CHARGES
    *      UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN  0.01  KCAL/A


         print torsions
         print hbond
         hbdist=3.4
         hbangl=120


       ARCHIVE AS FILE NUMBER istruct
       IF istruct .LT. nstructs THEN 200
!
       END
```

Input File A3.3: Second restrained MD run for Lom-AKH-I.

```
      BEGIN SIMULATION
    reduce output
    constrain using "lom.rstrnt"
!

      assign torsion name tor1
*     residue PGLU 1 atom C
*     residue LEU 1B atom N, atom CA atom C
      assign torsion name tor2
*     residue LEU 1B atom C
*     residue ASN 1C atom N, atom CA atom C
      assign torsion name tor3
*     residue ASN 1C atom C
*     residue PHE 1D atom N, atom CA atom C
      assign torsion name tor4
*     residue PHE 1D atom C
*     residue THR 1E atom N, atom CA atom C
      assign torsion name tor5
*     residue PRO 1F atom C
*     residue ASN 1G atom N, atom CA atom C
      assign torsion name tor6
*     residue ASN 1G atom C
*     residue TRP 1H atom N, atom CA atom C
     assign torsion name tor7
*     residue TRP 1H atom C
*     residue GLY 1I atom N, atom CA atom C
     assign torsion name tor8
*     residue GLY 1I atom C
*     residue THR 1J atom N, atom CA atom C
     assign torsion name tor9
*     residue THR 1E atom CA, atom C
*     residue PRO 1F atom N, atom CA
      print torsions
     force tor1 to -146.67 using 60.0 kcal/rad**2 in molecule 1
*     residue LEU 1B
     force tor2 to -149.84 using 60.0 kcal/rad**2 in molecule 1
*     residue ASN 1C
     force tor3 to -148.34 using 60.0 kcal/rad**2 in molecule 1
*     residue PHE 1D
     force tor4 to -155.35 using 60.0 kcal/rad**2 in molecule 1
*     residue THR 1E
     force tor5 to -148.81 using 100 kcal/rad**2 in molecule 1
*     residue ASN 1G
     force tor6 to -153.82 using 60.0 kcal/rad**2 in molecule 1
*     residue TRP 1H
     force tor7 to -168.52 using 60.0 kcal/rad**2 in molecule 1
```

```
*    residue GLY 1I
     force tor8 to -143.13 using 60.0 kcal/rad**2 in molecule 1
*    residue THR 1J
     force omega to trans in residues PGLU 1 to THR 1J
*    using 60.0 kcal/mol

     pull atom1 in molecule 1 residue THR 1E atom HA
*    and atom2 in molecule 1 residue PRO 1F atom HD1 to 2.4
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue THR 1E atom HA
*    and atom2 in molecule 1 residue PRO 1F atom HD2 to 2.8
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue THR 1E atom HN
*    and atom2 in molecule 1 residue THR 1E atom HB  to 2.54
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue THR 1E atom HN
*    and atom2 in molecule 1 residue THR 1E atom HG1 to 3.708
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue ASN 1C atom HN
*    and atom2 in molecule 1 residue LEU 1B atom HA to 2.500
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue PRO 1F atom HA
*    and atom2 in molecule 1 residue ASN 1G atom HN to 2.500
*    with a force = 40 kcal/mol
   pull atom1 in molecule 1 residue GLY 1I atom HN
*    and atom2 in molecule 1 residue TRP 1H atom HA to 3.000
*    with a force = 60 kcal/mol
   pull atom1 in molecule 1 residue PHE 1D atom HN
*    and atom2 in molecule 1 residue ASN 1C atom HA to 2.600
*    with a force = 60 kcal/mol

!

   MINIMIZE USING STEEPEST DESCENT FOR 1000 CYCLES
*    USING NO CROSS, NO MORSE, NO CHARGES
*    UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN 1000.0  KCAL/A
   MINIMIZE USING CONJUGATE GRADIENTS FOR 1000 CYCLES
*    USING NO CROSS, NO MORSE, NO CHARGES
*    UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN   10.0  KCAL/A

   INITIALIZE DYNAMICS AT 450 K FOR 10000 STEPS OF 1 FS
*    WITH NO CROSS AND NO MORSE AND NO CHARGES
*    WRITE HISTORY  EVERY 1000 STEPS
*    WRITE AVERAGES EVERY 1000 STEPS

   RESUME DYNAMICS AT 450 K FOR 10000 STEPS OF 1 FS
```

```
    *    WITH NO CROSS AND NO MORSE AND NO CHARGES
    *    WRITE HISTORY  EVERY 1000 STEPS
    *    WRITE AVERAGES EVERY 1000 STEPS
!
      nstructs=100

      istruct=0
100    istruct=istruct+1
       RESUME DYNAMICS AT 300 K FOR  5000 STEPS OF 1 FS
    *    WITH NO CROSS AND NO MORSE AND NO CHARGES
       ARCHIVE AS FILE NUMBER istruct
      IF istruct .LT. nstructs THEN 100

      istruct=0
200    istruct=istruct+1
       RETRIEVE FILE NUMBER istruct

         MINIMIZE USING STEEPEST DESCENT FOR 10000 CYCLES
    *      USING NO CROSS, NO MORSE, NO CHARGES
    *      UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN 1000.0  KCAL/A
         MINIMIZE USING CONJUGATE GRADIENTS FOR 10000 CYCLES
    *      USING NO CROSS, NO MORSE, NO CHARGES
    *      UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN  0.01  KCAL/A


         print torsions
         print hbond
         hbdist=3.4
         hbangl=120

       ARCHIVE AS FILE NUMBER istruct
       IF istruct .LT. nstructs THEN 200
!
       END
```

Input File A3.4: Third restrained MD run for Lom-AKH-I.

```
    BEGIN SIMULATION
  reduce output
  constrain using "lom.rstrnt"
!

    assign torsion name tor1
*    residue PGLU 1 atom C
*    residue LEU 1B atom N, atom CA atom C
    assign torsion name tor2
*    residue LEU 1B atom C
*    residue ASN 1C atom N, atom CA atom C
    assign torsion name tor3
*    residue ASN 1C atom C
*    residue PHE 1D atom N, atom CA atom C
    assign torsion name tor4
*    residue PHE 1D atom C
*    residue THR 1E atom N, atom CA atom C
    assign torsion name tor5
*    residue PRO 1F atom C
*    residue ASN 1G atom N, atom CA atom C
    assign torsion name tor6
*    residue ASN 1G atom C
*    residue TRP 1H atom N, atom CA atom C
   assign torsion name tor7
*    residue TRP 1H atom C
*    residue GLY 1I atom N, atom CA atom C
   assign torsion name tor8
*    residue GLY 1I atom C
*    residue THR 1J atom N, atom CA atom C
   assign torsion name tor9
*    residue THR 1E atom CA, atom C
*    residue PRO 1F atom N, atom CA
    print torsions
   force tor1 to -146.67 using 60.0 kcal/rad**2 in molecule 1
*    residue LEU 1B
   force tor2 to -149.84 using 60.0 kcal/rad**2 in molecule 1
*    residue ASN 1C
   force tor3 to -148.34 using 60.0 kcal/rad**2 in molecule 1
*    residue PHE 1D
   force tor4 to -155.35 using 60.0 kcal/rad**2 in molecule 1
*    residue THR 1E
   force tor5 to -148.81 using 100 kcal/rad**2 in molecule 1
*    residue ASN 1G
   force tor6 to -153.82 using 60.0 kcal/rad**2 in molecule 1
*    residue TRP 1H
   force tor7 to -168.52 using 60.0 kcal/rad**2 in molecule 1
```

```
*    residue GLY 1I
     force tor8 to -143.13 using 60.0 kcal/rad**2 in molecule 1
*    residue THR 1J
     force omega to trans in residues PGLU 1 to THR 1J
*    using 60.0 kcal/mol

     pull atom1 in molecule 1 residue THR 1E atom HA
*    and atom2 in molecule 1 residue PRO 1F atom HD1 to 2.4
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue THR 1E atom HA
*    and atom2 in molecule 1 residue PRO 1F atom HD2 to 2.8
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue THR 1E atom HN
*    and atom2 in molecule 1 residue THR 1E atom HB  to 2.54
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue THR 1E atom HN
*    and atom2 in molecule 1 residue THR 1E atom HG1 to 3.708
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue ASN 1C atom HN
*    and atom2 in molecule 1 residue LEU 1B atom HA to 2.500
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue PRO 1F atom HA
*    and atom2 in molecule 1 residue ASN 1G atom HN to 2.500
*    with a force = 40 kcal/mol
   pull atom1 in molecule 1 residue GLY 1I atom HN
*    and atom2 in molecule 1 residue TRP 1H atom HA to 3.000
*    with a force = 60 kcal/mol
   pull atom1 in molecule 1 residue PHE 1D atom HN
*    and atom2 in molecule 1 residue ASN 1C atom HA to 2.600
*    with a force = 60 kcal/mol

!


   MINIMIZE USING STEEPEST DESCENT FOR 1000 CYCLES
*    USING NO CROSS, NO MORSE, NO CHARGES
*    UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN 1000.0  KCAL/A
   MINIMIZE USING CONJUGATE GRADIENTS FOR 1000 CYCLES
*    USING NO CROSS, NO MORSE, NO CHARGES
*    UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN   10.0  KCAL/A

   INITIALIZE DYNAMICS AT 450 K FOR 10000 STEPS OF 1 FS
*    WITH NO CROSS AND NO MORSE AND NO CHARGES
*    WRITE HISTORY  EVERY 1000 STEPS
*    WRITE AVERAGES EVERY 1000 STEPS

   RESUME DYNAMICS AT 450 K FOR 10000 STEPS OF 1 FS
```

```
    *    WITH NO CROSS AND NO MORSE AND NO CHARGES
    *    WRITE HISTORY  EVERY 1000 STEPS
    *    WRITE AVERAGES EVERY 1000 STEPS
!
      nstructs=100

      istruct=0
100    istruct=istruct+1
       RESUME DYNAMICS AT 300 K FOR  5000 STEPS OF 1 FS
    *     WITH NO CROSS AND NO MORSE AND NO CHARGES
       ARCHIVE AS FILE NUMBER istruct
      IF istruct .LT. nstructs THEN 100

      istruct=0
200    istruct=istruct+1
       RETRIEVE FILE NUMBER istruct

         MINIMIZE USING STEEPEST DESCENT FOR 10000 CYCLES
    *      USING NO CROSS, NO MORSE, NO CHARGES
    *      UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN 1000.0  KCAL/A
         MINIMIZE USING CONJUGATE GRADIENTS FOR 10000 CYCLES
    *      USING NO CROSS, NO MORSE, NO CHARGES
    *      UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN  0.01  KCAL/A


         print torsions
         print hbond
         hbdist=3.4
         hbangl=120

        ARCHIVE AS FILE NUMBER istruct
        IF istruct .LT. nstructs THEN 200
!
        END
```

Input File A3.5: Restraint file for Lom-AKH-I.

```
!BIOSYM restraint 1
!
#NOE_distance
1:NH2_1K:HN1   1:ASN_1C:HA    2.10  10.00  0.000 100.000 100.000  1000.0
1:PHE_1D:HN    1:THR_1E:HG1   4.10  15.00  0.000 100.000 100.000  1000.0
1:PHE_1D:HN    1:THR_1E:HB    4.10  15.00  0.000 100.000 100.000  1000.0
1:PGLU_1:HN    1:THR_1J:HA    4.10  15.00  0.000 100.000 100.000  1000.0
1:PGLU_1:HN    1:THR_1J:HB    4.10  15.00  0.000 100.000 100.000  1000.0
1:PGLU_1:HA    1:GLY_1I:HN    4.10  15.00  0.000 100.000 100.000  1000.0
1:PRO_1F:HA    1:TRP_1H:HN    4.10  15.00  0.000 100.000 100.000  1000.0
1:ASN_1G:HA    1:TRP_1H:HN    2.20   2.40  0.000 100.000 100.000  1000.0
1:PHE_1D:HA    1:THR_1E:HN    2.20   2.80  0.000 100.000 100.000  1000.0
1:LEU_1B:HN    1:PGLU_1:HA    3.00   3.90  0.000 100.000 100.000  1000.0
1:LEU_1B:HN    1:THR_1J:HB    3.00  10.00  0.000 100.000 100.000  1000.0
```

Input File A3.6: Restrained MD run for structure A of Lom-AKH-II.

```
    BEGIN SIMULATION
 reduce output
 constrain using "lom2.rstrnt"
    assign torsion name tor1
*   residue PGLU 1 atom C
*   residue LEU 1B atom N, atom CA atom C
    assign torsion name tor2
*   residue LEU 1B atom C
*   residue ASN 1C atom N, atom CA atom C
    assign torsion name tor3
*   residue ASN 1C atom C
*   residue PHE 1D atom N, atom CA atom C
    assign torsion name tor4
*   residue PHE 1D atom C
*   residue SER 1E atom N, atom CA atom C
    assign torsion name tor5
*   residue SER 1E atom C
*   residue ALA 1F atom N, atom CA atom C
    assign torsion name tor6
*   residue GLY 1G atom C
*   residue TRP 1H atom N, atom CA atom C
    assign torsion name tor7
*   residue PGLU 1 atom O, atom C
*   residue LEU 1B atom N, atom HN
    assign torsion name tor8
*   residue LEU 1B atom O, atom C
*   residue ASN 1C atom N, atom HN
    assign torsion name tor9
*   residue ASN 1C atom O, atom C
*   residue PHE 1D atom N, atom HN
    assign torsion name to10
*   residue PHE 1D atom O, atom C
*   residue SER 1E atom N, atom HN
    assign torsion name to11
*   residue SER 1E atom O, atom C
*   residue ALA 1F atom N, atom HN
    assign torsion name to12
*   residue ALA 1F atom O, atom C
*   residue GLY 1G atom N, atom HN
    assign torsion name to13
*   residue GLY 1G atom O, atom C
*   residue TRP 1H atom N, atom HN
     print torsions
    force tor1 to -149.75 using 100 kcal/rad**2 in molecule 1
*   residue LEU 1B
    force tor2 to -151.29 using 100 kcal/rad**2 in molecule 1
```

```
*    residue ASN 1C
     force tor3 to -147.18 using 100 kcal/rad**2 in molecule 1
*    residue PHE 1D
     force tor4 to -150.02 using 100 kcal/rad**2 in molecule 1
*    residue SER 1E
     force tor5 to -150.11 using 100 kcal/rad**2 in molecule 1
*    residue ALA 1F
     force tor6 to -150.48 using 100 kcal/rad**2 in molecule 1
*    residue TRP 1H
     force tor7 to 180.00 using 50 kcal/rad**2 in molecule 1
*    residue PGLU 1
     force tor8 to 180.00 using 50 kcal/rad**2 in molecule 1
*    residue LEU 1B
     force tor9 to 180.00 using 50 kcal/rad**2 in molecule 1
*    residue ASN 1C
     force to10 to 180.00 using 50 kcal/rad**2 in molecule 1
*    residue PHE 1D
     force to11 to 180.00 using 50 kcal/rad**2 in molecule 1
*    residue SER 1E
     force to12 to 180.00 using 50 kcal/rad**2 in molecule 1
*    residue ALA 1F
     force to13 to 180.00 using 50 kcal/rad**2 in molecule 1
*    residue GLY 1G


     pull atom1 in molecule 1 residue GLY 1G atom HN
*    and atom2 in molecule 1 residue GLY 1G atom HA1 to 3.3
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue GLY 1G atom HN
*    and atom2 in molecule 1 residue GLY 1G atom HA2 to 3.10
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue ASN 1C atom HA
*    and atom2 in molecule 1 residue ASN 1C atom HN to 3.15
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue ASN 1C atom HN
*    and atom2 in molecule 1 residue LEU 1B atom HA to 2.50
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue LEU 1B atom HN
*    and atom2 in molecule 1 residue LEU 1B atom HA to 2.55
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue PGLU 1 atom HA
*    and atom2 in molecule 1 residue LEU 1B atom HN to 2.85
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue GLY 1G atom HN
*    and atom2 in molecule 1 residue ALA 1F atom HA  to 3.35
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue SER 1E atom HA
*    and atom2 in molecule 1 residue ALA 1F atom HN  to 2.80
```

```
    *   with a force = 60 kcal/mol
        pull atom1 in molecule 1 residue ASN 1C atom HA
    *   and atom2 in molecule 1 residue TRP 1H atom HN to 2.2
    *   with a force = 60 kcal/mol
        pull atom1 in molecule 1 residue PHE 1D atom HN
    *   and atom2 in molecule 1 residue PHE 1D atom HA to 3.50
    *   with a force = 60 kcal/mol




      MINIMIZE USING STEEPEST DESCENT FOR 10000 CYCLES
    *   USING NO CROSS, NO MORSE, NO CHARGES
    *   UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN 10.0  KCAL/A
      MINIMIZE USING CONJUGATE GRADIENTS FOR 1000 CYCLES
    *   USING NO CROSS, NO MORSE, NO CHARGES
    *   UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN  0.1  KCAL/A

      INITIALIZE DYNAMICS AT 700 K FOR 10000 STEPS OF 1 FS
    *   WITH NO CROSS AND NO MORSE AND NO CHARGES
    *   WRITE HISTORY  EVERY 10000 STEPS
    *   WRITE AVERAGES EVERY 10000 STEPS


!
      nstructs=100
      istruct=0
100   istruct=istruct+1
      RESUME DYNAMICS AT 700 K FOR 5000 STEPS OF 1 FS
    *   CONSTANT ENERGY
    *   WITH NO CROSS AND NO MORSE AND NO CHARGES
    ARCHIVE AS FILE NUMBER istruct
      IF istruct .LT. nstructs THEN 100

      istruct=0
200   istruct=istruct+1
      RETRIEVE FILE NUMBER istruct

        MINIMIZE USING STEEPEST DESCENT FOR 10000 CYCLES
    *     USING NO CROSS, NO MORSE, NO CHARGES
    *     UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN 10.0  KCAL/A
        MINIMIZE USING CONJUGATE GRADIENTS FOR 10000 CYCLES
    *     USING NO CROSS, NO MORSE, NO CHARGES
    *     UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN  0.001  KCAL/A

        print torsions
```

```
hbdist=3.0
hbangl=120
print hbond

ARCHIVE AS FILE NUMBER istruct
IF istruct .LT. nstructs THEN 200
!
END
```

Input File A3.7: Restrained MD run for structure B of Lom-AKH-II.

```
    BEGIN SIMULATION
 reduce output
 constrain using "lom2.rstrnt"
    assign torsion name tor1
*   residue PGLU 1 atom C
*   residue LEU 1B atom N, atom CA atom C
    assign torsion name tor2
*   residue LEU 1B atom C
*   residue ASN 1C atom N, atom CA atom C
    assign torsion name tor3
*   residue ASN 1C atom C
*   residue PHE 1D atom N, atom CA atom C
    assign torsion name tor4
*   residue PHE 1D atom C
*   residue SER 1E atom N, atom CA atom C
    assign torsion name tor5
*   residue SER 1E atom C
*   residue ALA 1F atom N, atom CA atom C
    assign torsion name tor6
*   residue GLY 1G atom C
*   residue TRP 1H atom N, atom CA atom C
    assign torsion name tor7
*   residue PGLU 1 atom O, atom C
*   residue LEU 1B atom N, atom HN
    assign torsion name tor8
*   residue LEU 1B atom O, atom C
*   residue ASN 1C atom N, atom HN
    assign torsion name tor9
*   residue ASN 1C atom O, atom C
*   residue PHE 1D atom N, atom HN
    assign torsion name to10
*   residue PHE 1D atom O, atom C
*   residue SER 1E atom N, atom HN
    assign torsion name to11
*   residue SER 1E atom O, atom C
*   residue ALA 1F atom N, atom HN
    assign torsion name to12
*   residue ALA 1F atom O, atom C
*   residue GLY 1G atom N, atom HN
    assign torsion name to13
*   residue GLY 1G atom O, atom C
*   residue TRP 1H atom N, atom HN
    print torsions
    force tor1 to -149.75 using 100 kcal/rad**2 in molecule 1
*   residue LEU 1B
    force tor2 to -151.29 using 100 kcal/rad**2 in molecule 1
```

```
*    residue ASN 1C
     force tor3 to -147.18 using 100 kcal/rad**2 in molecule 1
*    residue PHE 1D
     force tor4 to -150.02 using 100 kcal/rad**2 in molecule 1
*    residue SER 1E
     force tor5 to -150.11 using 100 kcal/rad**2 in molecule 1
*    residue ALA 1F
     force tor6 to -150.48 using 100 kcal/rad**2 in molecule 1
*    residue TRP 1H
     force tor7 to 180.00 using 100 kcal/rad**2 in molecule 1
*    residue PGLU 1
     force tor8 to 180.00 using 100 kcal/rad**2 in molecule 1
*    residue LEU 1B
     force tor9 to 180.00 using 100 kcal/rad**2 in molecule 1
*    residue ASN 1C
     force to10 to 180.00 using 100 kcal/rad**2 in molecule 1
*    residue PHE 1D
     force to11 to 180.00 using 100 kcal/rad**2 in molecule 1
*    residue SER 1E
     force to12 to 180.00 using 100 kcal/rad**2 in molecule 1
*    residue ALA 1F
     force to13 to 180.00 using 100 kcal/rad**2 in molecule 1
*    residue GLY 1G


     pull atom1 in molecule 1 residue GLY 1G atom HN
*    and atom2 in molecule 1 residue GLY 1G atom HA1 to 3.3
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue GLY 1G atom HN
*    and atom2 in molecule 1 residue GLY 1G atom HA2 to 3.10
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue PHE 1D atom HN
*    and atom2 in molecule 1 residue PHE 1D atom HA to 3.50
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue TRP 1H atom HA
*    and atom2 in molecule 1 residue ALA 1F atom HN to 2.20
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue SER 1E atom HA
*    and atom2 in molecule 1 residue ALA 1F atom HN to 2.60
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue ASN 1C atom HN
*    and atom2 in molecule 1 residue LEU 1B atom HA to 2.50
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue LEU 1B atom HN
*    and atom2 in molecule 1 residue LEU 1B atom HA to 2.55
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue LEU 1B atom HN
*    and atom2 in molecule 1 residue PGLU 1 atom HA to 2.85
```

```
*   with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue SER 1E atom HN
*   and atom2 in molecule 1 residue SER 1E atom HB1 to 3.30
*   with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue GLY 1G atom HN
*   and atom2 in molecule 1 residue SER 1E atom HB2 to 3.05
*   with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue GLY 1G atom HN
*   and atom2 in molecule 1 residue ALA 1F atom HA to 3.00
*   with a force = 60 kcal/mol



    MINIMIZE USING STEEPEST DESCENT FOR 10000 CYCLES
*    USING NO CROSS, NO MORSE, NO CHARGES
*    UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN 10.0  KCAL/A
    MINIMIZE USING CONJUGATE GRADIENTS FOR 1000 CYCLES
*    USING NO CROSS, NO MORSE, NO CHARGES
*    UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN  0.1  KCAL/A

    INITIALIZE DYNAMICS AT 900 K FOR 10000 STEPS OF 1 FS
*    WITH NO CROSS AND NO MORSE AND NO CHARGES
*    WRITE HISTORY  EVERY 10000 STEPS
*    WRITE AVERAGES EVERY 10000 STEPS

    RESUME DYNAMICS AT 900 K FOR 5000 STEPS OF 1 FS
*    CONSTANT ENERGY
*    WITH NO CROSS AND NO MORSE AND NO CHARGES

!
    nstructs=50
    istruct=0
100  istruct=istruct+1
    RESUME DYNAMICS AT 900 K FOR 10000 STEPS OF 1 FS
*    CONSTANT ENERGY
*    WITH NO CROSS AND NO MORSE AND NO CHARGES
   ARCHIVE AS FILE NUMBER istruct
     IF istruct .LT. nstructs THEN 100

    istruct=0
200  istruct=istruct+1
     RETRIEVE FILE NUMBER istruct

      MINIMIZE USING STEEPEST DESCENT FOR 10000 CYCLES
*      USING NO CROSS, NO MORSE, NO CHARGES
*      UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN 10.0  KCAL/A
      MINIMIZE USING CONJUGATE GRADIENTS FOR 10000 CYCLES
*      USING NO CROSS, NO MORSE, NO CHARGES
```

```
*       UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN  0.001  KCAL/A

        print torsions
        hbdist=3.0
        hbangl=120
        print hbond

        ARCHIVE AS FILE NUMBER istruct
        IF istruct .LT. nstructs THEN 200
!
        END
```

Input File A3.8: Restrained MD run for structure C of Lom-AKH-II.

```
    BEGIN SIMULATION
 reduce output
 constrain using "lom2.rstrnt"
    assign torsion name tor1
*   residue PGLU 1 atom C
*   residue LEU 1B atom N, atom CA atom C
    assign torsion name tor2
*   residue LEU 1B atom C
*   residue ASN 1C atom N, atom CA atom C
    assign torsion name tor3
*   residue ASN 1C atom C
*   residue PHE 1D atom N, atom CA atom C
    assign torsion name tor4
*   residue PHE 1D atom C
*   residue SER 1E atom N, atom CA atom C
    assign torsion name tor5
*   residue SER 1E atom C
*   residue ALA 1F atom N, atom CA atom C
    assign torsion name tor6
*   residue GLY 1G atom C
*   residue TRP 1H atom N, atom CA atom C
    assign torsion name tor7
*   residue PGLU 1 atom O, atom C
*   residue LEU 1B atom N, atom HN
    assign torsion name tor8
*   residue LEU 1B atom O, atom C
*   residue ASN 1C atom N, atom HN
    assign torsion name tor9
*   residue ASN 1C atom O, atom C
*   residue PHE 1D atom N, atom HN
    assign torsion name to10
*   residue PHE 1D atom O, atom C
*   residue SER 1E atom N, atom HN
    assign torsion name to11
*   residue SER 1E atom O, atom C
*   residue ALA 1F atom N, atom HN
    assign torsion name to12
*   residue ALA 1F atom O, atom C
*   residue GLY 1G atom N, atom HN
    assign torsion name to13
*   residue GLY 1G atom O, atom C
*   residue TRP 1H atom N, atom HN
    print torsions
    force tor1 to -149.75 using 100 kcal/rad**2 in molecule 1
*   residue LEU 1B
    force tor2 to -151.29 using 100 kcal/rad**2 in molecule 1
```

```
*    residue ASN 1C
     force tor3 to -147.18 using 100 kcal/rad**2 in molecule 1
*    residue PHE 1D
     force tor4 to -150.02 using 100 kcal/rad**2 in molecule 1
*    residue SER 1E
     force tor5 to -150.11 using 100 kcal/rad**2 in molecule 1
*    residue ALA 1F
     force tor6 to -150.48 using 100 kcal/rad**2 in molecule 1
*    residue TRP 1H
     force tor7 to 180.00 using 100 kcal/rad**2 in molecule 1
*    residue PGLU 1
     force tor8 to 180.00 using 100 kcal/rad**2 in molecule 1
*    residue LEU 1B
     force tor9 to 180.00 using 100 kcal/rad**2 in molecule 1
*    residue ASN 1C
     force to10 to 180.00 using 100 kcal/rad**2 in molecule 1
*    residue PHE 1D
     force to11 to 180.00 using 100 kcal/rad**2 in molecule 1
*    residue SER 1E
     force to12 to 180.00 using 100 kcal/rad**2 in molecule 1
*    residue ALA 1F
     force to13 to 180.00 using 100 kcal/rad**2 in molecule 1
*    residue GLY 1G

     pull atom1 in molecule 1 residue GLY 1G atom HN
*    and atom2 in molecule 1 residue GLY 1G atom HA1 to 3.3
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue GLY 1G atom HN
*    and atom2 in molecule 1 residue GLY 1G atom HA2 to 3.10
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue ASN 1C atom HA
*    and atom2 in molecule 1 residue ASN 1C atom HN to 3.15
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue LEU 1B atom HA
*    and atom2 in molecule 1 residue ASN 1C atom HN to 2.50
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue GLY 1G atom HN
*    and atom2 in molecule 1 residue ALA 1F atom HA  to 3.15
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue SER 1E atom HN
*    and atom2 in molecule 1 residue TRP 1H atom HA to 2.20
*    with a force = 80 kcal/mol
     pull atom1 in molecule 1 residue LEU 1B atom HN
*    and atom2 in molecule 1 residue LEU 1B atom HA to 2.55
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue LEU 1B atom HN
*    and atom2 in molecule 1 residue PGLU 1 atom HA to 2.85
*    with a force = 60 kcal/mol
```

```
      pull atom1 in molecule 1 residue ALA 1F atom HN
*     and atom2 in molecule 1 residue SER 1E atom HA to 2.50
*     with a force = 60 kcal/mol
      pull atom1 in molecule 1 residue PHE 1D atom HN
*     and atom2 in molecule 1 residue PHE 1D atom HA to 3.50
*     with a force = 60 kcal/mol
      pull atom1 in molecule 1 residue SER 1E atom HN
*     and atom2 in molecule 1 residue SER 1E atom HB1 to 3.30
*     with a force = 60 kcal/mol
      pull atom1 in molecule 1 residue GLY 1G atom HN
*     and atom2 in molecule 1 residue SER 1E atom HB2 to 3.50
*     with a force = 60 kcal/mol

      MINIMIZE USING STEEPEST DESCENT FOR 10000 CYCLES
*      USING NO CROSS, NO MORSE, NO CHARGES
*      UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN 10.0  KCAL/A
      MINIMIZE USING CONJUGATE GRADIENTS FOR 1000 CYCLES
*      USING NO CROSS, NO MORSE, NO CHARGES
*      UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN  0.1  KCAL/A

      INITIALIZE DYNAMICS AT 900 K FOR 10000 STEPS OF 1 FS
*      WITH NO CROSS AND NO MORSE AND NO CHARGES
*      WRITE HISTORY  EVERY 10000 STEPS
*      WRITE AVERAGES EVERY 10000 STEPS

      RESUME DYNAMICS AT 900 K FOR 5000 STEPS OF 1 FS
*      CONSTANT ENERGY
*      WITH NO CROSS AND NO MORSE AND NO CHARGES

!
      nstructs=50
      istruct=0
100   istruct=istruct+1
      RESUME DYNAMICS AT 900 K FOR 10000 STEPS OF 1 FS
*      CONSTANT ENERGY
*      WITH NO CROSS AND NO MORSE AND NO CHARGES
     ARCHIVE AS FILE NUMBER istruct
       IF istruct .LT. nstructs THEN 100

      istruct=0
200   istruct=istruct+1
       RETRIEVE FILE NUMBER istruct

         MINIMIZE USING STEEPEST DESCENT FOR 10000 CYCLES
*         USING NO CROSS, NO MORSE, NO CHARGES
*         UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN 10.0  KCAL/A
         MINIMIZE USING CONJUGATE GRADIENTS FOR 10000 CYCLES
*         USING NO CROSS, NO MORSE, NO CHARGES
```

```
*       UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN  0.001  KCAL/A

        print torsions
        hbdist=3.0
        hbangl=120
        print hbond

        ARCHIVE AS FILE NUMBER istruct
        IF istruct .LT. nstructs THEN 200
!
        END
```

Input File A3.9: Restraint file for Lom-AKH-II.

```
!BIOSYM restraint 1
!
#NOE_distance
1:GLY_1G:HN    1:PHE_1D:HA    4.00  15.00  0.000 100.000 100.000  1000.0
1:GLY_1G:HN    1:TRP_1H:HA    4.00  15.00  0.000 100.000 100.000  1000.0
1:SER_1E:HA    1:GLY_1G:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:GLY_1G:HA1   1:TRP_1H:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:GLY_1G:HA2   1:TRP_1H:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:PHE_1D:HA    1:ALA_1F:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:PHE_1D:HN    1:TRP_1H:HA    4.00  15.00  0.000 100.000 100.000  1000.0
1:PGL_1:HA     1:ASN_1C:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:LEU_1B:HA    1:PHE_1D:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:ASN_1C:HA    1:PHE_1D:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:PHE_1D:HA    1:SER_1E:HN    4.00  15.00  0.000 100.000 100.000  1000.0
```

Input File A3.10: Restrained MD run for alternative structure of Lom-AKH-II.

```
    BEGIN SIMULATION
 reduce output
 constrain using "phe.rstrnt"
    assign torsion name tor1
*   residue PGLU 1 atom C
*   residue LEU 1B atom N, atom CA atom C
    assign torsion name tor2
*   residue LEU 1B atom C
*   residue ASN 1C atom N, atom CA atom C
    assign torsion name tor3
*   residue ASN 1C atom C
*   residue PHE 1D atom N, atom CA atom C
    assign torsion name tor4
*   residue PHE 1D atom C
*   residue SER 1E atom N, atom CA atom C
    assign torsion name tor5
*   residue SER 1E atom C
*   residue ALA 1F atom N, atom CA atom C
    assign torsion name tor6
*   residue GLY 1G atom C
*   residue TRP 1H atom N, atom CA atom C
    assign torsion name tor7
*   residue PGLU 1 atom O, atom C
*   residue LEU 1B atom N, atom HN
    assign torsion name tor8
*   residue LEU 1B atom O, atom C
*   residue ASN 1C atom N, atom HN
    assign torsion name tor9
*   residue ASN 1C atom O, atom C
*   residue PHE 1D atom N, atom HN
    assign torsion name to10
*   residue PHE 1D atom O, atom C
*   residue SER 1E atom N, atom HN
    assign torsion name to11
*   residue SER 1E atom O, atom C
*   residue ALA 1F atom N, atom HN
    assign torsion name to12
*   residue ALA 1F atom O, atom C
*   residue GLY 1G atom N, atom HN
    assign torsion name to13
*   residue GLY 1G atom O, atom C
*   residue TRP 1H atom N, atom HN
    print torsions
    force tor1 to -149.75 using 100 kcal/rad**2 in molecule 1
*   residue LEU 1B
    force tor2 to -151.29 using 60 kcal/rad**2 in molecule 1
```

```
*    residue ASN 1C
     force tor3 to -150.48 using 60 kcal/rad**2 in molecule 1
*    residue PHE 1D
     force tor4 to -150.02 using 60 kcal/rad**2 in molecule 1
*    residue SER 1E
     force tor5 to -150.11 using 100 kcal/rad**2 in molecule 1
*    residue ALA 1F
     force tor6 to -147.18 using 60 kcal/rad**2 in molecule 1
*    residue TRP 1H
     force tor7 to 180.00 using 100 kcal/rad**2 in molecule 1
*    residue PGLU 1
     force tor8 to 180.00 using 100 kcal/rad**2 in molecule 1
*    residue LEU 1B
     force tor9 to 180.00 using 100 kcal/rad**2 in molecule 1
*    residue ASN 1C
     force to10 to 180.00 using 100 kcal/rad**2 in molecule 1
*    residue PHE 1D
     force to11 to 180.00 using 100 kcal/rad**2 in molecule 1
*    residue SER 1E
     force to12 to 180.00 using 100 kcal/rad**2 in molecule 1
*    residue ALA 1F
     force to13 to 180.00 using 100 kcal/rad**2 in molecule 1
*    residue GLY 1G


     pull atom1 in molecule 1 residue GLY 1G atom HN
*    and atom2 in molecule 1 residue GLY 1G atom HA1 to 3.4
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue GLY 1G atom HN
*    and atom2 in molecule 1 residue GLY 1G atom HA2 to 3.4
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue GLY 1G atom HN
*    and atom2 in molecule 1 residue SER 1E atom HB1 to 3.4
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue ASN 1C atom HA
*    and atom2 in molecule 1 residue ASN 1C atom HN to 3.15
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue ASN 1C atom HN
*    and atom2 in molecule 1 residue LEU 1B atom HA to 2.60
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue LEU 1B atom HN
*    and atom2 in molecule 1 residue LEU 1B atom HA to 2.8
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue PGLU 1 atom HA
*    and atom2 in molecule 1 residue LEU 1B atom HN to 3.2
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue GLY 1G atom HN
*    and atom2 in molecule 1 residue ALA 1F atom HA  to 3.350
```

```
*   with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue SER 1E atom HA
*   and atom2 in molecule 1 residue ALA 1F atom HN  to 2.70
*   with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue ASN 1C atom HA
*   and atom2 in molecule 1 residue PHE 1D atom HN to 3.4
*   with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue PHE 1D atom HA
*   and atom2 in molecule 1 residue SER 1E atom HN to 2.7
*   with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue PHE 1D atom HA
*   and atom2 in molecule 1 residue ALA 1F atom HN to 2.52
*   with a force = 60 kcal/mol




     MINIMIZE USING STEEPEST DESCENT FOR 10000 CYCLES
*     USING NO CROSS, NO MORSE, NO CHARGES
*     UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN 10.0  KCAL/A
     MINIMIZE USING CONJUGATE GRADIENTS FOR 1000 CYCLES
*     USING NO CROSS, NO MORSE, NO CHARGES
*     UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN  0.1  KCAL/A

     INITIALIZE DYNAMICS AT 600 K FOR 10000 STEPS OF 1 FS
*     WITH NO CROSS AND NO MORSE AND NO CHARGES
*     WRITE HISTORY  EVERY 10000 STEPS
*     WRITE AVERAGES EVERY 10000 STEPS

     RESUME DYNAMICS AT 600 K FOR 5000 STEPS OF 1 FS
*     CONSTANT ENERGY
*     WITH NO CROSS AND NO MORSE AND NO CHARGES

!
     nstructs=80
     istruct=0
100  istruct=istruct+1
     RESUME DYNAMICS AT 600 K FOR 10000 STEPS OF 1 FS
*     CONSTANT ENERGY
*     WITH NO CROSS AND NO MORSE AND NO CHARGES
   ARCHIVE AS FILE NUMBER istruct
     IF istruct .LT. nstructs THEN 100

     istruct=0
200  istruct=istruct+1
     RETRIEVE FILE NUMBER istruct
```

```
      MINIMIZE USING STEEPEST DESCENT FOR 10000 CYCLES
*       USING NO CROSS, NO MORSE, NO CHARGES
*       UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN 10.0  KCAL/A
      MINIMIZE USING CONJUGATE GRADIENTS FOR 10000 CYCLES
*       USING NO CROSS, NO MORSE, NO CHARGES
*       UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN  0.001  KCAL/A

      print torsions
      hbdist=3.0
      hbangl=120
      print hbond

      ARCHIVE AS FILE NUMBER istruct
      IF istruct .LT. nstructs THEN 200
!
      END
```

Input File A3.11: Restraint file for alternative structure of Lom-AKH-II.

```
!BIOSYM restraint 1
!
#NOE_distance
1:GLY_1G:HN    1:PHE_1D:HA    4.00  15.00  0.000 100.000 100.000  1000.0
1:GLY_1G:HA2   1:PHE_1D:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:GLY_1G:HA1   1:TRP_1H:HN    3.40  15.00  0.000 100.000 100.000  1000.0
1:GLY_1G:HA2   1:TRP_1H:HN    3.35  15.00  0.000 100.000 100.000  1000.0
1:GLY_1G:HN    1:TRP_1H:HA    4.00  15.00  0.000 100.000 100.000  1000.0
1:SER_1E:HA    1:GLY_1G:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:PHE_1D:HN    1:TRP_1H:HA    4.00  15.00  0.000 100.000 100.000  1000.0
1:PGLU_1:HA    1:ASN_1C:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:LEU_1B:HA    1:PHE_1D:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:TRP_1H:HN    1:PHE_1D:HA    4.00  15.00  0.000 100.000 100.000  1000.0
1:ASN_1C:HA    1:LEU_1B:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:ASN_1C:HN    1:PHE_1D:HA    4.00  15.00  0.000 100.000 100.000  1000.0
1:SER_1E:HA    1:PHE_1D:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:PHE_1D:HN    1:SER_1E:HB2   4.00  15.00  0.000 100.000 100.000  1000.0
1:PHE_1D:HN    1:SER_1E:HB1   4.00  15.00  0.000 100.000 100.000  1000.0
1:PHE_1D:HN    1:TRP_1H:HA    4.00  15.00  0.000 100.000 100.000  1000.0
1:SER_1E:HA    1:SER_1E:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:ALA_1F:HA    1:ALA_1F:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:PHE_1D:HA    1:PHE_1D:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:SER_1E:HB2   1:ALA_1F:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:SER_1E:HB1   1:ALA_1F:HN    4.00  15.00  0.000 100.000 100.000  1000.0
```

Input File A3.12: Initial high temperature restrained MD run for Lom-AKH-III.

```
    BEGIN SIMULATION
  reduce output
  constrain using "lom3.rstrnt"
     assign torsion name tor1
  *   residue PGLU 1 atom C
  *   residue LEU 1B atom N, atom CA atom C
     assign torsion name tor2
  *   residue LEU 1B atom C
  *   residue ASN 1C atom N, atom CA atom C
     assign torsion name tor3
  *   residue ASN 1C atom C
  *   residue PHE 1D atom N, atom CA atom C
     assign torsion name tor4
  *   residue PHE 1D atom C
  *   residue THR 1E atom N, atom CA atom C
     assign torsion name tor5
  *   residue PRO 1F atom C
  *   residue TRP 1G atom N, atom CA atom C
     assign torsion name tor6
  *   residue TRP 1G atom C
  *   residue TRP 1H atom N, atom CA atom C
     assign torsion name tor7
  *   residue PGLU 1 atom O, atom C
  *   residue LEU 1B atom N, atom HN
     assign torsion name tor8
  *   residue LEU 1B atom O, atom C
  *   residue ASN 1C atom N, atom HN
     assign torsion name tor9
  *   residue ASN 1C atom O, atom C
  *   residue PHE 1D atom N, atom HN
     assign torsion name to10
  *   residue PHE 1D atom O, atom C
  *   residue THR 1E atom N, atom HN
     assign torsion name to11
  *   residue PRO 1F atom O, atom C
  *   residue TRP 1G atom N, atom HN
     assign torsion name to12
  *   residue TRP 1G atom O, atom C
  *   residue TRP 1H atom N, atom HN
     assign torsion name to13
  *   residue THR 1E atom O, atom C
  *   residue PRO 1F atom N, atom CD
      print torsions
     force tor1 to -148.34 using 100 kcal/rad**2 in molecule 1
  *   residue LEU 1B
     force tor2 to -150.93 using 100 kcal/rad**2 in molecule 1
```

```
*    residue ASN 1C
     force tor3 to -148.24 using 100 kcal/rad**2 in molecule 1
*    residue PHE 1D
     force tor4 to -154.50 using 100 kcal/rad**2 in molecule 1
*    residue THR 1E
     force tor5 to -154.50 using 100 kcal/rad**2 in molecule 1
*    residue TRP 1G
     force tor6 to -148.34 using 100 kcal/rad**2 in molecule 1
*    residue TRP 1H
     force tor7 to 180.00 using 50 kcal/rad**2 in molecule 1
*    residue PGLU 1
     force tor8 to 180.00 using 50 kcal/rad**2 in molecule 1
*    residue LEU 1B
     force tor9 to 180.00 using 50 kcal/rad**2 in molecule 1
*    residue ASN 1C
     force to10 to 180.00 using 50 kcal/rad**2 in molecule 1
*    residue PHE 1D
     force to11 to 180.00 using 50 kcal/rad**2 in molecule 1
*    residue PRO 1F
     force to12 to 180.00 using 50 kcal/rad**2 in molecule 1
*    residue TRP 1G
     force to13 to 180.00 using 60 kcal/rad**2 in molecule 1
*    residue THR 1E


     pull atom1 in molecule 1 residue LEU 1B atom HN
*    and atom2 in molecule 1 residue PGLU 1 atom HA to 3.4
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue LEU 1B atom HA
*    and atom2 in molecule 1 residue ASN 1C atom HN to 2.95
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue PRO 1F atom HA
*    and atom2 in molecule 1 residue TRP 1G atom HN to 2.65
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue TRP 1G atom HN
*    and atom2 in molecule 1 residue TRP 1G atom HA to 3.3
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue TRP 1H atom HN
*    and atom2 in molecule 1 residue TRP 1H atom HA to 2.70
*    with a force = 60 kcal/mol
     pull atom1 in molecule 1 residue PHE 1D atom HN
*    and atom2 in molecule 1 residue ASN 1C atom HA to 2.90
*    with a force = 100 kcal/mol
     pull atom1 in molecule 1 residue PHE 1D atom HA
*    and atom2 in molecule 1 residue THR 1E atom HN to 2.60
*    with a force = 100 kcal/mol
     pull atom1 in molecule 1 residue THR 1E atom HB
*    and atom2 in molecule 1 residue THR 1E atom HN to 3.05
```

```
*   with a force = 100 kcal/mol


    MINIMIZE USING STEEPEST DESCENT FOR 10000 CYCLES
*    USING NO CROSS, NO MORSE, NO CHARGES
*    UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN 10.0  KCAL/A
    MINIMIZE USING CONJUGATE GRADIENTS FOR 1000 CYCLES
*    USING NO CROSS, NO MORSE, NO CHARGES
*    UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN  0.1  KCAL/A

    INITIALIZE DYNAMICS AT 600 K FOR 10000 STEPS OF 1 FS
*    WITH NO CROSS AND NO MORSE AND NO CHARGES
*    WRITE HISTORY  EVERY 10000 STEPS
*    WRITE AVERAGES EVERY 10000 STEPS


!
    nstructs=100
    istruct=0
100  istruct=istruct+1
    RESUME DYNAMICS AT 600 K FOR 2000 STEPS OF 1 FS
*    CONSTANT ENERGY
*    WITH NO CROSS AND NO MORSE AND NO CHARGES
   ARCHIVE AS FILE NUMBER istruct
    IF istruct .LT. nstructs THEN 100

    istruct=0
200  istruct=istruct+1
    RETRIEVE FILE NUMBER istruct

      MINIMIZE USING STEEPEST DESCENT FOR 10000 CYCLES
*      USING NO CROSS, NO MORSE, NO CHARGES
*      UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN 10.0  KCAL/A
      MINIMIZE USING CONJUGATE GRADIENTS FOR 10000 CYCLES
*      USING NO CROSS, NO MORSE, NO CHARGES
*      UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN  0.001  KCAL/A

      print torsions
      hbdist=3.0
      hbangl=120
      print hbond

      ARCHIVE AS FILE NUMBER istruct
      IF istruct .LT. nstructs THEN 200
!
      END
```

Input File A3.13: Second lower temperature restrained MD run for sample structures of Lom-AKH-III taken from families of the the previous higher temperature run.

```
    BEGIN SIMULATION
  reduce output
  constrain using "lom3.rstrnt"
     assign torsion name tor1
*    residue PGLU 1 atom C
*    residue LEU 1B atom N, atom CA atom C
     assign torsion name tor2
*    residue LEU 1B atom C
*    residue ASN 1C atom N, atom CA atom C
     assign torsion name tor3
*    residue ASN 1C atom C
*    residue PHE 1D atom N, atom CA atom C
     assign torsion name tor4
*    residue PHE 1D atom C
*    residue THR 1E atom N, atom CA atom C
     assign torsion name tor5
*    residue PRO 1F atom C
*    residue TRP 1G atom N, atom CA atom C
     assign torsion name tor6
*    residue TRP 1G atom C
*    residue TRP 1H atom N, atom CA atom C
     assign torsion name tor7
*    residue PGLU 1 atom O, atom C
*    residue LEU 1B atom N, atom HN
     assign torsion name tor8
*    residue LEU 1B atom O, atom C
*    residue ASN 1C atom N, atom HN
     assign torsion name tor9
*    residue ASN 1C atom O, atom C
*    residue PHE 1D atom N, atom HN
     assign torsion name to10
*    residue PHE 1D atom O, atom C
*    residue THR 1E atom N, atom HN
     assign torsion name to11
*    residue PRO 1F atom O, atom C
*    residue TRP 1G atom N, atom HN
     assign torsion name to12
*    residue TRP 1G atom O, atom C
*    residue TRP 1H atom N, atom HN
     assign torsion name to13
*    residue THR 1E atom O, atom C
*    residue PRO 1F atom N, atom CD
     print torsions
    force tor1 to -148.34 using 100 kcal/rad**2 in molecule 1
*    residue LEU 1B
```

```
    force tor2 to -150.93 using 100 kcal/rad**2 in molecule 1
*   residue ASN 1C
    force tor3 to -148.24 using 100 kcal/rad**2 in molecule 1
*   residue PHE 1D
    force tor4 to -154.50 using 100 kcal/rad**2 in molecule 1
*   residue THR 1E
    force tor5 to -154.50 using 100 kcal/rad**2 in molecule 1
*   residue TRP 1G
    force tor6 to -148.34 using 100 kcal/rad**2 in molecule 1
*   residue TRP 1H
    force tor7 to 180.00 using 50 kcal/rad**2 in molecule 1
*   residue PGLU 1
    force tor8 to 180.00 using 50 kcal/rad**2 in molecule 1
*   residue LEU 1B
    force tor9 to 180.00 using 50 kcal/rad**2 in molecule 1
*   residue ASN 1C
    force to10 to 180.00 using 50 kcal/rad**2 in molecule 1
*   residue PHE 1D
    force to11 to 180.00 using 50 kcal/rad**2 in molecule 1
*   residue PRO 1F
    force to12 to 180.00 using 50 kcal/rad**2 in molecule 1
*   residue TRP 1G
    force to13 to 180.00 using 60 kcal/rad**2 in molecule 1
*   residue THR 1E


    pull atom1 in molecule 1 residue LEU 1B atom HN
*   and atom2 in molecule 1 residue PGLU 1 atom HA to 3.4
*   with a force = 60 kcal/mol
    pull atom1 in molecule 1 residue LEU 1B atom HA
*   and atom2 in molecule 1 residue ASN 1C atom HN to 2.95
*   with a force = 60 kcal/mol
    pull atom1 in molecule 1 residue PRO 1F atom HA
*   and atom2 in molecule 1 residue TRP 1G atom HN to 2.65
*   with a force = 60 kcal/mol
    pull atom1 in molecule 1 residue TRP 1G atom HN
*   and atom2 in molecule 1 residue TRP 1G atom HA to 3.3
*   with a force = 60 kcal/mol
    pull atom1 in molecule 1 residue TRP 1H atom HN
*   and atom2 in molecule 1 residue TRP 1H atom HA to 2.70
*   with a force = 60 kcal/mol
    pull atom1 in molecule 1 residue PHE 1D atom HN
*   and atom2 in molecule 1 residue ASN 1C atom HA to 2.90
*   with a force = 100 kcal/mol
    pull atom1 in molecule 1 residue PHE 1D atom HA
*   and atom2 in molecule 1 residue THR 1E atom HN to 2.60
*   with a force = 100 kcal/mol
    pull atom1 in molecule 1 residue THR 1E atom HB
```

```
*   and atom2 in molecule 1 residue THR 1E atom HN to 3.05
*   with a force = 100 kcal/mol



    MINIMIZE USING STEEPEST DESCENT FOR 10000 CYCLES
*    USING NO CROSS, NO MORSE, NO CHARGES
*    UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN 10.0  KCAL/A
    MINIMIZE USING CONJUGATE GRADIENTS FOR 1000 CYCLES
*    USING NO CROSS, NO MORSE, NO CHARGES
*    UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN  0.1  KCAL/A

    INITIALIZE DYNAMICS AT 300 K FOR 10000 STEPS OF 1 FS
*    WITH NO CROSS AND NO MORSE AND NO CHARGES
*    WRITE HISTORY  EVERY 10000 STEPS
*    WRITE AVERAGES EVERY 10000 STEPS


!
    nstructs=100
    istruct=0
100 istruct=istruct+1
    RESUME DYNAMICS AT 300 K FOR 2000 STEPS OF 1 FS
*    CONSTANT ENERGY
*    WITH NO CROSS AND NO MORSE AND NO CHARGES
  ARCHIVE AS FILE NUMBER istruct
    IF istruct .LT. nstructs THEN 100

    istruct=0
200 istruct=istruct+1
    RETRIEVE FILE NUMBER istruct

      MINIMIZE USING STEEPEST DESCENT FOR 10000 CYCLES
*      USING NO CROSS, NO MORSE, NO CHARGES
*      UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN 10.0  KCAL/A
      MINIMIZE USING CONJUGATE GRADIENTS FOR 10000 CYCLES
*      USING NO CROSS, NO MORSE, NO CHARGES
*      UNTIL THE MAXIMUM DERIVATIVE IS LESS THAN  0.001  KCAL/A

      print torsions
      hbdist=3.0
      hbangl=120
      print hbond

      ARCHIVE AS FILE NUMBER istruct
      IF istruct .LT. nstructs THEN 200
!
      END
```

Input File A3.14: Restraint file for alternative structure of Lom-AKH-III.

```
!BIOSYM restraint 1
!
#NOE_distance
1:PGLU_1:HN    1:TRP_1H:HA    4.00  15.00  0.000 100.000 100.000  1000.0
1:PGLU_1:HN    1:THR_1E:HA    4.00  15.00  0.000 100.000 100.000  1000.0
1:LEU_1B:HA    1:THR_1E:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:PHE_1D:HA    1:ASN_1C:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:LEU_1B:HA    1:LEU_1B:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:ASN_1C:HA    1:ASN_1C:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:LEU_1B:HA    1:ASN_1C:HN    2.80   3.00  0.000 100.000 100.000  1000.0
1:PHE_1D:HA    1:PHE_1D:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:TRP_1G:HA    1:TRP_1H:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:TRP_1H:HA    1:TRP_1G:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:THR_1E:HB    1:TRP_1H:HN    4.00  15.00  0.000 100.000 100.000  1000.0
1:THR_1E:HA    1:PRO_1F:HD1   2.40   3.60  0.000 100.000 100.000  1000.0
1:THR_1E:HA    1:PRO_1F:HD2   2.40   3.60  0.000 100.000 100.000  1000.0
```