Structure Determination of Contrast Agents for Angiogenesis Imaging by High Resolution Nuclear Magnetic Resonance Spectroscopy

Marc Lenoir

A thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh
2006

Declaration

Unless otherwise stated, the work described in this thesis is my own work and has not been submitted in whole or as part for a degree or other qualification at this or any other University.

Abstract

Structure of bicyclic pseudo octapeptides containing a sequence arginine-glycine-aspartic acid (RGD peptides) was investigated by high-resolution NMR spectroscopy. The compounds, synthesized by General Electric Healthcare, are used as contrast agents in imaging of cancer. They bind primarily the $\alpha_v\beta_3$ integrin that is up-regulated during angiogenesis initiated by tumors reaching to existing blood vessels. Their basic peptide skeleton is conjugated with polyethylene glycol moieties of varying length which improve their bioavailability. The peptides also carry a metal binding site that is used to chelate a radioactive metal.

The structures were investigated by standard methods of solution NMR spectroscopy, relying primarily on the determination of distance restraints via NOESY and ROESY experiments. Structures were generated using different protocols in vacuum or explicit solvent. The best results were obtained through the analysis of a series of NOESY spectra acquired using varying mixing times. Full relaxation matrix analysis in explicit solvent was performed on these data yielding a consistent set of structures. The polyethylene glycol moiety and the metal binding site did not show any signs of tertiary structure.

The distance restraints for structure calculations were supplemented by dihedral restraints provided by the analysis of vicinal coupling constants. Towards this end, an extensive set of coupling constants related to dihedral angles of amino acids was obtained. Various experiments, including HMBC, HSQC and HETLOC, were evaluated yielding a basic set of experiments suitable for the measurement of coupling constants of peptides with natural abundance of isotopes.

The structures of the studied peptides were found to be similar and to adopt a reverse γ-turn centered on the aspartic acid of the RGD motif. Despite the bicycling nature of these peptides the compounds showed signs of flexibility, which was more pronounced in H_2O. Side chains of the compounds are flexible as implied from the analysis of the distributions of χ^1 angles.

Parameter such as chemicals shifts; temperature coefficients of H^N protons and Ramachandran plots were evaluated to validate the structural homogeneity of the final structures. Unrestrained molecular dynamics was carried out on the final structures and they were found to be stable during 1 ns runs.

Structures were determined in two solvents, H_2O and DMSO. It was found that the water-based structures are more extended, while the DMSO based structure are more compact. The DMSO-based structure of one peptide was compared with a published NMR structure of a similar peptide and a good agreement was found for the common RGD loop. The water-based structures are similar to a published X-ray structure of a RGD peptide in complex with $\alpha_v\beta_3$ integrin.

Acknowledgements

Thanks to:

My supervisor, Dušan Uhrín, for his patience and involvement throughout this project;

David Grace and Alan Cuthbertson and G.E. Healthcare for an interesting project and financial support,

Dr. David Norman for his help for the full relaxation matrix calculations,

Dr. Michelle Reed for acquiring spectra in Birmingham,

Dr. Tran Pham, Dr. Graeme Ball, Phil Robinson, Chris Fenton, Lan Jin, Bärbel Blaum and Dr. Juraj Bella for making the lab. a nice place to work in,

And, of course, to my parents, cheerful in any circumstances, and my brothers, Eric and Luc.

Contents

List of Figures

List of Tables

Part I

INTRODUCTION

Chapter 1

Diagnostic Tools for Angiogenesis

This project deals with the structure determination of contrast agents used in imaging of angiogenesis associated with cancer. Contrast agents bind to proteins that are up regulated during angiogenesis thereby enhancing the intensity of the radiations at the position of a tumor using radio-tracers. This forms a basis of a non-invasive method for diagnosis of cancer. The following chapter explains the biological relevance of the target proteins, integrins, describes the process of cancer progression and the clinical applications of medical imaging.

1.1 Angiogenesis and cancer

1.1.1 Cancer, a public health problem

Cancer is one of the leading causes of death in the Western world. Despite advances in diagnosis and treatment, the survival of patients remains poor. Constant advances in therapy and diagnosis did not prevent the number of deaths from cancer from increasing over the past fifty years [1]. Cancers can be defined as an abnormal growth of tissue. The malignant neoplasm show a higher degree of anaplasia and has properties of metastasis and invasion compared to benign neoplasms. Malignant cancer cells can escape the primitive tumor, reach either the blood or the lymphatic network, and create metastasis in other locations. Despite these generic features, cancers cover more than one hundred diseases. It is this diversity that is a challenge for early detection and treatment. Frequencies and types of cancer vary with sex, age and geography.

Since many cancers can be treated by surgery, chemotherapy or radiotherapy, chances of a cure increase significantly if cancer is detected in early stages. Early detection of cancer, development of efficient drugs that inhibit or kill cancer cells and the research of specific cancer target molecule are the major topics of cancer research. Because most cancers display angiogenic pathways, when developing diagnostic tools, it is worth targeting angiogenesis rather than the tumor itself.

1.1.2 Angiogenesis and cancer progression

Blood vessels are fundamentally composed of endothelial cells. These cells form tubes which control and direct the blood flow and maintain tissue perfusion. Blood vessels develop via two distinct processes: vasculogenesis and angiogenesis. Vasculogenesis takes place in embryo during the formation of blood vessels. The endothelial cells are formed from precursors, angioblasts. In adults, the blood vessels are produced only through angiogenesis. Angiogenesis is the formation of new blood vessels from pre-existing ones. Vasculature in adult mammals is quiescent under normal conditions except during the female reproductive cycles (ovulation, menstruations, pregnancy). Outside of female reproductive cycles, angiogenesis in adult is largely confined to pathological situations such as wound healing, tumor growth, diabetic retinopathy or inflammation-related diseases [2–6].

In adults, endothelial cells are one of the longest-living cells in the body. The estimated turn over time of endothelial cells is measured in years [2]. The importance of angiogenesis for solid tumor growth and metastasis is well documented. Implants of tumor cells in avascular sites such as the cornea of a rabbit eye, have shown this conclusively. The implants attracted capillaries and the tumor grew [7]. If the implants were prevented from reaching the blood supply or if the angiogenesis was inhibited, the tumor was considerably reduced.

Under non-pathological conditions, angiogenesis is tightly regulated by a balanced action of endogenous angiogenic stimulators and inhibitors. It was postulated that angiogenesis starts when the balance between pro-angiogenic factors (growth factors, cytokines, enzymes, adhesion molecules, metals, hypoxia or peptide fragments) and antiangiogenic factors (proteolytic fragments, cytokines, chemokines, soluble receptors,

collagenase inhibitors, vitamins, tumor suppressor genes) is disturbed by a growing tumor. This is referred to as the "angiogenic switch" [3–5].

Abnormal angiogenesis is a hallmark of cancer progression. It is generally accepted that solid tumors in avascular phase can grow until they reach 2-3 mm (i.e 10^5-10^6 cells), supplied in nutrients and oxygen by diffusion trough the existing vasculature. Beyond this size, tumors promote the formation of blood vessels. Hypoxic stress seems to be one of the principal stimuli for the tumor cells to start the angiogenesis cascade. The access to the blood network has two major consequences: (i) efficiently supplied in nutrient and oxygen, the tumors start growing exponentially and (ii) these new vessels can transport anaplastic tumor cells and form metastasis. Angiogenesis mediated by tumor growth is a complex continuous process involving many different molecules and can schematically be divided into three stages [8, 9]: initiation, invasion/proliferation and maturation/differentiation (Fig. 1.1). During the initiation phase, the tumor cells release angiogenic factors into the extra cellular matrix (ECM).

These factors bind to various receptors of the endothelial cells (tyrosine kinase receptor, vascular endothelial growth factor receptor, etc.) inducing activation signals for the endothelial cells. The activation is characterized by increased proliferation, invasion, differentiation and expression of receptors and secretion of proteolytic enzymes in the ECM. Several molecules have been suggested to promote the invasion and proliferation including selectins and integrins. Proteases such as matrix metalloproteinases or serine proteinases degrade the matrix and allow the endothelial cells to migrate towards the tumor cells. Changes in the ECM can modulate ECM-cell interactions. One group of molecules that can mediate the remodeling of the ECM and is involved in adhesion and proliferation events, are the integrins-cell adhesion receptors. Eventually, the endothelial cells align into vascular cords and form a new lumen for the newly created blood vessels.

The proteolysis of the ECM is associated with the release of angiogenic factors normally sequestered by the ECM. However, it appears now that the degradation of large ECM proteins also play an important part in angiogenesis regulation. These proteins are, under normal conditions, not ligands for the receptors involved in angiogenesis and expose, after proteolysis, the cryptic arginine-glycine-aspartic acid (RGD) sequences

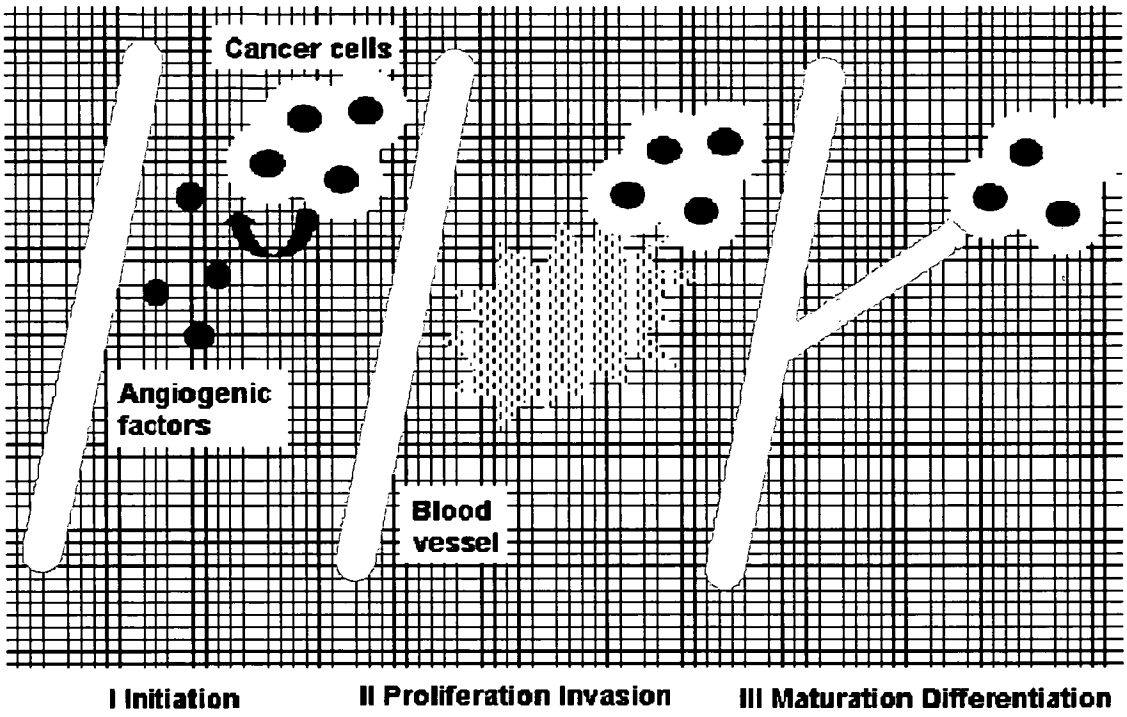


Figure 1.1: **Angiogenic cascade:** Angiogenesis is initiated by the release of angiogenic factors in the ECM by the tumor. The invasion-proliferation is characterized by the modification of the adhesive properties of the endothelial cells and the actions of the proteolytic enzymes that destroy the ECM. The cascade finishes with a maturation-differentiation phase which creates new mature blood vessels.

[2, 9]. These are recognized by several receptors, including integrins during the proliferative and invasive phases. This RGD sequence is found in many proteins of the ECM (vitronectin, fibrinogen, fibronectin, collagen, laminin, etc.). These interactions activate cell adhesions mediated by integrins. Integrins, which are up-regulated during angiogenesis are therefore very good candidates for tumor imaging.

1.2 The integrin family

Integrins are heterodimeric transmembrane glycoproteins, formed by two non-covalently associated α and β subunits. To date, the integrin family is comprised of 24 members which are the result of combinations of at least 18 α and 8 β subunits (these numbers vary depending on the author). The subunits arise from alternative splicing of mRNA and further postranslational modifications [10]. Integrins have various functions: they are involved in cell-cell, cell-ECM interactions and signal transduction [11–13]. Their

binding specificity is determined by the combination of subunits. Sometimes the same ligand is recognized by several integrins [10, 11, 14, 15]. For instance, the RGD sequence is a common primary binding motif for various integrins [14, 16] including α_v containing integrins and $\alpha_{IIb}\beta_3$, $\alpha_5\beta_1$, $\alpha_8\beta_1$. Cells control the activity of integrins, although the details of this regulation are unclear. This mechanism is particularly important for white cells and platelets. The regulation of adhesion allows white cell to migrate in the body until they meet the stimulus that will activate their anchoring. The platelets will express/activate β_3 subunit through a contact with damaged blood vessels [17]. Many integrins are associated with angiogenesis, α_v integrins in particular [8, 16, 18].

In order to localize angiogenesis, one must use a target which is upregulated and present at significant levels at the sites of angiogenesis and genetically stable. With this regard the integrin $\alpha_v\beta_3$ is a potential marker for angiogenesis.

The $\alpha_v\beta_3$ integrin and its interactions with the RGD motif

In adults, the $\alpha_v\beta_3$ integrin (or vitronectin receptor) has a limited distribution and appears at low levels in several cells including intestinal cells, vascular cells, activated leukocytes and macrophages. On the other hand, endothelial cells in tissues undergoing angiogenesis present high levels of $\alpha_v\beta_3$ integrin [19–21] and hence represent a potent marker of angiogenesis. The role of this integrin during angiogenesis has now been investigated for two decades. $\alpha_v\beta_3$ was first thought to directly promote angiogenesis and ligands were designed to block its activity. However, experiments carried out by Hynes et al. [22, 23] showed that the presence of gene coding for the β_3 subunit in mice is not essential for neovascularization and in fact the absence of this gene leads to an increase of angiogenic response. The authors suggested that $\alpha_v\beta_3$ acts as a trans dominant inhibitor over other integrins of the epithelial cells involved in angiogenesis, namely $\alpha_1\beta_1$, $\alpha_5\beta_1$ and $\alpha_2\beta_1$. In the absence of integrin $\alpha_v\beta_3$ the inhibition can not occur, which explains the increased angiogenic behavior of mice lacking $\alpha_v\beta_3$ integrins. As pointed out by Carmeliet [24], the lack of β_3 integrins could be compensated through other mechanisms and the author concluded that our understanding of the role of integrins in angiogenesis is incomplete.

Vitronectin receptor is a type I transmembrane protein with an extracellular N-terminus and cytoplasmic C-terminus consisting of a 125 kDa α_v subunit and a 105 kDa β_3 subunit. The vitronectin receptor shares common features with other members of the integrin family i.e. a seven helix-protein with a short cytoplasmic tail and a large extracellular domain. Figure 1.2 schematically shows the different parts of this integrin.

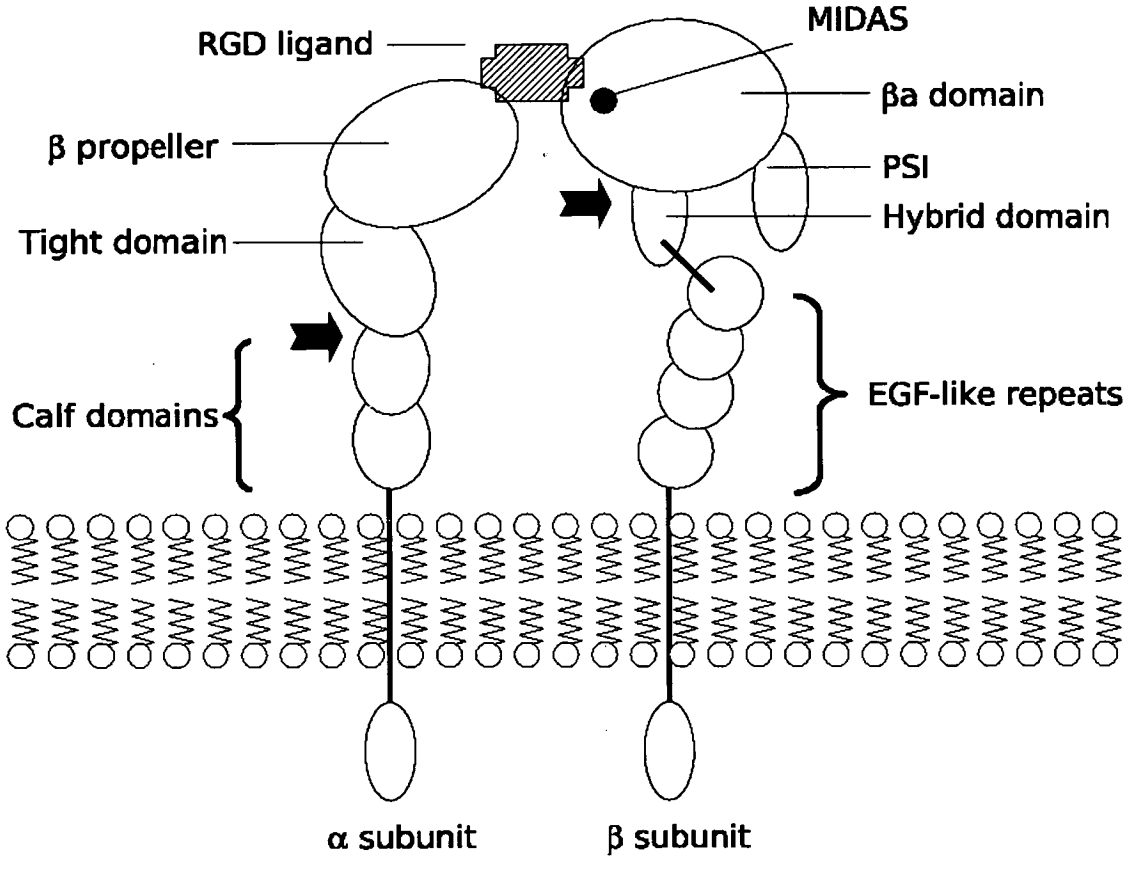


Figure 1.2: **Schematic representation of the $\alpha_v\beta_3$ integrin:** The α subunit is composed from calf domains (β-sandwich domains), a tight domain (Ig-like domains) and a β propeller domain. The β subunit consists of a plexin, semaphorin and integrin (PSI) domain, repeats of epidermal growth factor (EGF) domains and a βA domain. The black arrows indicate the borders of the C-terminal construct that yielded the crystal structure [25].

Both α and β units are required for the formation of an RGD binding site. The structure of an extracellular segment of $\alpha_v\beta_3$ has been solved at 3.1 Å by X-ray crystallography with and without an RGD ligand, cyclo(RGDf-N(Me)V)[1] also called Cilengitide

[1] Small letter "f" stands for D-phenylalanine

[25–27]. This provides an important insight into the mode of binding and the changes occurring upon the formation of the complex.

The heterodimer is formed by two nearly parallel tails that are ended by a "head" which brings both parts together at an angle of 135^o creating a V shape binding site. This head forms the putative RGD binding site by participation of a seven bladed β propeller domain from α_v and a β_A domain from β_3 (Fig. 1.3). The β_A contains three divalent ions which are part of the metal ion-dependent adhesion site (MIDAS), an adjacent MIDAS (AMIDAS) and a ligand-associated metal binding site (LIMBS)[25]. In these sites the cations are coordinated by a characteristic sequence of amino acids (DXSXS) [28].

In the RGD-integrin complex , the RGD motif bridges both units of the $\alpha_v\beta_3$ integrin. The arginine guanidium group forms salt bridges with the carboxylate groups of Asp218 and Asp120 of the α subunit of the propeller. The side chain of the arginine is partially exposed to the solvent. The carboxylic group of the aspartic acid of the ligand is totally buried in the integrin and its side chain has hydrophobic contacts with the side chain of Arg214 of the β_A domain. One of the oxygens of carboxylate group of the peptide forms a polar interaction with Mn^{2+} of MIDAS in the β_A domain. The other oxygen forms hydrogen bonds with the backbone amide protons of Tyr122 and Asn215. The glycine residue lies at the interface between the two subunits forming hydrophobic interactions. The LIMBS and AMIDAS sites do not interact with the ligand but stabilize the conformation of the β_A domain [25]. It has also been suggested that the phenyl group of the RGD ligand forms a π-stacking interaction with the side chain of Tyr122 [29].

The RGD motif is the minimal recognition sequence as confirmed by cross-linking experiments [30] and crystallographic studies [25]. The RGD sequence is the cornerstone of many molecules aimed at targeting integrins, $\alpha_v\beta_3$ in particular.

1.3 RGD targets and imaging of angiogenesis

The strategy, which aims at using the newly formed endothelial cells during tumor formation for therapy/imaging, is attractive for several reasons: (i) endothelial cells,

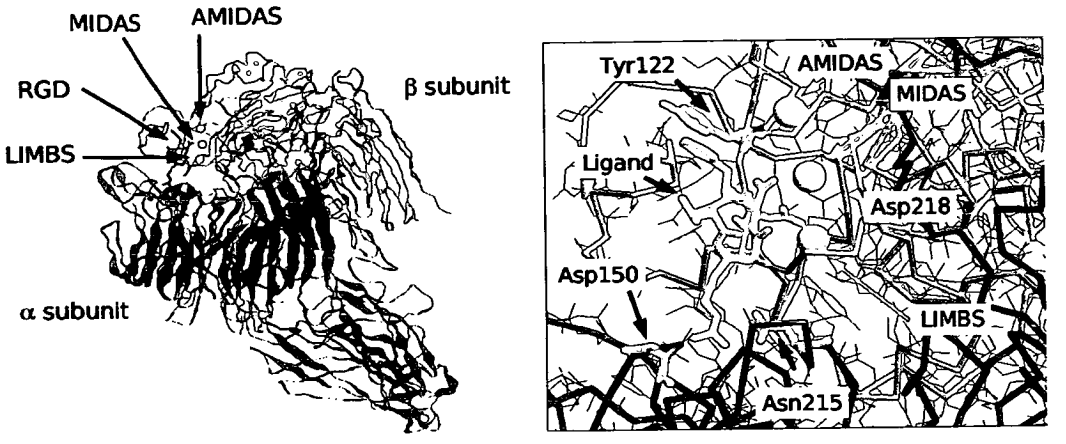


Figure 1.3: **Structure of the integrin $\alpha_v\beta_3$ in complex with an RGD ligand:** Structure solved by Xiong et al.[25] of the extracellular segment of integrin $\alpha_v\beta_3$ with RGD ligand and Mn^{2+} ions. The RGD ligand (green) bind mutually the α (blue) and β (cyan) subunits. The expansion of the binding site is shown on the right. The backbone and the side chains of the integrin are shown as thick lines and narrow lines, respectively. The residues indicated directly interact with the RGD ligand.

although contributing to tumor progression, are normal, stable cells ; (ii) the endothelium is readily accessible for commonly used therapeutics via the blood network. Cyclopeptides, containing the RGD sequence and RGD-like compounds, have been investigated to improve their affinity and specificity for targeted integrins. Mainly $\alpha_v\beta_3$ has been studied for angiogenesis related to cancer progression. Nevertheless other targets have been defined including $\alpha_v\beta_5$ or β_1 integrins [31]. RGD-based molecules have three potential applications in cancer therapy [11]:

- Diagnosis and prognosis: tumor imaging.
- Therapeutics: inhibition of angiogenesis with antagonists.
- Site-directed delivery, conjugating the RGD ligand with other drugs for optimal efficacy and safety [32, 33].

Numerous RGD-mimicking or RGD-containing peptides have been synthesized to specifically target the $\alpha_v\beta_3$ integrin at the surface of activated endothelial cells. However, the RGD recognition is shared by several integrins and it appears that subtle conformational changes in the RGD motif are responsible for integrin specificity. The specificity and affinity of the ligands are often compared with that of vitronectin, a protein natu-

rally occurring in the ECM (the $\alpha_v\beta_3$ integrin was formerly called vitronectin receptor). Their specificity is established by binding studies using various integrins. Cyclic peptides are advantageous since they resist exopeptidases and at the same time restrain the conformation of RGD loops. Incorporation of amino acid analogues based on RGD templates or D-amino acids also prevent degradation of the molecules [6, 27, 34–41].

Imaging techniques are non-invasive methods for *in vivo* investigations of tissues. Angiogenesis can be observed by using magnetic resonance imaging, computed tomography, positron emission tomography, ultrasonography and optical imaging (see reference [42] for review). Imaging of angiogenesis with RGD-based molecules makes use of radioactive labels that are employed within the context of standard medical methods such as scintigraphy (^{18}F, ^{99}Tc, ^{111}In or ^{90}Y) [43–45] or positron emission tomography (^{64}Cu or ^{123}I) [46, 47]. These techniques can locate sites of angiogenesis, the amount of blood vessel growth and elucidate features that distinguish normal vessels from angiogenic ones. RGD containing cyclic peptides must therefore be modified in order to carry radioactive elements.

The pharmacokinetics can also be improved by conjugation of polyethylene glycol (PEG), hydrophilic amino acids or sugar moieties to the peptides [48]. PEG also serves to increase the molecular mass of small molecules to avoid excretion via the renal route. PEG protects molecules from enzymatic degradations by its steric effects and increases the solubility of compounds in water. PEG is not toxic and does not induce immune response [49]. To date, two $\alpha_v\beta_3$ binding agents have entered clinical trials:

- The humanized monoclonal antibody LM609 to $\alpha_v\beta_3$ (Vitaxin) developed at the Scripps Research Institute, La Jolla, California, USA and licensed to Applied Molecular Evolution, San Diego, California, USA and MedImmune, Inc., Gaithersburg, Maryland, USA for cancer treatment.

- Cilengitide, EMD 121974 or EMD 85189 from Merck KgaA, Darmstadt, Germany a synthetic cyclic peptide mimicking the RGD sequence, cyclo(L-arginyl-glycyl-L-aspartyl-D-phenylalanyl-N-N-methyl-L-valyl) tested for cancer treatment

1.4 Aims of the work

G. E. Healthcare has developed a series of bicyclic octapeptides aimed as contrast agents for imaging of angiogenesis during cancer progression [50]. They are conjugated with PEG moieties, in order to improve the bioavailability, and with a chelating site to carry a radioactive metal for the detection by imaging techniques.

The aims of this study are:

- Solution structure determination of series of RGD-based peptides designed by GE Healthcare using NMR spectroscopy.
- Assessment of possible conformation changes of the peptide core as a result of the presence of functional groups.
- Assessment of the flexibility of the peptides in water and DMSO.
- A comparison with existing NMR or X-ray structures of related compounds.
- Binding interaction of the peptides with $\alpha_v\beta_3$ integrin using automated docking.

Chapter 2

High Resolution Nuclear Magnetic Resonance

We used Nuclear Magnetic Resonance (NMR) as a tool to study the conformation of small, cyclic peptides. This technique is non-destructive and allows to work with mg quantities of material. NMR is capable of providing three dimensional coordinates in solution and also indicating flexibility. For small constrained peptides, flexibility might be an important part of the binding process [51]. The following chapter describes the nuclear magnetic resonance phenomenon and focuses on the methods that were employed during this investigation.

2.1 The nuclear magnetism

Energy levels and populations

NMR spectroscopy relies on the magnetic properties of nuclei. In general, NMR of peptides makes use of ^{1}H, ^{13}C, ^{15}N nuclei which all possess spin 1/2, thus, only spin 1/2 will be considered when placed in a strong magnetic field ($\vec{B_0}$), spins with a spin quantum number 1/2 will create a two-state distribution with different energies. According to Boltzmann distribution, the equilibrium population of spins at the two levels can be calculated as:

$$\frac{N_\beta}{N_\alpha} = e^{(\frac{-\Delta E}{kT})} \tag{2.1}$$

where N_α and N_β are the number of nuclei aligned with and against the external magnetic field, ΔE is the energy gap between these states, k is the Boltzmann's constant

and T is the absolute temperature in Kelvin. Spins will precess around the external magnetic field which is aligned with the z-axis. Associated with the spin angular momentum is the collinear nuclear magnetic dipole moment ($\vec{\mu}$) defined by:

$$\vec{\mu} = \gamma \vec{I} \tag{2.2}$$

where γ is the gyromagnetic ratio. The projection of $\vec{I}$ onto the z axis (I_z) can take $2I + 1$ discrete values in units of $\hbar = h/2\pi$ where h is the Planck's constant. For spins $1/2$ nuclei I_z can take only two values of $\pm\hbar/2$. The energy of the interaction of a spin magnetic moment $\vec{\mu}$ with the external magnetic field $\vec{B_o}$ is defined by:

$$E = -\vec{\mu}.\vec{B_o} \tag{2.3}$$

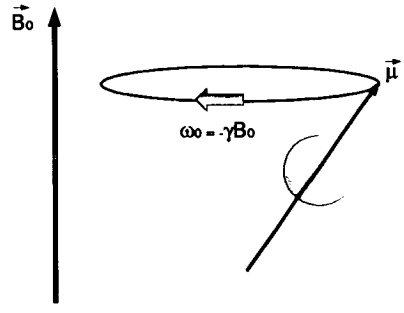


Figure 2.1: **Spin behavior in a static magnetic field:** A spin with a positive gyromagnetic ratio precesses clockwise at Larmor frequency (ω_0) around B_0. The magnetic $\vec{\mu}$ moment is indicated by an arrow.

Using a projection onto the z axis, the energy gap between the two states can be calculated as $\Delta E_{\beta\to\alpha} = E_\alpha - E_\beta = -\hbar\gamma B_0$. When combined with Planck's equation $\Delta E = h\nu$, the Larmor frequency (ν_0) of nuclear precession is obtained:

$$\nu_0 = -\frac{\gamma B_0}{2\pi}, \tag{2.4}$$

sometimes expressed using the angular velocity as $\omega_0 = -\gamma B_0$. At a macroscopic level, precessing spins will create a "bulk magnetization" along the z axis (M_z) due to the excess of their population at a lower energy level (Fig. 2.2).

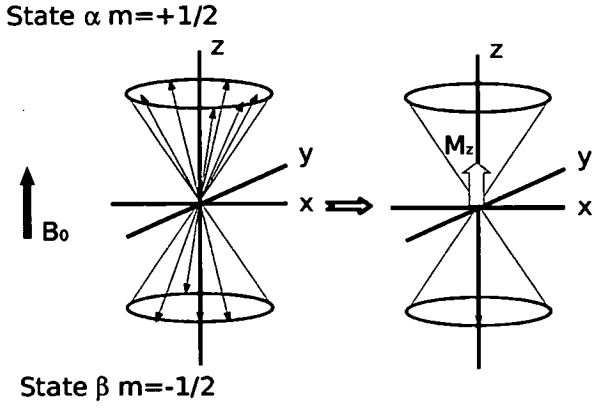


Figure 2.2: **Population difference and bulk magnetization:** The population difference between α and β states is created in the external magnetic field B_0. The resulting bulk magnetization (M_z) is indicated by a grey arrow.

This magnetization is the source of NMR signal and its size is related to the population difference between the two states. Combining equations 2.1 and 2.4, an expression that the incorporates gyromagnetic ratio can be derived:

$$\frac{N_\beta}{N_\alpha} = e^{(\frac{-\gamma h B_0}{2\pi kT})} \approx 1 - \frac{\gamma h B_0}{2\pi kT} \tag{2.5}$$

Using this equation, the sensitivity of different nuclei can be assessed. When calculating the relative sensitivity, the natural abundance of individual nuclei must also be considered (Table 2.1).

Table 2.1: Nuclear properties of nuclei used in this study.

Nucleus	Natural abundance (%)	Gyromagnetic ratio ($\times 10^8$)	Relative sensitivity
^{1}H	99.98	2.67	1.000
^{13}C	1.11	0.67	0.016
^{15}N	0.36	-0.27	0.001

2.2 Chemical shift

Each nucleus experiences a magnetic field that is the sum of the external magnetic field and local magnetic fields originating primarily from bonding electrons, but also from more remote sources. The local magnetic field resulting from shielding of nuclei can be described as:

$$B_{local} = B_0(1 - \sigma) \tag{2.6}$$

where σ is the shielding constant. The shielding constant is affected by surrounding atoms whose contribution can be divided into diamagnetic and paramagnetic components:

$$\sigma = \sigma_{diamagnetic} + \sigma_{paramagnetic} \tag{2.7}$$

For example, the magnetic field B_0 induces a diamagnetic current in the phenyl ring which opposes the external magnetic field. As a result, the protons in the region above or below the plane of the ring are shielded and the atoms in the plane of the ring are deshielded.

The shielding effects lead to dispersion of signals of the same nucleus in NMR spectra, opening a possibility for assigning the observed resonance to individual atoms. In order to compare spectra acquired at different field strengths, the chemical shift (δ) scale has been introduced. This is a relative scale describing the position of a signal relative to the position of a reference compound,

$$\delta(ppm) = \frac{\nu - \nu_{ref.}}{\nu_{ref.}} \times 10^6 \tag{2.8}$$

where ν and $\nu_{ref.}$ are the absolute frequencies.

Temperature dependence of chemical shifts

Temperature dependence of the amide protons is often used in peptides to assess the solvent accessibility of these protons and to identify donors of hydrogen bonds [27, 38, 52–54]. Due to the reasonable dispersion of the amide protons and a small number of residues in peptides, the temperature dependence is readily accessible using standard 1D proton spectra. Variations of chemical shifts versus the temperature are often linear

and calculated slopes are referred to as the temperature coefficient ($\Delta\delta_{H^N}/\Delta T$), often quoted in parts per billion per degree Kelvin (ppb.K^{-1}).

The temperature dependence of amide proton chemical shifts is related to the inverse third power of the distance between the donor and the acceptor [52]. It is assumed that, in proteins and peptides, all amide protons are involved either in inter or intra molecular interactions. The strength of these interactions decreases with increased molecular motion as a result of higher temperature. The average donor-acceptor distance increases causing amide proton chemical shift to move to their random coil chemical shifts. For those involved in hydrogen bonds, the effect induced by the acceptor persists and the protons are less temperature dependent. However, the temperature dependency of amide protons is also influenced by their environments [53]. The changing effect of the aromatic ring current as a result of the temperature change can lead to false interpretations of the temperature coefficients [55]. Finally exchange phenomena have to be considered when interpreting the experimental results.

2.3 Scalar couplings

NMR signals acquire a more complex character through the interaction of spins with close-by NMR active nuclei. These through-bond interactions are mediated by bonding electrons occupying overlapping orbitals and are quantified by scalar coupling constants (J couplings). The effect of scalar couplings can be illustrated conveniently by inspecting two coupled nuclei, A and X. Their spins can orient in a parallel or antiparallel fashion with respect to the external magnetic field yielding the following four combinations $A_\alpha X_\alpha$, $A_\alpha X_\beta$, $A_\beta X_\alpha$ or $A_\beta X_\beta$. Parallel orientations of spins are less energetically favorable than the antiparallel ones and their energies expressed in Hz are consequently incremented by +1/4J, while the energy of the antiparallel orientations will decrease by the same amount. Focusing on spin A, the frequency of its transition is not uniform any more and depending on the state of spin X will either increase or decrease by 1/2J (Fig. 2.3). Energy levels of spin X are changed in the same way as those of spin A producing splitting that also corresponds to the value of J.

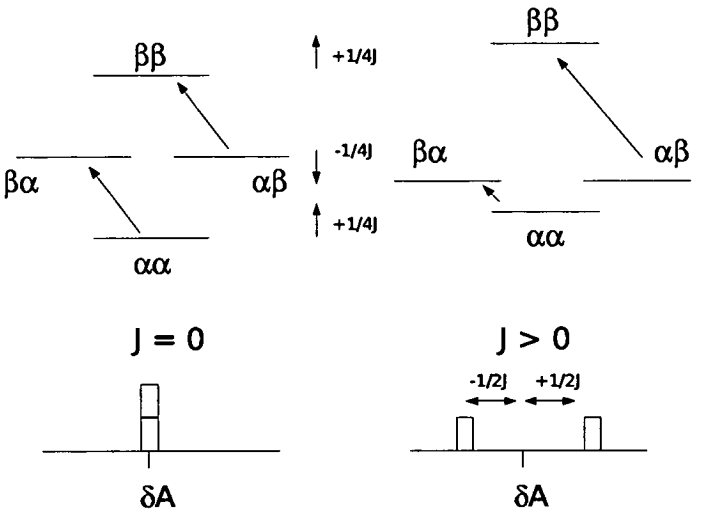


Figure 2.3: Energy diagram of a J-coupled two spin system. Shown are the transitions for spin A $\alpha\alpha \rightarrow \beta\alpha$ and $\alpha\beta \rightarrow \beta\beta$ for J=0 and J$\neq$0 together with corresponding spectra. The coupling constant arises due to change of energy levels caused by the presence of spin X.

Scalar coupling constants and geometry

The existence of J couplings indicate that the coupled nuclei are separated by few bonds which provides immediate structural information and forms the basis for numerous NMR experiments. Potentially even more useful for three dimensional structure determination of molecules, the sizes of coupling constants can be related to geometrical parameters. In particular the three-bond coupling constants can be related to dihedral angles via the semi-empirical Karplus relation.

The Karplus equation has the following general form:

$$^3J = Acos^2(\theta + \Delta) + Bcos(\theta + \Delta) + C \tag{2.9}$$

where A, B and C are adjustable parameters reflecting the effect of substituents and Δ is the phase difference between the ϕ, ψ and χ^1 dihedral angles and dihedral angles corresponding to the measured coupling constants. Karplus equations have been parametrized for ϕ, ψ and χ^1 angles of peptides. These dihedral angles are defined as C'_{i-1}-N-C^α-C', N-C^α-C'-N_{i+1} and N-C^α-C^β-C^γ (C^γ →S, cysteine a) for ϕ, ψ and χ^1 angles. The values of the Karplus parameters have been parametrized and are reviewed

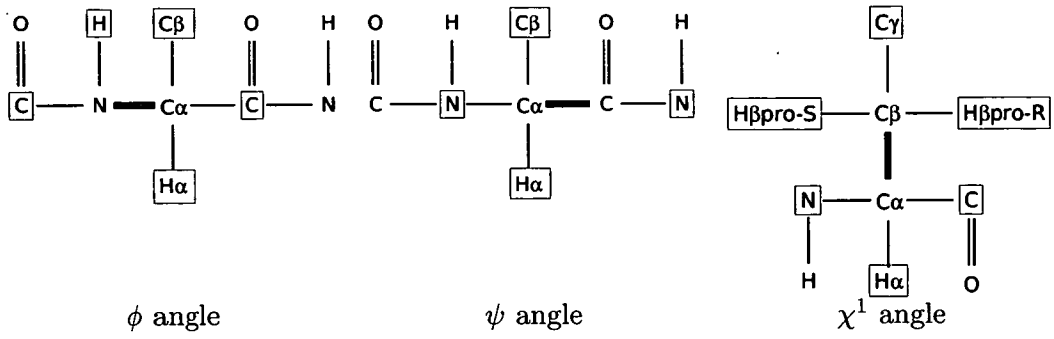


Figure 2.4: **Selected dihedral angles of peptides:** Central nuclei defining a particular dihedral angle are connected by a thick line. The boxes indicate atoms whose coupling constants can be related to the particular dihedral angle.

elsewhere [56, 57].

The coupling constants related to the χ^1 angle can also be used to determine the stereospecific assignment of the prochiral protons of the side chains. Side chains can adopt one of the three staggered conformations shown below.

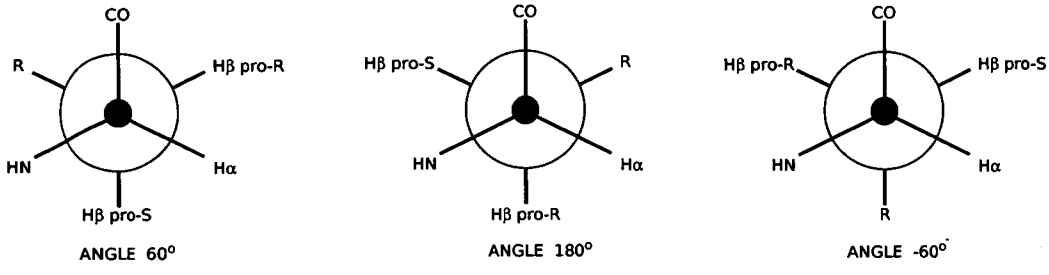


Figure 2.5: Newman projection of staggered conformations described by the χ^1 dihedral angle

The rotamer population is commonly calculated via the Pachler equations that assume that only trans and gauche conformations (180° and ±60°) are present. The model assumes that the observed coupling constant is a weighted sum of the coupling constants of individual rotamers:

$$J_{observed} = aJ(-60) + bJ(180) + cJ(60) \tag{2.10}$$

where a, b and c are the population fractions (a+b+c=1). Pachler's equation uses values of coupling constants synclinal and anticlinal to calculate the ratio of rotamers

based on two different experimental coupling constants. This leads to stereospecific assignment of prochiral protons.

2.4 Fourier transform NMR

Fourier transform NMR spectroscopy uses series of pulses to manipulate the macroscopic magnetization and rotate it from its equilibrium position. It is convenient to look at the evolution of magnetization using a reference frequency. Spins rotating at the same frequency as the reference frequency will seem to be immobile in this frame. This frame of reference is called the rotating frame and its frequency is identical to the carrier frequency. Application of a radiofrequency pulse (r.f.) from the transverse plane will flip the magnetization by an angle α:

$$\alpha = \gamma B_1 t_p / 2\pi \tag{2.11}$$

where, B_1 is the strength of the r.f. pulse and t_p the length of the pulse. In NMR pulse sequences 90^o and 180^o pulses are the most commonly used. Once **M** is moved away from the z-axis, the system is perturbed and must return to its Boltzmann's distribution of energy states. This is achieved through relaxation phenomena.

The energy absorbed by the spins is released to the environment or the lattice; this phenomenon is called longitudinal or spin-lattice relaxation. It is characterized by the time constant, T_1:

$$\frac{dM_z}{dt} = -\frac{M_z - M_0}{T_1} \tag{2.12}$$

where $R_1 = 1/T_1$ is the rate constant of the transition of the perturbed system to its equilibrium state. In addition to the longitudinal relaxation (z magnetization), the loss of phase coherence of the x-y component of the magnetization is characterized by the transverse relaxation sometimes called spin-spin relaxation. This process does not lead to any net energy transfer between spins. As spins return to equilibrium, the transverse component decays exponentially to zero at the rate $R_2 = 1/T_2$.

In FT NMR, the receiver detects oscillating electric currents induced by the rotating macroscopic magnetization in the x-y plane, which can be described as a mixture of sine and cosine waves decaying via spin-spin relaxation to zero. This signal is called the free

induction decay (FID). Fourier transform of the FID gives rise to the NMR spectrum with the peaks observed at frequencies relative to the carrier frequency. Detection of both real and imaginary components of signals allows determination of a relative position of the peaks with regard to the carrier position. Several operations can be performed on FIDs prior to Fourier transform in order to enhance either the resolution of the peaks or to increase the signal-to-noise ratio of spectra.

Weighting functions. A signal-to-noise ratio enhancing weighting function accelerates the decay of FID giving more weight to the initial parts of FID which contain less noise. Many functions are available for this purpose (exponential, sine, sine squared, etc.). However, if the signal decays too fast, the linewidth will increase. If the signal resolution is the priority, the latter parts of FID can be enhanced, leading to resolution enhancement. This is typically achieved by Gaussian functions. Two parameters of the Gaussian function can be optimized: the maximum of the Gaussian function (GB), and the line broadening (LB). The latter is set to correspond to the linewidths of the FT signal, while the first, expressed as an increasing fraction of the acquisition time, provides progressively stronger resolution enhancement.

Zero filling. The digital resolution of spectra can be increased by a procedure called zero filling. This consists of adding points beyond the end of the FID with zero intensity and leads to spectra with increased digital resolution. This is particularly useful for accurate measurement of small coupling constants performed in this work.

2.5 Two-dimensional NMR

Two-dimensional experiments can be represented by a general scheme (Fig. 2.6) containing four periods: the preparation, evolution, mixing and acquisition periods.

The second dimension is generated by incrementation of the evolution period by Δt_1. In this way, numerous FIDs will be acquired, modulated by various NMR parameters in t_1 and t_2. Two successive Fourier transforms of this 2D array of FIDs will create a two-dimensional spectrum (Fig. 2.7).

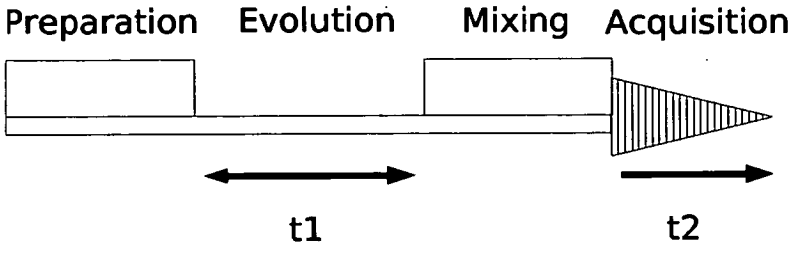


Figure 2.6: **General scheme of 2D NMR experiments:** The preparation period, also containing the relaxation interval, prepares spins into a desired state for the evolution interval. Here, either spin-spin couplings and/or chemical shifts are coded. Mixing transfers the magnetization between spins; this is detected in the directly-detected dimension during acquisition.

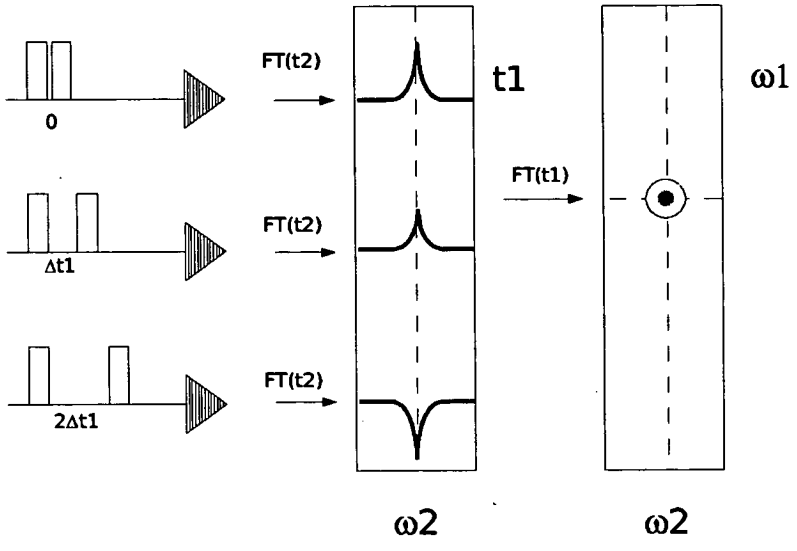


Figure 2.7: **Schematic representation of a 2D NMR experiment:** The signal acquired during t_2 is also modulated during t_1. The first FT creates a 2D interferogram, with peaks at their Larmor frequencies in F_2 modulated by different (cross peak) or identical (auto peak) Larmor frequencies in t_1. The second Fourier transform generated a 2D spectrum with peaks at their chemical shifts in both dimensions.

2.6 Polarization transfer

Homonuclear polarization transfer allows the magnetization to be transferred between two identical nuclei. The coherent transfer of magnetization between J couplings is behind the powerful 2D experiments used for resonance assignments. The incoherent transfer utilizing dipolar coupling leads to detection of spacial proximity between protons.

Polarization transfers play an important role in multidimensional NMR experiments. Heteronuclear transfer of magnetization between protons and nuclei with lower gyromagnetic ratio, dramatically increases the sensitivity of experiments involving heteronuclei. INEPT (Insensitive Nuclei Enhanced by Polarization Transfer) makes use of the scalar couplings between protons a heteronuclei such as ^{13}C or ^{15}N. Via creation of an antiphase magnetization, the INEPT pulse sequence transfers the magnetization between these different types of nuclei. Heterocorrelated experiments benefit greatly from two successive INEPT transfers thus utilizing the more favorable Boltzman distribution of protons and detection of signal at higher frequencies. This, for example, leads to sensitivity increase of ^{15}N detection by a factor of 300.

In INEPT type experiments single quantum coherences are utilized for chemical shift labeling of heteronuclei. HMQC (heteronuclear multiple-quantum correlation) uses multiple quantum coherence for chemical shift labeling.

2.7 Pulse field gradients

Pulse field gradients (PGF) are short, typically millisecond, pulses during which the static magnetic field B_0 is made inhomogeneous. This disturbance is not random. The strength of PFGs varies linearly along a certain axis, typically the z axis. During PFGs, spins experience different magnetic fields depending on their position in the NMR tube, resulting in a loss of coherence for any magnetization not aligned along the z axis. This magnetization, temporarily lost, can be recovered by applying an identical gradient with opposite sign, or as in polarization transfer experiments, with an appropriately scaled gradient. Gradients are also used for purging of unwanted magnetization, for example, after storing the useful magnetization along z-axis or by manipulating selected signals by selective pulses. The latter principle is used for water suppression in the WATERGATE experiment [58] and double-pulse field gradient spin echo (DPFGSE) experiment [59]. Gradients are also used for coherence selection e.g. in HMBC (Hetero multiple-bond correlation) experiments used in this work.

2.8 The Nuclear Overhauser Enhancement (NOE)

Structure calculations based on NMR data make an extensive use of the distance-intensity relationship contained in the NOE enhancements. The following section presents the NOE phenomenon and explains how it can be influenced by the motion of the molecule or the magnetic field of the spectrometer [60–62].

In principle, spin systems can return to equilibria via stimulated relaxation or spontaneous relaxation. It can be shown that spontaneous relaxation is too slow to account for the observed relaxation [60]. Relaxation is caused by local fluctuating magnetic fields that have the required frequencies to induce transitions.

NOE is a significant part of relaxation processes whereby spins return to thermal equilibria. NOE involves concerted flips of spins mediated by their mutual dipolar interactions and contributes to the changes of the longitudinal magnetization. It is therefore part of the longitudinal relaxation processes. The efficiency of NOE is closely related to the motion regime of the molecules in solution which is characterized by the rotational correlation time (τ_c). The rotational correlation time is defined as the time required to rotate the molecule from its initial position by 1 radian and can be approximated by the Debye expression for isotropic molecular tumbling:

$$\tau_c = \frac{4\pi\eta a^3}{3kT} \tag{2.13}$$

The rotational correlation time is measured in seconds (τ_c); a is the radius of a spherical molecule in m, η the viscosity of the solvent ($N.s.m^{-2}$), and T is the temperature in Kelvin. This rotational correlation time varies with the size of the molecule (tens of ps for small molecules, tens of ns for large proteins) or with the viscosity of the solvent for the same molecule (e.g. DMSO and water). Molecular motion determines the correlation function (G(t)) which, when Fourier transformed, yields the spectral density J(ω) function:

$$J(\omega) = \bar{B}_{loc}^2 \frac{2\tau_c}{1+\omega^2\tau_c^2} \tag{2.14}$$

also expressed via a reduced spectral density j(ω) as $J(\omega) = \bar{B}_{loc}^2 j(\omega)$ where $\bar{B}_{loc}$ are the local magnetic fields.

For a two spin system with dipolar interacting nuclei I and S (Fig 2.8), the relaxation

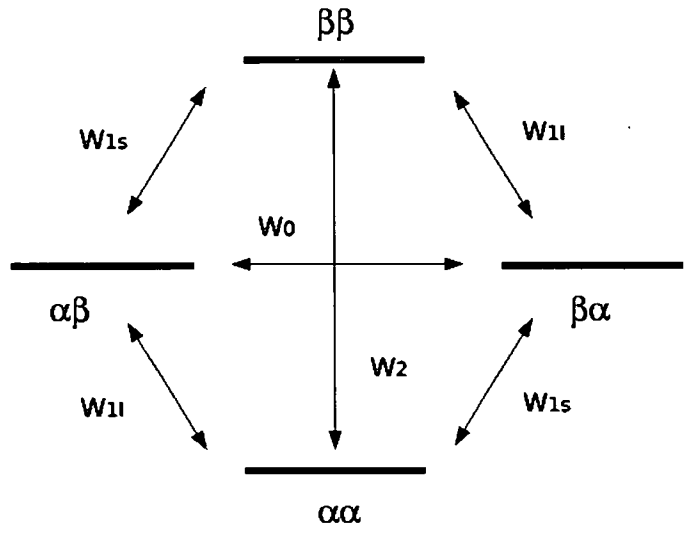


Figure 2.8: **Energy levels and transition probabilities for a two spin-system, IS:** Transition rate constants W_2 and W_0 correspond to transitions in which both spins are flipped, $\alpha\alpha \leftrightarrow \beta\beta$ and $\beta\alpha \leftrightarrow \alpha\beta$.

rates can be described by the Solomon equation. For I:

$$\frac{dI_z}{dt} = -(I_z - I_z^0)R_I - (S_z - S_z^0)\sigma_{IS} \tag{2.15}$$

where I_z and S_z represent the non-equilibrium z-magnetization of spin I and S, while I_0 and S_0 represent their equilibrium magnetization. $R_I = 2W_{1i} + W_2 + W_0$ is the self relaxation rate and $\sigma_{IS} = W_2 - W_0$ the cross relaxation rate. This cross relaxation rate describes the rate of dipole-dipole transitions giving rise to NOE enhancement.

The above relaxation rates are related to the spectral density via the following relationship:

$$W_{ij} = A_{ij}b^2 j(\omega_{ij}) \tag{2.16}$$

where A_{ij} is a coefficient derived from quantum mechanics, b^2 is related to the magnitude of the local field and $j(\omega_{ij})$ is the reduced spectral density. Considering two nuclei of the same kind, the rates are:

$$\begin{aligned} W_{1i} &= 1.5b^2 j(\omega_i) \\ W_0 &= b^2 j(0) \\ W_2 &= 6b^2 j(2\omega_i) \end{aligned} \tag{2.17}$$

where $b = (\mu_0\gamma^2)/(\hbar\pi r_{ij}^{-3})$, μ_0 is the permitivity in vacuum, γ is the gyromagnetic ration and r_{ij} is the distance separating the nuclei i and j. The above equations,

illustrate that the NOE enhancement depends on the molecular motion of the molecules characterized by the rotational correlation time, τ_c, the frequency of the spectrometer (ω_0) and the interatomic distance r_{ij}.

This motional model considers molecules as rigid spheres tumbling isotropically, which is not generally the case. To account for internal motion, Lipari-Szabo's often referred to as "model free" approach introduced an order parameter (S) and an internal correlation tine (τ_e) into the correlation function. If we call τ_{eff} the effective correlation time with $\tau_{eff}^{-1} = \tau_c^{-1} + \tau_e^{-1}$, the reduced spectra density becomes:

$$j(\omega) = S^2 \frac{\tau_c}{1+\omega^2\tau_c^2} + (1-S^2)\frac{\tau_{eff}}{1+\omega^2\tau_{eff}^2} \tag{2.18}$$

Where S^2 accounts for the rotation of the inter-proton vector relative to a molecule axis. If the molecule is rigid, $S^2 = 1$, and the movements of the vector follows only from the overall molecular motion. On the other hand, if $S^2 = 0$, a completely isotropically movement of the vector is superimposed on the movement of the molecule. If we consider that the internal motion is much faster than the overall tumbling ($\tau_e \ll \tau_c$). Thus, τ_{eff} is comparable to τ_e and the second term of the equation 2.18 becomes negligible. This is the approximation used in NOE based calculations (see chapter 3).

NOE in multispin system

In a multispin system, the cross relaxation between protons leads to apparent cross relaxation between protons that are not coupled by dipolar coupling. This also causes lowering of the intensity of NOE cross peaks. To account for these effects, arising particularly at long mixing times, the peak intensities can be calculated at any mixing time using a matrix of the relaxation rates. For a multispin system in transient NOE experiments, the Solomon equation can be written as:

$$\frac{dI_{nz}}{dt} = -(I_{nz} - I_{nz}^0)R_{In} - \sum_m (I_{mz} - I_{mz}^0)\sigma_{mn} \tag{2.19}$$

which can be expressed in a matrix form as:

$$\frac{d[\mathbf{a}(\tau_m)]}{d\tau_m} = -\mathbf{R}\mathbf{a}(\tau_m) \tag{2.20}$$

These differential equations can be solved to give:

$$\mathbf{a}(\tau_m) = e^{(\mathbf{R}\tau_m)}\mathbf{a_0} \tag{2.21}$$

where $\mathbf{a}(\tau_m)$ is the matrix of cross-peak intensities at mixing time τ_m, $\mathbf{R}$ is the relaxation matrix and $\mathbf{a_0}$ is the diagonal matrix of equilibrium which corresponds to the volumes of the diagonal peaks at (τ_m)=0.

$$\mathbf{R} = \begin{bmatrix} R_1 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1n} \\ \sigma_{21} & R_2 & \sigma_{23} & \cdots & \sigma_{2n} \\ \sigma_{31} & \sigma_{32} & R_3 & \cdots & \sigma_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_{n1} & \sigma_{n2} & \sigma_{n3} & \cdots & R_n \end{bmatrix} \tag{2.22}$$

The diagonal elements of the relaxation matrix are related to the longitudinal relaxation rates and the off-diagonal elements are the cross relaxation rate constants. The eigenvalue λ and the eigenvector χ of the matrix are then used to calculate the intensity of the NOE at any given mixing time.

$$\mathbf{a}(\tau_m) = \chi e^{(-\lambda\tau_m)}\chi^{-1}\mathbf{a_0} \tag{2.23}$$

where $\mathbf{a}$ is the matrix whose elements give the cross peak intensity between i and j, χ are the eigenvectors of the rate matrix describing the relaxation behavior of the system of spins. λ are the diagonal matrix elements giving the eigenvalues for the system.

This method can be used to calculate cross peak intensities in NOESY spectra. Given a motional model (using parameters τ_c and S^2) the reduced spectral densities can be calculated. This leads to the definition of relaxation rate constants of the relaxation matrix and cross peak intensities for any given mixing time. In practice, as not all the intensities can be obtained, the process iteratively compares the result of the calculated intensities with the observed ones [63, 64].

Nuclear Overhauser spectroscopy (NOESY)

The NOE enhancements are most conveniently detected by 2D NOESY experiments. In NOESY experiments, cross peaks arise between dipolarly coupled spins. At short mixing times, zero-quantum (ZQ) coherences generated during the NOESY experiment give rise to cross peaks between J coupled protons, which are superimposed on top of the NOESY cross peaks in a form of antiphase and dispersive peaks. NOE enhancements are the main source of geometrical information and the quality of the data at short mixing times is essential for structure calculations. Several methods have been described in the literature for the removal of the unwanted zero-quantum coherences.

The most efficient method uses a combination of swept-frequency 180^o pulse and a low intensity pulsed-field gradient [65]. The advantage of this method is that it works in one scan. The gradient leaves the z magnetization, which is present during the NOESY mixing time, unaffected. The swept-frequency 180^o pulse flips the spins at different positions in the sample at different times. Hence, the ZQ coherences acquire different phases as a function of the position of spins in the sample resulting in their averaging to zero.

NOE enhancement in the rotating frame

Rotating frame NOE experiment (ROESY) provides a possibility of observing rotating frame NOE enhancements (ROEs). A weak magnetic field B_1 is applied along the axis perpendicular to B_0 (e.g. y axis) resulting in spins precessing around an effective field that now forms an angle θ with the z axis. The consequence is that the extreme narrowing limit always applies ($\omega\tau_c \ll 1$) and the enhancements are thus always positive [61]. This is an alternative to NOESY experiments for molecules when the NOE enhancements are close to zero. ROESY experiments suffer from TOCSY transfers and off-resonance effects. The off-resonance effects arise because the tilt of each spin depends on its Larmor frequency. This angular dispersion affects the quantitative exploitation of cross peaks and their interpretation in terms of distances. A way to circumvent these effects is to apply the off-resonance irradiation alternatively at high and low frequencies. This averages the angular dispersion and makes the spins experience similar effective fields irrespective of their Larmor frequency. It has been shown when the off-resonance irradiations are placed at more than $1.5\delta_m$ the TOCSY transfers and the off-resonance effects are significantly decreased (δ_m is the frequency difference between the carrier frequency and the maximum spectral frequency) [66, 67].

Rotating frame nuclear Overhauser spectroscopy (ROESY)

ROESY experiments are an alternative to NOESY experiments when the NOE enhancements are weak. The first 90^o pulse of ROESY pulse sequence is followed by a weak spin-locking whose length is equivalent to the mixing time of the NOESY experiment. The irradiation can either be applied at the carrier frequency, i.e. in the

middle of the spectra, or off-resonance at equal and opposite frequencies with respect to the carrier frequency. The experiment using such double off-resonance irradiations is referred to as T-ROESY. This modification reduces the coherent magnetization transfer mediated by scalar couplings (TOCSY transfer) and the off-resonance effect which depends on the effective spin-lock angle.

2.9 Homonuclear methods

2.9.1 Correlation spectroscopy (COSY)

The COSY experiment is used to establish pairs of protons that have a mutual scalar coupling. The cross peaks in COSY spectra have a fine structure. The active coupling, which gave rise to the cross peak, is in antiphase ($sin(\pi Jt_1)$ modulation), while all the other passive couplings appear in phase ($cos(\pi Jt_1)$ modulation). The spectra can be used to identify, step by step, the protons belonging to the same spin system. COSY experiments are typically acquired with a double quantum filter (DQF). The advantage of the DQF-COSY experiment over a simple COSY experiments is that both diagonal and cross peaks can be phased to pure absorption. The DQF coherence pathway can be selected by applying two gradient pulses contained within spin echoes which remove the offset evolution taking place during the PGFs.

2.9.2 Total correlation spectroscopy (TOCSY)

The TOCSY experiment transfers the magnetization between protons that are not directly coupled. This experiment allows facile identification of protons belonging to the same spin-system. The advantage of TOCSY, compared to COSY, is that TOCSY cross peaks are in-phase, avoiding attenuation of signals seen in COSY spectra as a consequence of the antiphase nature of COSY cross peaks. In the TOCSY experiment, a strong magnetic field is applied along the y axis. This field locks the spins, effectively removing their chemical shift differences, initiating the magnetization transfer. In practice, multi-pulse sequences, such as DIPSI-2 [68], are used to achieve an efficient and uniform spin-lock within a complete range of proton chemical shifts. The length of the spin-lock, referred to as the mixing time, dictates the number of remote protons

to which the magnetization is transferred. This parameter has to be chosen carefully to detect all the protons of the spin system. erence, on or off resonance.

2.10 Heteronuclear methods

2.10.1 Heteronuclear single quantum correlation (HSQC)

HSQC - a double INEPT experiment - provides one bond chemical shift correlations between X (^{15}N or ^{13}C) nuclei and protons. Its main use is in the assignment of heteronuclei resonances, providing the proton signals have been assigned. In cases where a complete proton assignment is not possible via homonuclear methods, mainly because of overlaps, capitalizing on the dispersion of ^{13}C chemical shifts, the HSQC experiment greatly assists in completing this task. ^{1}H-^{15}N HSQC is also the method of choice for the determination of ^{1}H-^{1}H coupling constants. Due to the quadrupolar relaxation of ^{14}N, doublets are broadened considerably. Together with overlap of H^N signals, this makes coupling constant determination from 1D ^{1}H spectra often impossible. The quadrupolar broadening is absent in ^{15}N molecules and the dispersion of ^{15}N chemical shifts removes the overlap problem.

2.10.2 Heteronuclear multiple bond correlation (HMBC)

Heteronuclear long-range correlations are provided by HMBC experiments which utilize the evolution of multiple quantum coherences to achieve the heteronuclear chemical shift labeling. These experiments were extensively used in this project and will therefore be described here in detail.

The cross peaks in HMBC experiments show a mixed phase due to the evolution proton-proton couplings and proton chemical shift during the preparation delay. The shape of HMBC cross peaks is further complicated by the long-range heteronuclear coupling appearing in antiphase. As a result, HMBC multiplets have a complex phase making the determination of heteronuclear coupling constants difficult. Several methods have been proposed to achieve this task. They rely on fitting of the cross sections extracted from the HMBC by using templates [69–71].

Two approaches, used in this project, have been proposed by Edden *et al.*[69]. Long-range heteronuclear coupling constants can be determined by reconstructing the mixed-phase multiplet by using either 1D spectra or one-bond correlation cross peaks as templates. The latter are recorded simultaneously with long-range correlations in gradient-enhanced HMBC experiments (ge-HMBC).

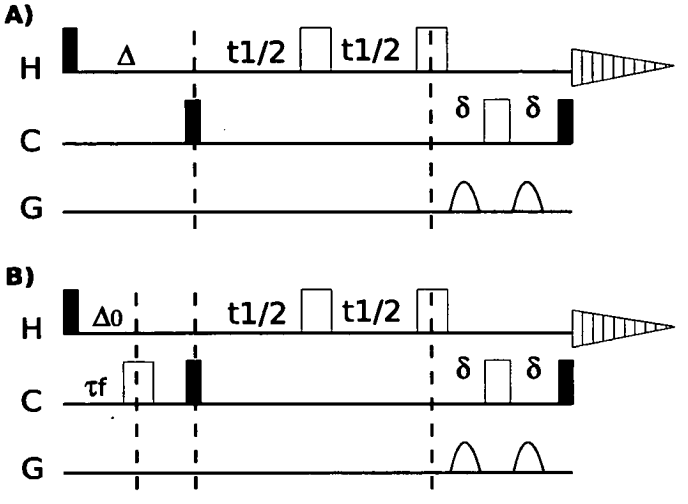


Figure 2.9: **Pulse sequences of HMBC experiments used for multiplet reconstruction:** Open and filled rectangles represent 180^o and 90^o pulses, respectively. The proton 180^o placed in the middle of the t_1 period refocuses proton chemical shifts. The final proton 180^o pulse ensures that the phase of HMBC multiplets is identical to the phase of left-shifted multiplets obtained from 1D TOCSY experiments. Figure A shows the pulse sequence used in conjunction with 1D templates. Figure B shows the pulse sequence with one-bond multiple-bond filter at the beginning of the sequence. This experiment also provides one-bond cross peaks that are used as templates.

Reconstructing the multiplets by simulating the mixed phase: In the first approach, the ge-HMBC cross peaks are reproduced by multiplets extracted from z-filtered 1D TOCSY spectra. These multiplets, excised from 2D spectra, are inverse Fourier transformed and the resulting FIDs are left-shifted by the length of the preparation delay Δ and the time accommodating the gradients (2δ) (Fig. 2.9 A). Followed by Fourier transformation, the phase properties of these 1D multiplets are identical to those of HMBC multiplets with respect to the homonuclear couplings and chemical shifts. Two such proton multiplets are then gradually shifted from their chemical shift positions in opposite directions and subtracted. In this way a reconstructed HMBC

multiplet is obtained. A fitting procedure is employed that minimizes the differences between HMBC and a reconstructed multiplet. The displacement yielding the best fit corresponds to the amplitude of the long-range heteronuclear coupling constant.

This method requires pure phased multiplets that are manipulated and compared with the HMBC multiplets in the frequency domain. A series of 1D TOCSY spectra are acquired using the same acquisition parameters as those used to acquire the HMBC spectrum (acquisition time, dwell time, offset and spectral width). The HMBC signal (S_{HMBC}) can be described as follows:

$$S_{HMBC}(t2) = A_{HMBC} \times S_{proton}(t2 + \Delta) \times sin(\pi J_{XH}(t2)) \tag{2.24}$$

where A_{HMBC} is the amplitude of the S_{HMBC} cross peak extracted from the from the F_2 dimension of an HMBC spectrum, S_{proton} is the proton multiplet without heteronuclear coupling and Δ is the evolution delay of the HMBC. The factor $sin(\pi J_{XH} t_2)$ converts in-phase proton into a multiplet with a heteronuclear coupling in antiphase. The fitting procedure reproduces the HMBC multiplets by using the signal, which is equivalent to an inphase multiplet acquired with a preacquisition delay:

$$S_{1D} = 90° - \Delta - Acquisition \tag{2.25}$$

The S_{1D} is generated using 1D TOCSY spectra. A proton multiplet of interest is extracted from a 1D spectrum and an inverse Fourier transform generates its the time domain signal. This FID is then left-shifted by the value of Δ in the time domain and Fourier transformed again. The resulting peak has an identical phase modulation as the cross peaks extracted from the HMBC spectrum and the scaling factor.

$$S_{HMBC} = A_{trial} \times ((S_{1D} + \frac{J_{XH_{trial}}}{2}) - (S_{1D} + \frac{J_{XH_{trial}}}{2})) \tag{2.26}$$

The extraction of the coupling constants is performed in the frequency domain. The trial J coupling is introduced and the inphase multiplet is reconstructed by shifting (J/2) and subtracting two S_{1D} multiplets. At the same time the scaling factor A_{trial} is optimized. The in-house fitting program "decoupantiphase" performs a Powell minimization of the function given by Eq. 2.27 in a two-dimensional space of the coupling constant.

$$\mathcal{A} = \sum_{i=Jmin}^{Jmax} A_{trial} \sqrt{[(S_{proton_0} - S_{proton_i} - S_{HMBC})^2]} \tag{2.27}$$

where Jmin and Jmax are the limits for the minimum and the maximum coupling constants, respectively. The program shifts S_{proton_i} point by point and compares it to S_{HMBC}. At the minimum value of $\mathcal{A}$, the program returns the value of the coupling constant and the value of A_{trial}.

Reconstructing the multiplets from one-bond templates: The second approach makes use of one-bond cross peaks which have the same phase properties as the long-range cross peaks. This reflects identical evolution of proton-proton couplings and chemical shifts for both types of cross peaks. HMBC spectra containing only long-range or one-bond cross peaks are obtained by manipulating four HMBC experiments in which a one-bond multiple-bond (OBMB) filter is implemented. The OBMB filter makes use of the difference between the values of the long-range and one-bond coupling constants (2-5 Hz and ~140 Hz, respectively). Small changes in the overall evolution delay will not affect the intensity of the long range couplings whereas these modifications will have a definite effect on the evolution of one bond couplings. The separation of one-bond and multiple-bond cross peaks relies on a movable 180^o ^{13}C pulse applied during the Δ_0 delay (Fig. 2.9B). This does not affect the evolution of the proton-proton couplings which evolve during $\Delta_0 + \Delta$ delay in all cases. On the other hand, the heteronuclear couplings will be affected by this pulse which forms a spin-echo and refocuses the heteronuclear couplings. Different position of this pulse leads to different effective evolution via $^1J_{HX}$. If Δ_0 is chosen to be $(2 \times (^1J_{XH})^{-1}$ or τ_1, two experiments acquired with the ^{13}C 180^o pulse placed a times 0 and τ_1 will result in an evolution of $^1J_{XH}$ equal to $(\Delta + \tau_1)$ and $(\Delta - \tau_1)$, i.e. the difference is $1/^1J_{XH}$. Hence, their subtractions yield a spectrum containing one-bond cross peaks while their sum yields a spectrum containing only the long-range cross peaks. However, due to the high frequency of the one-bond coupling evolution, it is possible that a particular setting of intervals could lead to zero intensity of one-bond cross peaks in both spectra. Another pair of experiments is therefore required in order to guarantee that, at least in one of the two pairs of HMBC spectra, a particular one-bond cross peak will be found. This is achieved by using $\tau_0 = (4 \times {}^1J_{XH})^{-1}$. Using the following delays in four experiments with $\tau_f = (0, \tau_0, \tau_1, \tau_1 + \tau_0)$ and labeling the experiments a, b, c and d, a combination of $(a - c) \pm (b - d)$ yields spectra containing one-bond

cross peaks whereas $a + b + c + d$ yields long-range cross peaks only.

Rather than simulating the phase-modulated multiplets from pure phase TOCSY multiplets, the templates used to extract the coupling constants could be acquired simultaneously with the long-range cross peaks. The one-bond cross peaks appearing in the HMBC will have acquired the same phase modulation as the long range ones. Both kinds of cross peaks contain heteronuclear coupling constants. In the case of the one-bond correlations, the antiphase components of cross peaks are separated by large $^1J_{CH}$ coupling constants. Their two parts are therefore modulated only by proton-proton coupling constants and proton chemical shifts and can be taken individually to reconstruct the long-range correlation cross peaks. The coupling constants are determined by minimizing the following equation.

$$\mathcal{A} = |S_{coupled} - A_{trial}(S_{one-bond_{leftshift}} - S_{one-bond_{rightshift}})| \tag{2.28}$$

where $S_{one-bond}$ is the one-bond template. The amplitude of the coupling constant corresponds to the difference in Hz between the right-shifted and the left-shifted templates.

The preparation delays (typically 60 to 120 ms) used in the HMBC experiments are tuned to long-range coupling constants (typically < 6 Hz in peptides). During this delay large one-bond coupling constants modulated the proton magnetization in an unpredictable manner. In order to ensure that some antiphase magnetization with respect to the one-bond coupling constants is generated before the conversion to multiple-quantum coherences, a one-bond multiple-bond filter (OBMB) needs to be set up. The filter is calibrated for a one-bond coupling constant. Four separate experiments are acquired with an effective evolution of one-bond coupling constants set to 3 μs, 1/4 ^{1}J, 2/4 ^{1}J and 3/4 ^{1}J. This ensures, that in one pair of experiments, the one-bond correlation cross peaks will be present and, within a pair, their phase will be opposite. At the same time, such small changes of evolution intervals do not affect the long-range cross peaks. which do not change their phase. Separation of the one-bond and the long-range cross peaks is therefore possible, which helps to remove overlaps.

The HMBC experiment can also be used in a different way for the measurement of long-range couplings. This method proposed by Uhrín *et al.* [72] uses two refocused

HMBC experiments yielding the heteronuclear coupling constants as in-phase components of HMBC multiplets. These otherwise identical experiments are acquired with and without carbon decoupling and pure phase multiplets with respect to heteronuclear couplings are obtained by applying purging ^{13}C pulses immediately before acquisition.

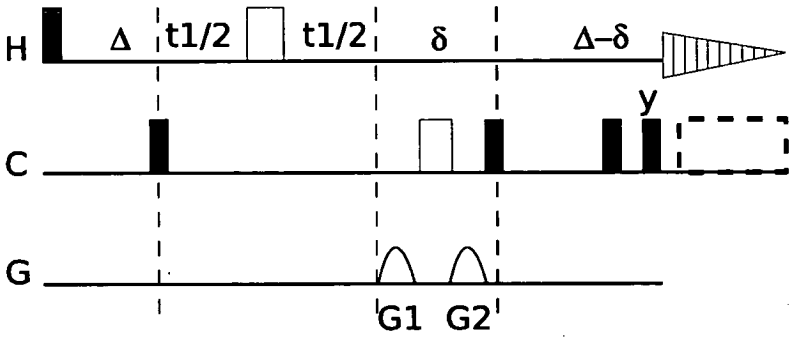


Figure 2.10: **Pulse sequence of refocused HMBC experiments:** See Fig. 2.9 for meaning of the symbols. The gradients G1 and G2 are set to -50% and 30%, and -30% and 50%, for real and imaginary data, respectively. The broad-band decoupling, indicated by a dashed box, is applied during the acquisition time of the decoupled experiment.

As a result, decoupled cross peaks will be twice as intense as the coupled ones. Knowing the scaling between cross peaks extracted from the two spectra facilitates the multiplet reconstruction process and the coupling constant determination. The reconstructed multiplet is obtained by shifting, in opposite directions, and adding two identical decoupled multiplets. The fitting process minimized the difference between the reconstructed and coupled multiplet. The coupling constants are calculated by minimizing the following equation:

$$\mathcal{A} = |S_{coupled} - 0.5(S_{decoupled_{leftshift}} + S_{decoupled_{rightshift}})| \tag{2.29}$$

The decoupled refocused HMBC experiment can also be used to estimate the values of coupling constants based on cross peak intensities. The intensities of cross peaks in refocused HMBC experiments is proportional to $sin^2(\pi J\Delta)$ where Δ is the preparation and the refocusing delay. For one particular proton, if one heteronuclear coupling

constant (J_A) is known, another coupling constant (J_B) can be calculated as:

$$\begin{aligned} A &= I_0 sin^2(\pi J_A \Delta) \\ B &= I_0 sin^2(\pi J_B \Delta) \\ J_B &= asin\left(\sqrt{sin^2(\pi J_A \Delta) \times (B/A)}\right)/\pi\Delta \end{aligned} \tag{2.30}$$

where A and B are the cross peak intensities. This equation leads to two solutions. In most of the cases, one of these solutions can be safely discarded by inspection of the splitting in the coupled spectrum.

2.10.3 Heteronuclear long range coupling (HETLOC)

Although categorized as a heteronuclear method, both axis of HETLOC spectra have proton chemical shifts. HETLOC experiment provides a mean to measure heteronuclear coupling constants which are smaller than the linewidth of the signals. This experiments is a (ω_1) half-filtered TOCSY. It produces E.COSY (Exclusive COSY) pattern of cross peaks split in F_1 by one-bond couplings and shifted in F_2 by the value of $^nJ_{XH}$ coupling constant. Not only can the amplitude of the couplings be measured but also their sign, by comparing the direction of the displacement of cross peaks with those of the diagonal peaks. The latter correspond to one-bond coupling constants, the sign of which are known. The original experiment proposed by Kessler [73] has been modified by Uhrín *et. al* [74]. In this modification, ^{13}C or ^{15}N attached protons are selected by a gradient-improved X-filter. A gradient enhanced bilinear rotation decoupling (G-BIRD) pulse applied in the middle of the t_1 periods refocuses the homonuclear couplings of CH protons leaving only the proton chemical shifts and the one bond couplings to evolve. This removes the proton-proton coupling constants in F_1 improving resolution and sensitivity. The t_1 period is followed by a proton-proton TOCSY block which transfers the magnetization to coupled protons of the same spin system. The heteronuclear long-range coupling constants are easily determined by shifting the two parts of the same cross peaks. The drawback of this method is that it relies on proton-proton couplings to build the magnetization of remote protons coupled to the heteronucleus. This method is therefore not suitable for the measurement of inter-residue coupling constants.

2.11 Structure calculations using NMR data

Molecular structures are calculated using molecular mechanics based on Newtonian physics. The potential energy of the system is calculated based on molecular mechanics force fields. This potential energy is the sum of individual energies that arise from covalently bonded atoms (bonds length, valence angles, dihedral angles), non-bonded interactions (van der Waals and electrostatic terms) and experimental restraints:

$$E_{total} = \sum E_{vdw} + E_{elec} + E_{bonds} + E_{angles} + E_{impropers} + E_{experimental} \quad (2.31)$$

This equation evaluates the energy as a function of the molecular geometry. Each term calculates the energy contribution arising from a deviation of individual parameters from its equilibrium value as described in the force field. The total energy without the experimental energies is often referred to as the empirical energy. The calculations make use of several variables and so a multidimensional energy landscape must be explored. This potential surface energy is minimized during structure calculation.

For large systems like peptides or proteins many minima exist for the potential energy and the aim of the calculation is is to sample a large conformational space in order to localize the global minimum. Molecular dynamics (MD) which uses Newton's law of motion is suitable for conformation searches in large systems. Kinetic energy is introduced to the simulation and Newton's law of motion is solved for each atom. Each atom moves in proportion to its kinetic energy and the force-field potential energy. This molecular dynamics can be combined with a simulated-annealing scheme (MDSA) which improves the search for global minima. At high temperature, large conformational changes are possible and as the temperature decreases, the structures are frozen out. These molecular dynamics steps are followed by minimizations that regularize the structure within the minima found by molecular dynamics [75–77].

Solutions can be modeled using explicit or implicit solvents, periodic boxes for instance. These calculations are aimed at reproducing the conditions in solution and therefore, generating more realistic structures. Electrostatic interactions are neglected in calculations performed in vacuum where the non-bonded interactions are governed by the volume of the atoms (Van der Waals radii) and purely repulsive forces. Calculations in solvents also account for the attractive forces, by means of atomic charges

and Leonard-Jones potentials. Such calculations should therefore lead to more realistic structures. The drawbacks of this approach are the computational cost and the fact that it is not possible to heat the system to temperatures of 1000 K or 2000 K, ordinarily used in simulated annealing protocols. This prevents random coordinates being used as a starting point for structure calculations. Because of this, calculations performed in solvents are used to refine structures.

2.11.1 Restraining function for experimental data

The structure determination is carried out by means of restrained minimization and restrained molecular dynamics (MD). The calculated structures have to be in agreement with experimental data, otherwise the potential energy of the system will increase. The experimental data report on an average conformation on the time scale of NMR experiments therefore the manner in which they are used can influence the outcome of calculations. Therefore it is important to understand how the experimental parameters are treated during the calculations.

Distance restraints: NOE

The distances are restrained during the calculations using a "soft-square" function. Within the "switching" region a square well function is set whereas outside of the switching region a "soft" asymptote is applied. The average distance, R, between the two sets of atoms (i and j) is computed as follow.

$$R_{ij} = \left(\sum_{ij} R^{-6} \right)^{-\frac{1}{6}} \tag{2.32}$$

The calculated distance is then compared to the distance, d, obtained from experimental data and the difference, Δ, is calculated as:

$$\Delta = \begin{cases} R - (d + d_{plus} - d_{off}) & \text{if} \quad d + d_{plus} - d_{off} < R \\ 0 & \text{if} \quad d - d_{minus} < R < d + d_{plus} - d_{off} \\ d - d_{minus} - R & \text{if} \quad R < d - d_{minus} \end{cases} \tag{2.33}$$

where d_{plus} and d_{minus} are the upper and lower limits and d_{off} is the offset function (often set to 0). The "soft square" restraining function is a combination of two functions. Within the limit defined by d, d_{minus} and d_{plus}, the energy penalty is null.

Inside of the switching region the energy penalty is a power function while outside of this region the energy increases linearly ($softexp = 1$) as illustrated in Fig. 2.11. In practice, d is set to the maximum distance, d_{plus} is 0 (except for degenerate protons) and d_{minus} is equal to d. The energy resulting from the distance restraints derived from NOESY or ROESY spectra E_{noe} is calculated as follows:

$$E_{NOE} = min(ceil, SC) \begin{cases} a + \frac{b}{\Delta^{softexp}} + c\Delta & R > (d + d_{plus} - d_{off} + r_{sw}) \\ \Delta^{exp} & R < (d + d_{plus} - d_{off} + r_{sw}) \end{cases} \quad (2.34)$$

This function is suitable for the calculations of distance restraints as the exact distance remains unknown. The use of upper and lower limits efficiently absorbs errors arising from miscalibrated cross peak volumes, the effect of spin diffusion or the distance variations due to the local dynamics of the molecule.

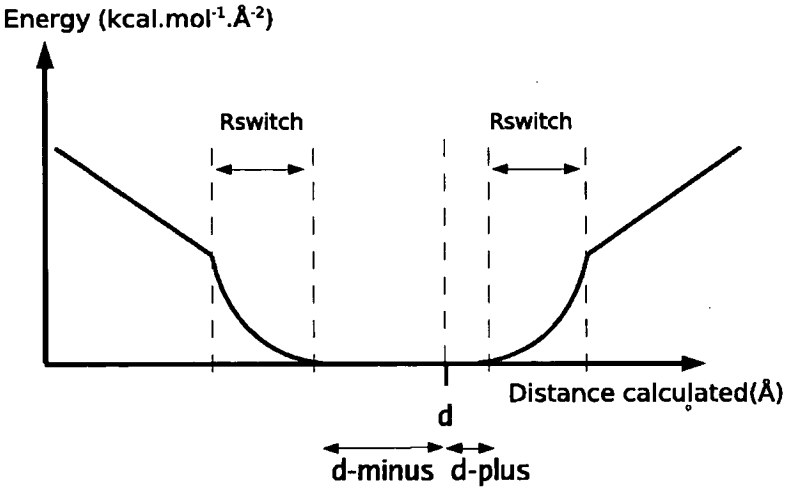


Figure 2.11: **Restraining function used for the calculation of NOE energy:** See text for details.

The treatment of the distance restraints outlined above takes into account several factors that influence the accuracy of the experimental distances.

Dihedral restraints: CDIH

The dihedral angles are restrained by a square-well penalty function defined by the experimentally determined centroid (θ), an angle range ($\Delta\theta$) accounting for errors of the measurement, inaccuracies in the parameterization of Karplus curves or limited flexibility, a weighting factor (W), the exponent (a) of the square-well function and

the energy constant (C in kcal mol^{-1} rad^{-2}). The dihedral angle energy restraint (E_{CDIH}) is calculated using θ as:

$$E_{CDIH} = W \sum C\ well(modulo_{2\pi}(\theta - \theta_o)\Delta\theta)^a \tag{2.35}$$

where the $well(\theta, \theta_o)$ function is calculated as:

$$\begin{cases} \theta - (\theta_o + \Delta\theta) & \text{if} \quad \theta > (\theta_o + \Delta\theta) \\ 0 & \text{if} \quad -(\theta_o + \Delta\theta) < \theta < (\theta_o + \Delta\theta) \\ \theta - \theta_o - \Delta\theta a + b & \text{if} \quad \theta < -(\theta_o + \Delta\theta) \end{cases} \tag{2.36}$$

The energy profile of the dihedral restraint function is shown in Fig. 2.12.

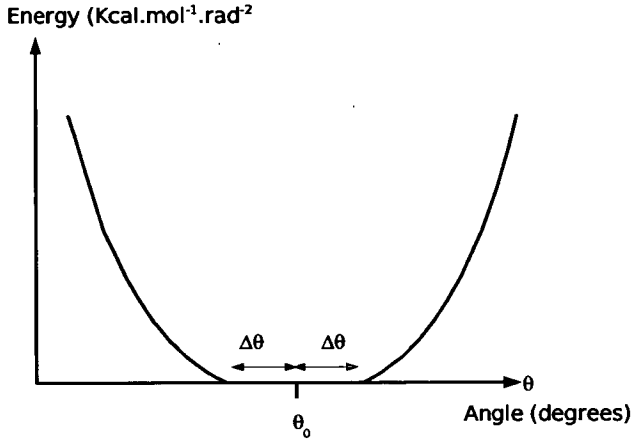


Figure 2.12: **Restraining function for dihedral angles:** Within a range of $\Delta\theta$ around the experimentally-determined dihedral angle θ, the dihedral angle energy is zero. Outside of the range, the energy follows a parabolic function.

Full relaxation matrix: FRM

The principle underlying the FRM formalism is to back-calculate the spectra using three-dimensional structures and to minimize the difference between the observed NOE intensities and those calculated by the full relaxation matrix. The target function $E_{relaxation}$ is defined in XPLOR (X-ray Crystallography and NMR) as:

$$E_{relaxation} = W_N \sum_{Spectra} \sum_{i=1}^{N_S w_i} \cdot \mathcal{F}(I_i^c, k_S I_i^o, \Delta_i, n)^m \tag{2.37}$$

where W_N is the overall energy constant for the relaxation term, I_i^c and I_i^o are respectively the calculated and observed intensities, Δ_i is an error estimate for I_i^o, which was

set to 0 with the use of a parabolic function ($\mathcal{F}$). It is suggested to use m=2 (parabolic function) and n=1/6 for NMR [78]. w_i is a weighting factor for individual spectra. This allows the contribution individual spectra. In calculations, the weights were uniform for all spectra to be weighted. k_S is the calibration factor which is determined for each spectrum by:

$$k_S = \frac{\sum_{i=1}^{N_S} I_i^c}{\sum_{i=1}^{N_S} I_i^o} \tag{2.38}$$

where N_S is the number of cross peaks in each spectrum. The function $\mathcal{F}(a, b, \Delta, n)$ is defined, for a parabolic function, as the difference between the n^{th} powers of I_i^o and I_i^c as $|I_i^{o,n} - I_i^{i,n}|$. By analogy with the R value used in crystallography, a R value is calculated to measure the fit of the refined structures to the NOE intensities.

$$R^n = \frac{\sum_{Spectra} \sum_{i=1}^{N_S} w_i \cdot \mathcal{F}(I_i^c, k_S I_i^o, \Delta_i, n)}{\sum_{Spectra} \sum_{i=1}^{N_S} w_i \cdot (k_S I_i^o)^n} \tag{2.39}$$

Part II

MATERIALS AND METHODS

Chapter 3

Materials and Methods

This chapter introduces the molecules studied during the course of tis work and the procedures that were used to collect, process and analyze the data obtained by NMR spectroscopy.

3.1 The cyclic peptides

The compounds studied were designed, synthesised supplied by G.E. Healthcare. They are four bicyclic pseudo-peptides labeled throughout the work as peptide 1, 2, 3 and 4. The samples, chemically pure, were stored under dry conditions at -20 °C as powders.

Structure and numbering of the peptides

The cyclic peptides have a common octapeptide core with the following sequence of amino acids: KCRGDCFC (Fig. 3.1). There are two rings in each molecule. The first ring is formed by a disulphide bridge between the central cysteines 2 and 6. The second cyclisation is achieved through a head-to-tail linkage between the N and C terminal lysine and cysteine. This linkage, NHCO-CH_2-S- can be thought of as being formed between the amine of the lysine and a modified side chain of the cysteine 8. Peptides 2-4 are obtained from peptide 1 by the derivatisation of the C-terminal cysteine and/or side chain of the lysine. Cysteine carries a polyethylene glycol chain of increasing length in compounds 2-4 and a metal chelating moiety is attached to lysine in compounds 3 and 4. It was decided to describe these hybrid molecules purely as peptides linked by

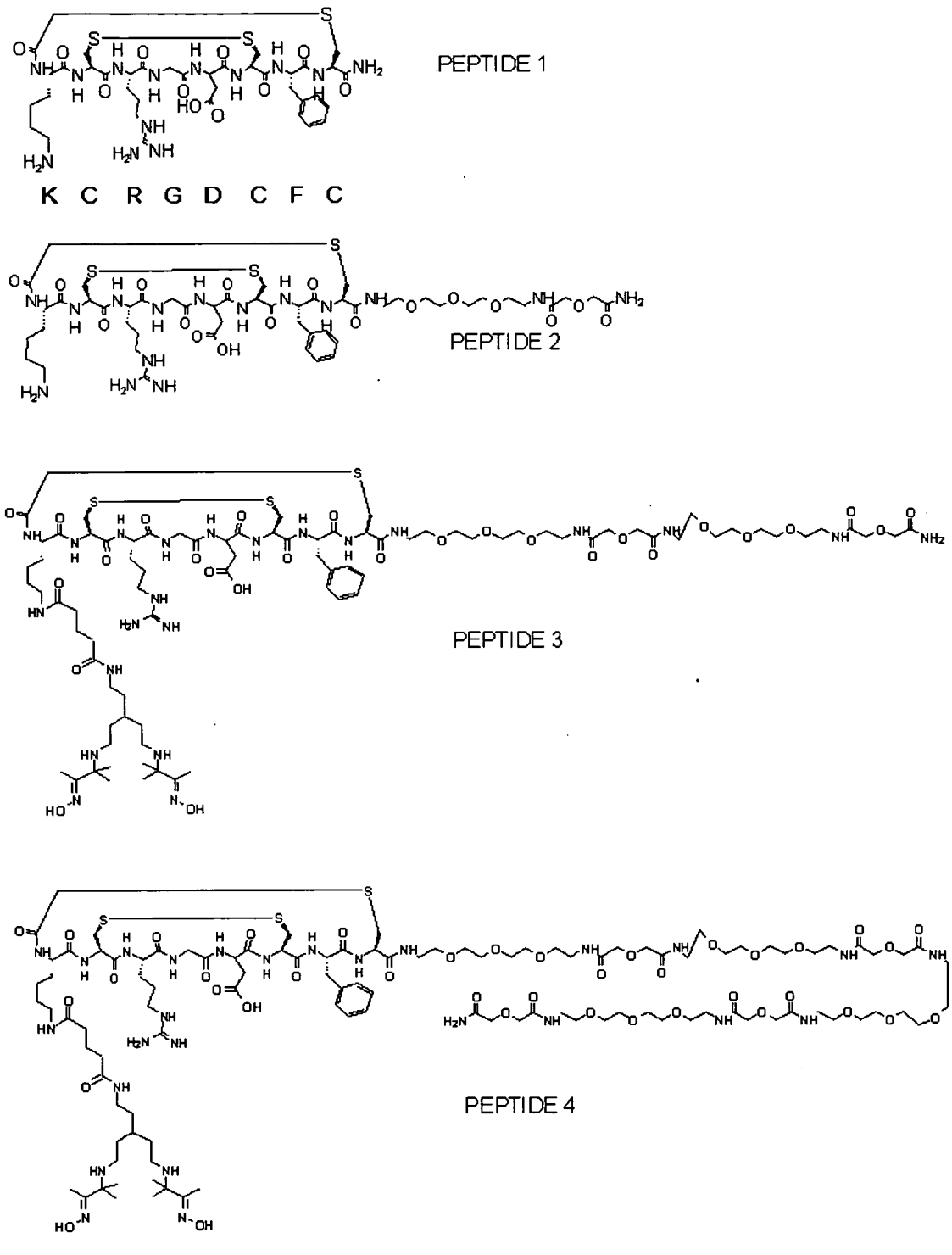


Figure 3.1: **Primary structures of bicyclic pseudo-peptides 1-4.**

amide bonds. It can be seen in Fig. 3.1 that the polyethylene glycol motif of peptide 2 is repeated two times in peptide 3 and four times in peptide 4, each of these motifs

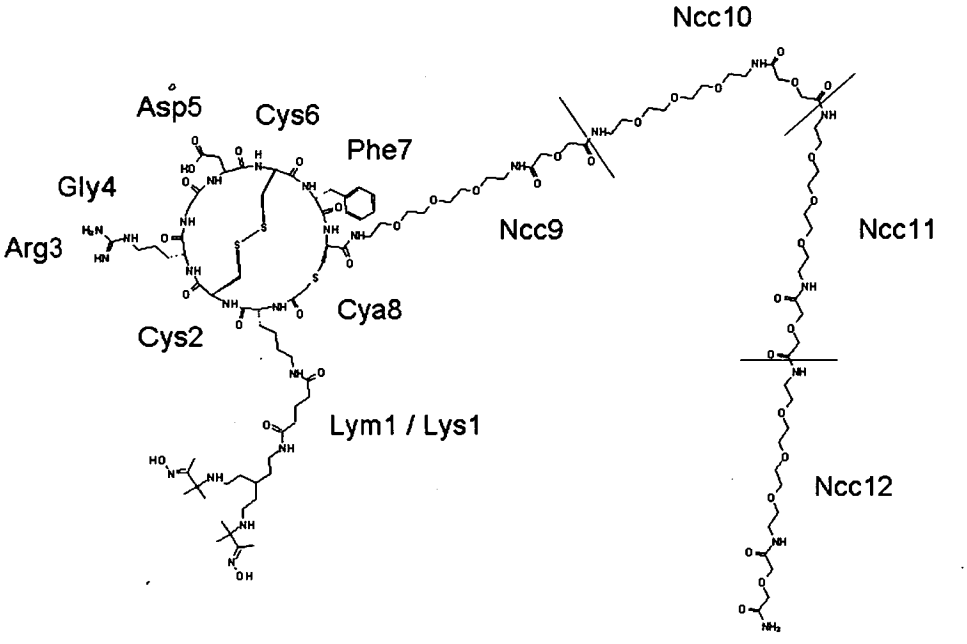


Figure 3.2: Residue numbering of peptide 4:

being separated by an amide bond. This motif was called "Ncc" and was formally treated as an amino-acid. The reasons behind this arbitrary descriptions are:

- Limit the number of new parameters that will be required for the modified force field.
- Ease the assignment of resonances which is usually performed by using the fingerprint region (one amide proton usually corresponds to one residue)

The nomenclature and the structure of these unusual amino-acids were described using three and one letter codes as described below.

Residue 8, Cya or B: The final residue of the peptide core was classified as a modified cysteine that carries an ethanoic group attached to the sulfur of the side chain (Fig. 3.3). The connection with the N-terminal of Lys1 is achieved via an amide bond between the side-chain of residue 8 and the amine of residue 1. This link was named "CLL".

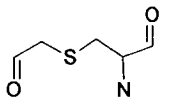


Figure 3.3: Cya residue.

Polyethylene glycol tail, Ncc or X: The polyethylene glycol motifs (Fig. 3.4) are linked to Cya8 or to themselves through amide bonds.

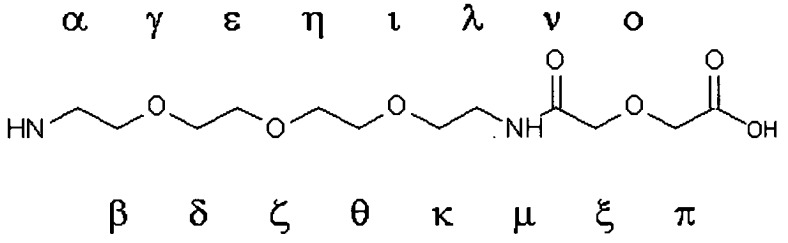


Figure 3.4: Ncc residue.

Modified lysine, Lym or J: These modifications occur only in peptides 3 and 4; this is a regular lysine in peptides 1 and 2. Similarly to Cya these modifications concern the side chain of an amino-acid. This residue was described as a sequence of three moieties lysine-link-MBS (MBS: metal binding site). The "link" can be described as a pentan-(1,5)-dioic and the MBS is a 3,3,11,11-tetramethyl -4,9- diazatridecane -2,12- dione dioxime bearing an ethylamine in position 7.

The amide group of the C terminal: The amide group attached at the C-terminus of the peptides was considered as part of the last residue (Cya in peptide 1 and Ncc in peptides 2-4).

This nomenclature outlined above will be used throughout this thesis. The one-letter and three-letter codes are summarized in appendix A.

3.2 Preparation of NMR samples

The study was carried out in water and per-deuterated dimethyl sulfoxide (DMSO-d6). All peptides were dissolved in water (H_2O:D_2O 9:1). The samples in water were

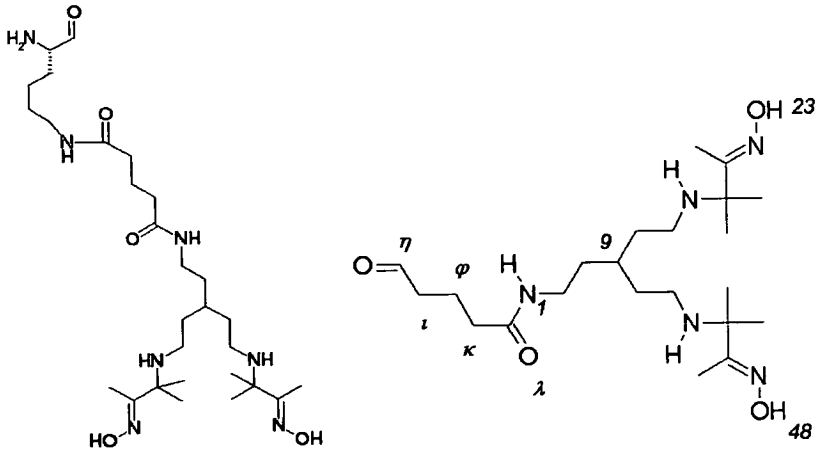


Figure 3.5: **Lym residue:** lysine, link and MBS moieties are represented in red, green and black. The numbering of the unusual part of the modified lysine is shown on the left. Detailled numbering is shown in chapter 5.

stored at 4°C between the experiments while the sample in DMSO was stored at room temperature. Peptide 2 was also dissolved in pure D_2O and pure DMSO-d6. The structures of peptides 1, 2, 3 and 4 obtained in water are referred to throughout the thesis as P1w, P2w, P3w and P4w while symbol P2d is used for the structure of peptide 2 in DMSO.

Peptides were dissolved in a solution containing 10 mM deuterated acetic acid/acetate. Sodium azide (NaN_3, 0.01% v/v) was added to prevent bacterial contamination and biological degradation of the samples. The samples were stable in solution and no substantial degradation was observed (Fig. 3.6). The deuterated acetic acid and the deuterated water were purchased from Sigma-Aldrich and were analytically pure (> 99.9 %).

Based on a series of one dimensional proton spectra acquired by G.E. Healthcare at various pH, the pH of the samples were adjusted to between 3.3 and 3.6. This range has been chosen because at higher pH the amide protons were exchanging too fast with the water protons and their signals were severely attenuated when water presaturation experiments were used. The final pH and concentrations of the samples were for peptide 1: pH 3.58, 11.21 mM; peptide 2: pH 3.3, 3.71 mM (another 44.7 mM sample was prepared for acquisition of NOESY spectra); peptide 3: pH 3.65, 15 mM; P4w pH 3.60, 4.45 mM and pH 3.3, 8.9 mM). The variations of concentrations used are

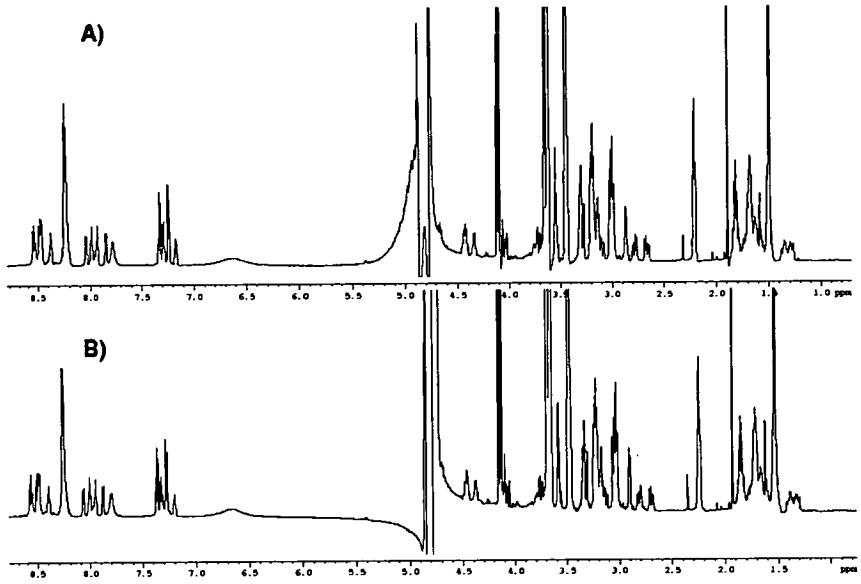


Figure 3.6: **Stability of peptide 4 in water:** One dimensional spectra acquired of peptide 4 acquire immediately after sample preparation at 298K and 800MHz (A) and three months later (B) using an identical receiver gain. No degradation of the sample or signal loss were observed.

explained by the varying amount of peptides available and the need to keep a constant, small sample volume for NMR experiments. Shigemi tubes and sample volumes of 320 μL were used in order to maximize the concentration.

A series of one-dimensional spectra was acquired at various temperatures. A temperature of 298K was chosen for structural investigation of peptides according to the largest dispersion of the H^N signals in proton spectrum of peptide 2. This relatively low temperature was also chosen in order to minimize possible conformational exchange. Lowering the temperature further lead to broadening of H^N resonances due to increased efficiency of quadrupolar relaxation of ^{14}N.

The sample of peptide 2 prepared in 90% v/v H_2O was lyophilised and re-dissolved in 350 μL D_2O keeping the concentration of buffer ions and thus pD values identical to the original sample.

Small peptides are often studied by NMR in DMSO. Dimethyl sulfoxide is advantageous for NMR experiments on small peptides because:

- The solubility of peptides at acidic pH in aqueous solutions is uncertain and amide protons cannot be observed at higher pH (fast exchange with H_2O) [79]. DMSO overcomes this problem ; it is a polar but aprotic solvent and the amide protons can be observed since the chemical exchange of labile peptide protons with protons of the residual water is slow.

- The viscosity of the DMSO (twice that of water) slows down the molecular tumbling of peptides in solution facilitating the observation of NOEs.

- The signal of residual water appears between the H^{β} and the H^{α} regions, in the part of the proton spectrum that contains few resonances, thus reducing the overlap problems. The residual DMSO signal appears at the edge of the H^{β} region (high field) and overlaps with few signals.

- The NMR structure of a related peptide has been solved in DMSO [38]. It was therefore of interest to compare it with the structures of peptides 1-4. Peptide 2 was taken forward for complete structure determination in DMSO.

Peptide 2 was dissolved in 500 μL DMSO-d6 (>99.5% pure, purchased from Sigma-Aldrich). This peptide was very soluble in DMSO and the final concentration of the sample of 13.6 mM was achieved. Similarly to the experiments carried out in water, a series of spectra were acquired at various temperatures in order to optimize the dispersion of the H^{N} and H^{α} protons. A temperature of 303 K was chosen for the structure determination.

3.3 NMR experiments

An extensive set of NMR experiments were used to obtain structural restraints, local geometry and to assess flexibility of peptides. NMR experiments were carried out on Bruker Avance 600 MHz and 800 MHz spectrometers equipped with a triple-resonance cryoprobe (600 MHz) and a 800 MHz triple resonance probe with triaxial pulse field gradients. Two NOESY spectra of peptides 1 and 3 were acquired on a Varian INOVA 900 MHz spectrometer (The Henry Wellcome Building for Biomolecular NMR, Birmingham).

3.4 Processing NMR spectra

NMR data were processed using two programs: AZARA [80] and XWINNMR (Bruker). Unless specified otherwise, all spectra were processed in XWINNMR. AZARA was used to prepare files for ANSIG [81] (volume measurement of NOE cross peaks). The AZARA package includes programs PROCESS, PLOT2, CONTOUR and CONNECT. PROCESS has all the facilities for processing FIDs (zero-filling, phase correction, baseline correction, weighting functions, etc.) and is also capable of converting the XWINNMR processed spectra to a format compatible with ANSIG. PLOT2 was used to visualize the spectra created by PROCESS. CONTOUR created the contour file required by the assignment program ANSIG. CONNECT converted the assignments and volumes from ANSIG to a readable format for XPLOR or CNS.

NOESY spectra acquired on Bruker spectrometers were processed in XWINNMR and "blocked" in PROCESS. Data were processed using a Gaussian window function. Spectra used for coupling constant and temperature coefficients determination were processed in XWINNMR and zero filled to 32K points in the directly-detected dimension to achieve a digital resolution of approximately 0.2 Hz per point. Data acquired on the VARIAN spectrometer were Fourier transformed with PROCESS. Resolution enhancement was achieved by multiplying the data with a Gaussian function in the directly-detected dimension and with a sine squared function in the F_1 dimension. The digital resolution was doubled by zero filling.

3.5 Resonance assignments

The assignment of proton resonances within the same residue was achieved through the analysis of homonuclear 2D COSY spectra. A summary of experimental parameters for J-based homonuclear experiments is reported in Table 3.1.

3.6 Distance restraints

The assignment and integration of the NOE cross peaks were performed in ANSIG. A contour file was created using processed spectra by the program CONTOUR [80] and

Table 3.1: **Experimental parameters of 2D homonuclear experiments used for the assignment of protons resonances:** Differences between individual peptides arise from varying availability of spectrometers time and sample concentrations. The spectra were acquired at 800 MHz. Water suppression was achieved using DPFGSE for samples in H_2O.

Peptide	Experiment	Mixing time (ms)	NS[a]	t_{1max} (ms)
1 (H_2O)	DQF-COSY		32	85..2
	TOCSY	80	2	170.5
2 (H_2O)	DQF-COSY		2	204.8
	TOCSY	65	4	102.4
3 (H_2O)	DQF-COSY		12	160
	TOCSY	60	8	146.3
4 (H_2O)	DQF-COSY		12	75
	TOCSY	60	32	75.3
2 (DMSO)	DQF-COSY		2	142.7
	TOCSY	60	32	35.5

[a] Number of scans per increment.

imported into ANSIG. The cross peak volumes were calculated using the integration procedure in ANSIG. The accuracy of the volumes was improved by employing variable box sizes that were adjusted manually in order to avoid over estimation by inclusion of neighboring cross peaks. When possible, the volume of partially overlapped cross peaks was determined by estimating the contribution of each cross peak. In such case the total volume of both cross peaks was measured and the contribution of each cross peak was determined by using peak areas of individual peaks obtained by a resolution-enhancing window function. Alternatively, line deconvolution was performed using 1D cross sections in XWINNMR. When working with a series of NOESY spectra acquired with different mixing times, the sizes of integral boxes were fixed and no volume prediction was done.

For the smallest peptides, ROESY experiments were acquired to circumvent the problem of weak NOE enhancements. ROESY experiments were acquired for peptides 1, 2 and 3 in water. The distance restraints were extracted as for the NOESY spectra.

The T-ROESY experiment was acquired for peptide 2 in water at 800 MHz. According to Maliavin et al. [67], the offsets for off-resonance irradiation were placed at $|1.5\delta_m|$ of the carrier frequency. The maximum frequency between the carrier frequency and the

Table 3.2: **Acquisition parameters for 2D NOESY spectra used for sequential assignment and initial structure calculations:** The water suppression in H_2O samples was achieved by DPFGSE [59]. The residual water in D_2O was suppressed by a soft presaturation of the water signal. Unless specified otherwise, the spectra were acquired at 800 MHz.

Peptide	NS	Mixing time (ms)	t_{1max} (ms)
1 (H_2O)[a]	48	250	109
2 (H_2O)	64	200	98
2 (D_2O)	16	200	150
3 (H_2O)[a]	48	250	109
4 (H_2O)	32	250	160
2 (DMSO)	16	100	160

[a] Spectra acquired at 900 MHz

Table 3.3: **Acquisition parameters for 2D ROESY experiments:** Unless specified otherwise, the spectra were acquired at 800 MHz.

Peptide	NS	Mixing time (ms)	t_{1max} (ms)
1 (H_2O)	32	200	151
2 (H_2O)	8	400	146
2 (H_2O)[a]	16	200	102
3 (H_2O)	32	200	128

[a] *acquired at 600 MHz*

spectral frequencies was set to 3000 Hz. The carrier frequency was measured at 3814 Hz which gives the off-resonance frequencies at 8314 Hz and -686 Hz. The irradiation strength was set to 2.45 KHz. These settings yielded a tilt angle of 50^o. A 200 ms mixing time T-ROESY was acquired with 32 scans for t_{1max} of 68.9 ms.

Distance restraints obtained from a series of NOESY spectra: XWINNMR automatically scales each spectrum based on the intensity of the largest signals. The spectra were therefore re-scaled manually in order to ensure a direct comparison of intensities at various mixing times. Integration of cross peak volumes leads to build-up curves, showing the kinetics of NOE enhancements. The volumes were calculated using constant area (0.01×0.02 ppm) boxes in order to avoid inconsistencies between individual spectra. Only unambiguously assigned cross peaks were integrated. Absolute intensities were used to generate the XPLOR restraint file, which allowed direct

comparison between individual spectra.

Table 3.4: **Series of 2D NOESY spectra with increasing mixing time**: Unless specified otherwise, the spectra were acquired at 600 MHz.

Peptide	NS	Mixing times (ms)	t_{1max} (ms)
2 (H_2O)	16	55, 75, 100, 200, 400	149
2 (D_2O)[a]	16	55, 75, 100, 200, 400	150
3 (H_2O)	16	55, 75, 100, 150, 200, 400	149
4 (H_2O)	16	55, 75, 100, 150, 200, 400	160
2 (DMSO)	16	55, 75, 100, 200, 400	160

[a] *acquired at 800 MHz*

3.7 Measurement of coupling constants

Table 3.5: NMR Experiments for measurement of scalar coupling constants.

Experiments	**Vicinal coupling constants**
1D proton spectrum	$^3J_{H^\alpha H_2^\beta}$, $^3J_{H^\alpha H_3^\beta}$
1D TOCSY	$^3J_{H^N H^\alpha}$, $^3J_{H^\alpha H_2^\beta}$, $^3J_{H^\alpha H_3^\beta}$
1D COSY	$^3J_{H^N H^\alpha}$, $^3J_{H^\alpha H_2^\beta}$, $^3J_{H^\alpha H_3^\beta}$
2D ^{13}C HETLOC	$^3J_{H^N C'}$, $^3J_{H^\alpha C^\gamma}$
2D ^{15}N HETLOC	$^3J_{NH_2^\beta}$, $^3J_{NH_3^\beta}$
2D ^{15}N HSQC	$^3J_{H^N H^\alpha}$
2D ^{13}C HMBC decoupled and refocused	$^3J_{H^N C'}$, $^3J_{H^N C^\beta}$, $^3J_{H^\alpha C^\gamma}$, $^3J_{H_3^\beta C'}$, $^3J_{H_2^\beta C'}$, $^3J_{H^\alpha C'_{i-1}}$
2D ^{13}C ge HMBC	$^3J_{H^N C'}$, $^3J_{H^N C^\beta}$, $^3J_{H^\alpha C^\gamma}$, $^3J_{H_3^\beta C'}$, $^3J_{H_2^\beta C'}$, $^3J_{H^\alpha C'_{i-1}}$
2D ^{15}N ge HMBC	$^3J_{NH_2^\beta}$, $^3J_{NH_3^\beta}$, $^3J_{H^\alpha N_{i+1}}$

Scalar coupling constant were extracted from the one-dimensional and two-dimensional spectra listed in Table 3.5. Different experiments provide coupling constants with different efficiency. The factors influencing this vary from experiment to experiment. A major factor usually, but not always, is the size of the coupling constant. Some experiments are limited by their poor sensitivity at natural abundance of ^{13}C and ^{15}N. Imperfect suppression of H_2O protons renders some (usually H^α) protons unobservable. The coupling pattern of H^β of lysine and arginine residues led to complicated cross

peaks from which the coupling constants were difficult to extract. In addition, proton and carbon chemical shifts of CH^{β} moieties H^{β} were partially overlapped. A series of 1D TOCSY and 1D COSY experiments with selective excitation of H^{α} protons solved this problem since the chemical shift of these protons for the two residues were resolved. At the same time, these two spectra offered a convenient way for coupling constant determination from such poorly-resolved multiplets. Both spectra were acquired using the same experimental parameters (acquisition time, number of scans, receiver gain, temperature, offset, selective pulse). The peaks in the COSY spectra were antiphase with respect to $^{3}J_{H^{\alpha}H^{\beta}}$ couplings. The 1D TOCSY yielded pure phase multiplets in which all the coupling constants were in-phase.

In principle, the methods employed for the determination of long-range coupling constants employed in this work can be divided into two groups:

1. **Direct methods.** The coupling constants are directly measured from the splitting of multiplets. This method was used only when the peaks were in pure absorption mode. In particular it was used to extract $^{3}J_{H^{N}H^{\alpha}}$ coupling constants from 1D ^{1}H spectra (basic, 1D TOCSY and ^{1}H-^{15}N HSQC spectra). Simple reading of the resonance frequencies was also employed when the coupling constant was coded as the frequency shift between two otherwise identical multiplets

2. **Extraction of coupling constants by reconstructing multiplets.** This method was applied to the analysis of ge-HMBC spectra, where the signals are antiphase with respect to the heteronuclear coupling but have a mixed phase with regard to proton-proton couplings.

A home-written program [1] performing a two-dimensional search (scaling and shifting) was used for the extraction of coupling constants from mixed phase multiplets. The details of this method are discussed in chapter 4.

[1] The program was written by Dr. Tran Pham. It performs a Powell minimization with two variables, the scaling factor and the frequency shifting. The function that is minimized is the root square of the difference between the reconstructed (added, shifted and scaled) signal and the experimental multiplets.

^{15}N coupled ^{1}H^{15}N-HSQC experiments

At natural abundance, molecules contain 99.63% of ^{14}N. This nucleus, has a spin quantum number $I = 1$ and associated with a non-spherical charge distribution. This induces quadrupolar relaxation of nearby nuclei and thus broadens the lines of H^N protons attached to the ^{14}N. $^{3}J_{H^NH^\alpha}$ coupling constants are therefore more accurately measured from ^{1}H-^{15}N HSQC experiments where ^{1}H attached to ^{15}N are detected [82]. In order to increase the digital resolution of the directly-detected dimension, the spectra were acquired without ^{15}N decoupling. This eliminated heating and allowed the use of acquisition times of 0.5 to 0.6 s. The H^N multiplets were therefore, in addition to $^{3}J_{H^NH^\alpha}$ coupling constants, also split by one-bond ^{1}H-^{15}N coupling constants. The latter splitting appears in antiphase. An additional advantage of this approach is that the overlap often seen in 1D spectra is eliminated due to dispersion of ^{15}N chemical shifts.

Parameters used for ^{15}N coupled ^{1}H^{15}N-HSQC are presented in table 3.6. Coupling constants were measured from the 1D cross sections in XWINNMR. Spectra were acquired at 600 MHz.

Table 3.6: Experimental parameters for ^{15}N coupled ^{1}H-^{15}N HSQC.

Peptide	NS	t_{1max} (ms)	AQ[a] (ms)
1 (H_2O)	128	45	526
2 (H_2O)	512	53	621
3 (H_2O)	128	70	620
4 (H_2O)	128	70	619
2 (DMSO)	64	140	310

[a] Aquisition time in the directly detected dimension.

HMBC experiments

HMBC experiments for multiplet reconstruction from one-bond templates: The experimental time for each HMBC experiment was 16 hours. Each experiment was acquired with 512 increments and 64 scans per increment. The preparation delay Δ was set to 80 ms and the overall evolution delay for proton chemical shift and

Table 3.7: **Experimental parameters for HMBC experiments:** The ^{13}C HMBC were acquired using spectral widths of 6250 and 27156 Hz in F2 and F1, respectively. For ^{15}N HMBC these parameters were set to 6250 and 2128 Hz. Four delays Δf for HMBC experiments using OBMB filter were set to 3μs, 1.786 ms, 3.571 ms and 5.357 ms. All spectra were acquired at 600 MHz with a 328 ms acquisition time in the directly-detected dimension.

Peptide	X	Evolution delay (ms)	NS	$t_1 max$ (ms)
HMBC experiments using 1D TOCSY template				
1 (H_2O)	^{13}C	82.4	72	29.5
3 (H_2O)	^{15}N	104.64	1856	16.9
	^{13}C	86.631	128	18
HMBC experiments using one-bond template				
3 (H_2O)	^{13}C	80	64	9.4
4 (H_2O)	^{13}C	80	64	9.4

proton-proton coupling constants was identical to the experiment described in 4.3.2 i.e. 86.631ms. Cross peaks, either long-range or one-bond, should present a similar shape compared to the reconstructed multiplets and the HMBC experiments described above.

All spectra were processed using identical parameters. No phase correction in the F_2 dimension was applied and spectra were zero-filled to create a 2D matrix of 32K×1K points. The addition and subtraction of 2D spectra was performed in XWINNMR using the "add2d" command. The spectra were processed to alternatively remove the one-bond or the long-range cross peaks as shown in Fig. 3.7. The different combination of HMBC spectra yielded different intensities for one bond cross peaks. Figure 3.7 shows the reconstructed HMBC spectra containing separated long-range and one-bond cross peaks. For well-resolved and intense cross peaks, both methods returned the same values of coupling constants. The method using one-bond cross peaks as templates produced better results than when 1D TOCSY multiplets were used. However, few long-range coupling constants could be extracted from HMBC experiments in general, which was particularly true for ^1H-^{15}N coupling constants.

There are two main problems associated with both methods: (i) the low signal-to-noise ratio which accompanies cross peaks mediated by small coupling constants and again, (ii) a particular problem for small coupling constants, the possibility of the

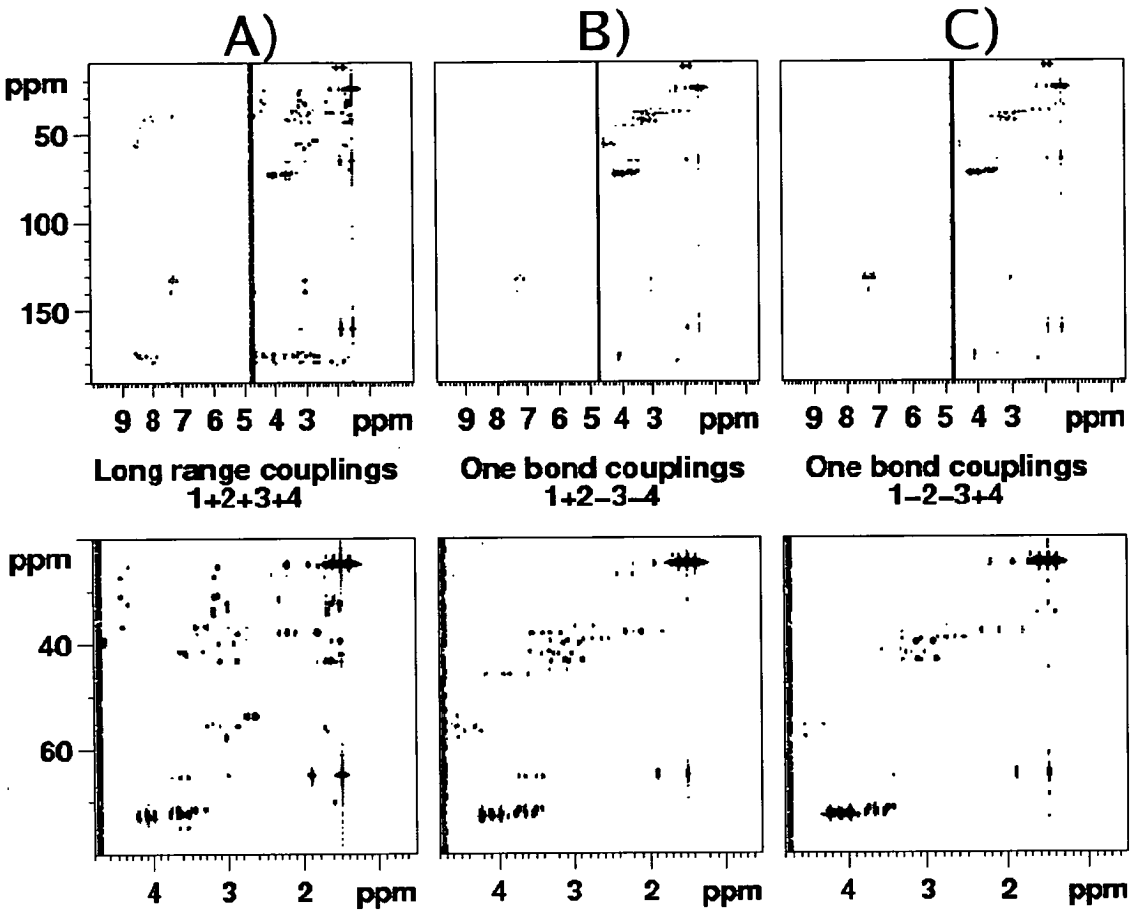


Figure 3.7: **OBMB HMBC spectra of peptide 3:** The spectra on the top show the separation of (A) long-range cross peaks (1+2+3+4) and (B and C) one bond cross peaks (1+2-3-4 and 1-2-3+4). The spectra at the bottom show expansions of the aliphatic region illustrating the efficiency of the filter.

analysis converging on incorrect scaling factor/coupling constant pair. It was found that the results could vary dramatically depending on the region taken to extract the coupling, and/or window function applied. For small peptides with natural abundance of isotopes the extraction of long-range coupling constants is challenging mainly due to the increased linewidths compared to small organic molecules (< 500 g/mol) such as glucose where these methods work better.

For these reasons, another implementation of HMBC was investigated and applied to the measurement of heteronuclear long-range coupling constants.

HMBC refocused and decoupled experiments The low signal-to-noise ratio observed for cross peaks mediated by small coupling constants is caused, to a large extent, by the cancellation of lines with opposite phase in HMBC spectra. This leads to an inability of program algorithms to determine correctly the scaling factor. A method that would present the heteronuclear coupling constants in phase and would also remove the necessity of determining the scaling factor would be more appropriate to the measurement of long-range heteronuclear coupling constants in peptides. This is possible to achieve within a framework of HMBC experiment, but in its refocused form. These experiments were acquired for all peptides using identical parameters summarized in Table 3.8.

Table 3.8: **Experimental parameters for refocused and decoupled HMBC experiments:** The spectral width was set to 6614×2715 Hz for all the experiments.

Peptide	NS	t_{1max} (ms)	AQ (ms)
1 (H_2O)	128	9.4	200
2 (H_2O)	192	7.1	174
2 (DMSO)	192	7.1	174
3 (H_2O)	256	5.9	200
4 (H_2O)	192	9.4	200

HETLOC experiments

HETLOC experiments were used to determine the intra-residue heteronuclear couplings constants. The long-range proton-carbon coupling constant corresponds to the frequency difference between the corresponding TOCSY peaks in the low and high field parts of a multiplet separated by $^1J_{XH}$ in F_1. In principle, their values can be determined by a simple reading of peak frequencies, and the HETLOC method is therefore included in direct methods. In practice, the coupling constants were determined using a modified version of "decantiphase". This program minimizes the difference between two peaks that were substracted and shifted (Fig. 4.6). The value of the optimal shift corresponds to the value of a coupling constant. Very small coupling constants could be measured from the ω_1-X half filtered TOCSY spectra.

The parameters used to acquire the X-filtered HETLOC experiments are given in Table

3.9. The scaling factor in F_1 was set to one and the $^1J_{CH}$ and $^1J_{NH}$ were set to 140 Hz and 100 Hz, respectively. All spectra were acquired at 600 MHz.

Table 3.9: Parameters for X-filtered experiments.

Peptide	X	t1*max* (ms)	NS	Mixing time (ms)	AQ (ms)
1 (H_2O)	^{13}C	78	64	60	492
	^{15}N	74	136	60	492
2 (H_2O)	^{13}C	78	72	60	492
	^{15}N	74	136	60	492
3 (H_2O)	^{13}C	78	128	80	492
	^{15}N	74	128	80	492
4 (H_2O)	^{13}C	78	64	60	492
	^{15}N	57	192	60	492
2 (DMSO)	^{13}C	43	80	60	171
	^{15}N	74	88	60	492

3.8 Referencing the proton spectra

The temperature sensors of spectrometers measure the temperature below the sample and not directly in the sample. Also the set value may not correspond accurately to the actual temperature. Therefore the temperature of the sample must be calibrated in order to obtain accurate temperatures. The correlation between the set and the actual temperature was obtained using a sample of glycol. The separation of the peaks of the ethane and hydroxyl protons is proportional to the temperature. The macro "tecalib" in XWINNMR was used to calculate the temperature of the sample according to the frequency difference between the two signals.

The chemical shifts were referenced using the solvent signal (water or DMSO). In water, the carrier frequency was set on the water signal and its chemical shift (δ_{H_2O}) [83] was calculated according to the equation as:

$$\delta_{H_2O} = 4.766 - 0.0119(t - 25) - 0.002(pH - 7) - 0.009[salt] \quad (3.1)$$

where t is the temperature in ^oC. XWINNMR macro "ref", which calculates the chemical shift of water, was used to reference the spectra.

The chemical shift of the DMSO signal also varies with the temperature. This variation

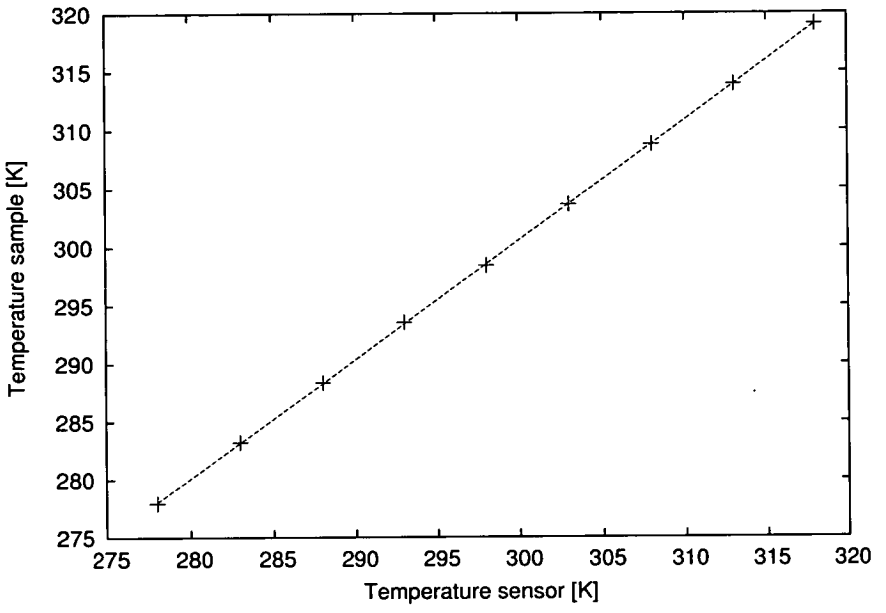


Figure 3.8: **Temperature calibrations:** Calculated temperature of the sample versus the temperature returned by the sensor of the spectrometer. The crosses represent the experimental points. The dotted line is the best fit between the temperature of the sample and the target temperature.

was assessed using a sample containing DMSO and silicone oil, an internal standard. The chemical shift of silicone oil signal is temperature-independent. The chemical shift of the DMSO signal was set to 2.510 at 295 K and sampled every 2 °C from 298 K to 340 K. The figure 3.9 shows the chemical shift changes of DMSO signal with the temperature relative to that of the silicon oil. The obtained correlation is shown in Fig. 3.10.

A linear regression was applied to fit the temperature versus the chemical shift of the DMSO signal. The resulting equation 3.2 was used to reference the proton spectra for the experiments carried out in DMSO.

$$\delta_{DMSO}(ppm) = -6.047 \times 10^{-4} \times T + 2.6798 \tag{3.2}$$

3.9 Structure calculation protocols

The following section describes the protocols used in this work. The three-dimensional structures were calculated using the programs CNS (Crystallography and NMR system)

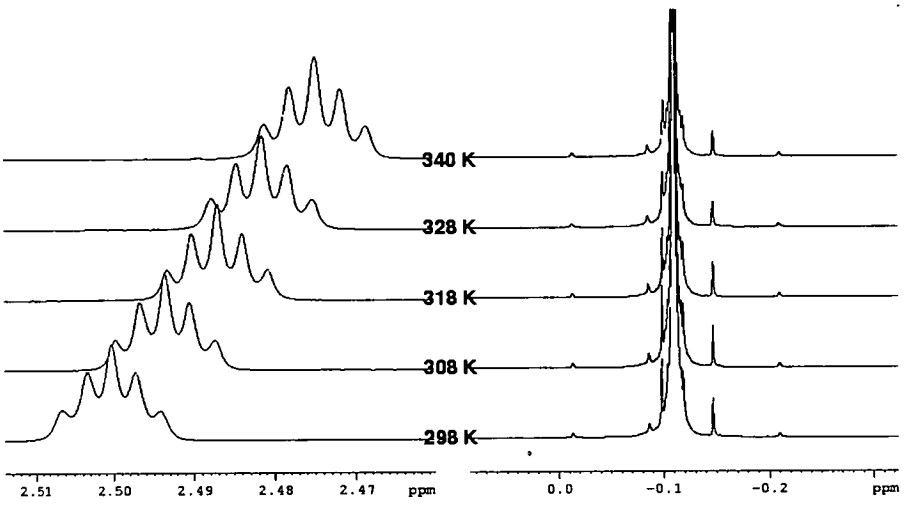


Figure 3.9: **Expanded regions of DMSO and silicon oil signals at different temperatures:** The series of spectra on the left shows the chemical shift variation of DMSO signal with the temperature whereas the series on the right shows that the chemical shift of the silicon oil signal does not vary with the temperature.

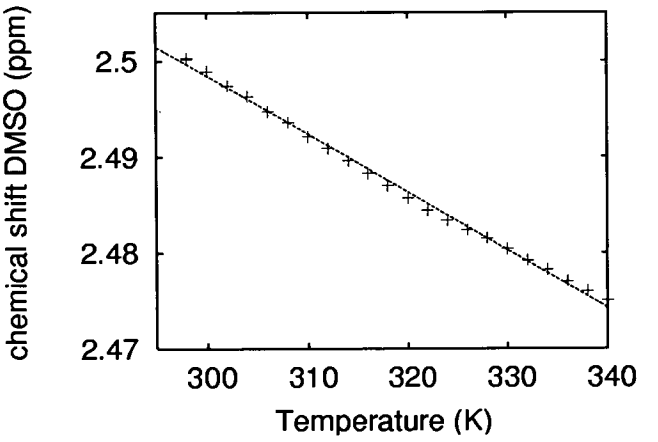


Figure 3.10: **Temperature dependence of the chemical shift of DMSO:** The crosses represent experimental points and the straight line is the best linear fit.

version 1.0 [84] and XPLOR-NIH version 2.1[78]. Various force fields can be used within CNS and XPLOR (PARALLDG, AMBER/OPLS, CHARMM, PROSLQ) [85]. The commonly employed PARALLHDG force field, whose geometric parameters are derived from the Cambridge Structural Database [86], was used.

For distance restraints, parameters used during the course of the calculation are summarized in Table 3.10. The parameters that were fixed during the calculations are

given in bold.

Table 3.10: **Parameter for distance restraints.** The table describes parameters of the CNS restraining function used to calculate NOE energy.

Parameter	Definition	Statement[a]
ceil	Ceiling value for energy constants	ceiling = 100
S	Scaling factor for NOE classes	scale
C	Additional scaling factor	sqconstant = 1
softexp	Exponent for the soft square function	soexponent = 1
a and b	Constants determined by the program (the function is smooth at edge of the switching region) at R = (d + d_{plus} - d_{off} + r_{sw})	No statement
c	Slope of asymptote	asymptote
R	Effective distance	defined in 2.32
exp	Exponent for the square well function within the switching region	sqexponent = 2
$\mathbf{d}_{off}$	Offset	Sqoffset = 0
d, $\mathbf{d}_{plus}$, $\mathbf{d}_{minus}$	Distance, upper and lower limits	ASSIGN
$\mathbf{r}_{sw}$	Switching region between asymptote and square	Rswitch = 1

[a] Unless specified otherwise, these values were used in CNS scripts

For dihedral restraints, unless specified otherwise, the energy constant (C) and the weighting factor (W) were set to 20 and the exponent (a) to 2. These settings led to rather stringent restraints that kept the conformation of peptides in the regions defined by the experimentally-determined dihedral angles. The dihedral restraints out-weighted the distance restraints because they were determined accurately by combining coupling constants related to the same dihedral angle. This also meant that the conformational search was performed within regions strictly defined by the dihedral restraints. In cases where distance and dihedral restraints would not converge toward the same conformation, distance restraints would be violated rather than the dihedral restraints or, in a situation where no violations of the experimental parameters were observed, the restrained dihedral angles would be found at the limit of the region defined by the dihedral restraints. The structure determination strategy is comprised of three sequential steps which are summarized in Fig. 3.11: creation of structure files that included the unusual residues, calculation of the initial structures and structure

refinements.

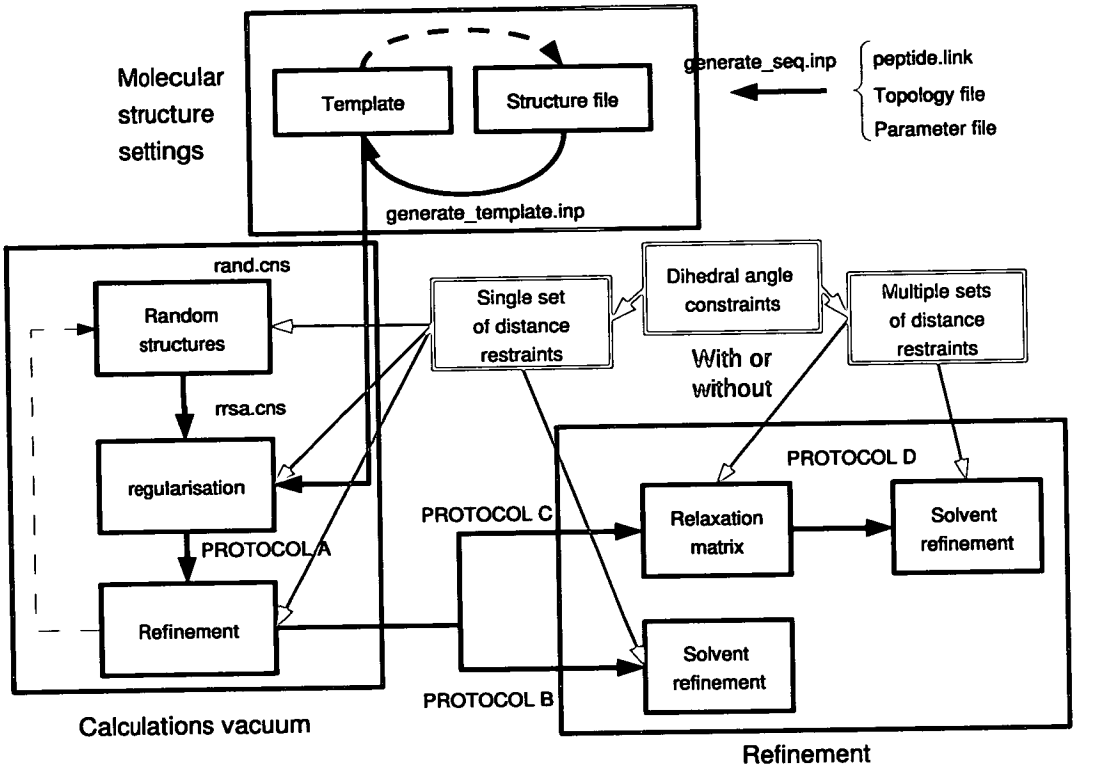


Figure 3.11: **Structure determination.** The figure summarizes the different steps used for NMR structure determination. The scripts are underlined, the black arrows show the flow of the calculations, the dashed lines indicate a loop aimed at tuning the parameters, the red boxes represent the experimental data extracted by NMR, the blue boxes indicate the calculated structures or parameters. The figure is adapted from the XPLOR manual [78].

Generating the template and the structure files

The structure calculation process requires a structure file that defines and quantifies empirical parameters and a template file that is used to regularize the structures. The structure file is created from the parameter, topology and link files. The template file contains the coordinates of the molecule calculated via free molecular dynamics and energy minimization in CNS and XPLOR. These calculations allow possible mistakes in the definitions of the chemical parameters to be revealed.

The topology file was first created for the residues and pre-residues in order to link the

different residues by defining the atom types, atom names, geometrical relations (angles, dihedral angles, improper angles, bonds), local charges and atomic masses. Two methods were used to create the parameters for the unusual parts of the molecules. The parameters for residues Cya and Ncc and the pre-residue CLL were derived from existing parameters of the force field or taken from the Hic-Up database [87]. The structure of residue Lym was constructed and minimized using the program HYPERCHEM. The equilibrium parameters were calculated using Cartesian coordinates of the energy-minimized structure and the definitions of the topology file using protocol *learn.inp* (appendix B). The new parameters were appended to the original parameter file. The topology (appendix C), parameter (appendix D) and link (appendix E) files were employed to create a new structure file reflecting the amino acid sequence with unusual residues. The uniqueness of the parameters was checked and a template structure was generated.

Calculating the first vacuum structures

The first set of structures were generated from random coordinates using restrained molecular dynamics in vacuum. Starting from random coordinates, three steps are necessary to calculate the structure [88]. The first stage *rand.cns* [89] minimizes the energy (Evdw) of the molecule at high temperature (2000 K and 1500 K) using atom volumes. The calculations start with a simplified representation of the molecules using 1 or 2 atoms per residue, typically C^{α} and C^{β}. The radii of the chosen atoms are large while the other atoms have no volume. The covalent energies are set to low values and no impropers are included in the energy term. When the other atoms are added, the covalent and the experimental energies are progressively increased whilst the radii are brought back to their normal values. The second stage regularizes the structures by slowly introducing planarity and chirality. A test checking the correctness of the enantiomers is performed by comparing the energy of two enantiomers for each chiral center. The enantiomer showing the lowest energy is selected. This procedure is adapted to prochiral groups where the protons have been arbitrarily stereospecifically assigned [90]. The regularization is achieved by molecular dynamics and simulated annealing (MDSA) (2000 K to 100 K) in *rrsa.cns*.

The final step refines the regularized structures using a linear slow-cooling scheme (see protocol A page 64 for details). The overall length of the molecular dynamics is 2.43 ps followed by a test for correct enantiomers and Powell minimizations of the structures obtained from molecular dynamics. These scripts used were developed by Nilges and are available on the XPLOR website [78]. The weighting of the experimental parameters were kept constant (NOE 25 kcal/mol/Å^2, CDIH 20 kcal/mol/rad^2) throughout the simulation.

Structure refinements

The structure refinements are aimed at giving a more accurate and more realistic representation of the peptides in solution. This step was performed in two different ways. The vacuum structures were refined in solvent and both the vacuum and water structures were also calculated using the full relaxation matrix. The molecules were refined in an explicit solvent where the dielectric interactions are accounted for [2].

A full relaxation matrix protocol calculates the relaxation matrix from NOESY cross peaks obtained at various mixing times and therefore leads to more accurate calibration of intermolecular distances. Therefore this protocol was used to refine the structures calculated using single mixing time NOESY spectra.

Vacuum: protocol A

The first structures were calculated in vacuum using a single set of distance restraints. The NOE restraints were divided into four distance classes according to their normalized intensities (very weak, weak, medium and strong). This process is discussed in detail in section 5.2. The experimental restraints were introduced in the early stages of the structure calculation starting from random coordinates as detailed in Table 3.11. During the first cooling stage of the refinement, the NOE asymptote was linearly increased in order to improve the convergence of the structures. The prochiral protons were not assigned stereospecifically; prochiral assignment was achieved by swapping their assignment and comparing the energy of the resulting enantiomers. This proce-

[2] In vacuum the atoms are treated as purely repulsive spheres whereas in solution both the repulsive and attractive interactions are used.

dure worked well for β protons, but was not efficient for α protons of the glycine. A filter was therefore set up for this residue as described in section 5.2.

Table 3.11: **Stages of simulated annealing in vacuum:** The high temperature conformational search is followed by two simulated annealing runs with a linear cooling scheme. The weight of the restraining functions was kept constant for NOE and dihedral restraints at 25 and 20 respectively.

Cooling stage	High temperature	Cooling stage 1	Cooling stage 2
temperature(K)	2000	2000 to 1000	1000 to 100
length (ps)	150	1800	480
NOE asymptote	0.1	0.1 to 1.0	1.0

Water refinement: protocol B

The structures obtained using protocol A were "soaked" in water or DMSO. The TIP3 model was used for water. The parameters for DMSO were taken from the examples provided by the program XPLOR. Non-bonded parameters for water were similar to OPLS (Optimized Potential for Liquid Simulations) parameters and the PARALLHDG5.3. force field was employed to account for solvatation effects and dielectric interactions. The dihedral angle term, DIHE [3] was removed from the empirical parameters as it was not included in vacuum and undefined in the topology file [91]. The distance restraints parameters used in structure calculations in explicit solvent differ from those employed in vacuum (Table 3.12). The slope of the asymptote of the restraining function is 2, rather than 1 and the scaling factor is 50 instead of 25. These parameters are chosen to make NOEs more significant during the refinement process. This is justified because of the energy increase arising from the water molecules surrounding the peptides and concomitant increase of the electrostatic energy terms.

A thin layer of solvent was created around the solute with a minimum distance of 4 Å between the oxygen atoms of the solvent and a maximum distance of 8 Å. Solvent boxes of 18.856 Å (TIP3 model) containing 216 water molecules [92] or 29.096 Å (DMSO) containing 208 molecules [93] were duplicated until the solute was immersed. The

[3] DIHE and CDIH energies are used in XPLOR. DIHE energy increases if the dihedral angles are not in the allowed region whereas CDIH energy increases when dihedral angles differ from those derived experimentally.

Table 3.12: **Restrained molecular dynamics protocol used for refinement in explicit solvent**: The coordinate restraints are indicated by fixed (F) and harmonic (H) positional restraints. The atom name and the scaling factor are also indicated for the harmonic restraints.

Stage	Pos. Res.	Length (ps)	Temp. (K)	K_{ang} (kcal.mol^{-1}rad^{-2})	K_{impr} (kcal.mol^{-1}rad^{-2})
Mini	F	20			
Mini	H C^{α}=10	40			
Heating	H C^{α}=10	1.25	100→500	50	5
Refinement	0	12.5	500	50	5
Cooling	0	1.25	500→100	50→500	5→500
Mini	0	100			

refinement in explicit solvent starts with the relaxation of the solvent around the solute (two minimizations). The solute is initially kept at a fixed position and progressively released. Three stages of MDSA bring the system from 100 K to 500 K and cool it down again. The weight of the improper and angle terms is lowered during the heating and at high temperature stages. Without these modifications, the conformation would stay very close to the initial one. During the cooling stage the energy constants are brought back to the values defined by the force field.

Full relaxation matrix: protocol C

The relaxation matrix calculations were performed in XPLOR using NOE intensities at various mixing times (55 ms to 400 ms). During the FRM calculation, the comparison between the back-calculated and the observed distances is minimized (R factor). The "model-free" approach was chosen to describe the relationship between the NOE intensities and the flexibility of the molecule. The backbone protons were assumed to be less flexible than the side chain protons. This was implemented in the protocol by setting the order parameter to 0.85 for backbone-backbone cross peaks, 0.80 for backbone-side chain cross peaks and 0.65 for side chain-side chain cross peaks. The overall correlation times of all peptides were set to 0.75 ns.

The refinement of the structures calculated using protocol A was achieved by MDSA followed by three minimizations. The temperature was linearly decreased from 1000 K to 75 K during 1 ps followed by three cycles of minimization of 30 steps each.

Full relaxation matrix: protocol D

Protocols B and C were combined yielding the full relaxation matrix calculations in explicit solvents.

3.10 Docking of peptides 1-4 into $\alpha_v\beta_3$ integrin

Docking of the water structure of peptides 1-4 was performed using the AUTODOCK 3.05 [94]. The following section describes the parameters used to dock peptides into the binding site of $\alpha_v\beta_3$ integrin.

Preparation of the receptor. The crystal structure of the extracellular domain of the $\alpha_v\beta_3$ integrin co-crystallized with Cilengide (PDB entry code = 1L5G) was used for docking. Calculations were performed at the RGD binding site of of $\alpha_v\beta_3$, which is clearly identified in this crystal structure.

Hydrogens were added to the receptor using a modified version of the generate_easy.inp script in CNS (available at the CNS website). Subsequently, conformation of side chains was optimized using a 200 steps Powell minimization keeping the position of the backbone atoms fixed.

Manganese ions (Mn^{2+}) of the original X-ray structure were replaced by calcium ions (Ca^{2+}) for which parameters are available for AUTODOCK [94]. This substitution was also used by Marinelli *et al.* [29] in the docking of Cilengitide to the $\alpha_v\beta_3$ integrin.

Kollman charges were used to calculate partial atomic charges and solvation parameters were automatically added by AUTODOCK. Calcium ions were given a charge of +2. Parameters of calcium ions, required for the calculation of the grid map of the receptor, were taken from "lj4.py" script available at the AUTODOCK website. This script calculates Lennard-Jones parameters for various pairs of atoms (Table 3.13). These parameters were manually added to the grid file. The grid map was calculated using AUTOGRID with 121^3 points (intact peptide 1 to 4) and 61^3 points (Cilengitide, intact peptides 1, truncated peptides 2-4), respectively. A grid-point spacing of 0.375 Å was used and the grid was centered on the center of mass of Cilengitide complexed with the

$\alpha_v\beta_3$ integrin. AUTOGRID was used to calculate the electrostatic (distance-dependent dielectric constants of -0.5) and steric maps of the receptor.

Table 3.13: Lennard-Jones parameters for Ca^{2+} in AUTODOCK.

atoms i-j	Radius i-j (Å)	eps i-j (kcal.mol^{-1})
Ca-C	2.99	0.0426195
Ca-N	2.74	0.0441045
Ca-O	2.59	0.049302
Ca-S	2.99	0.049302
Ca-H	1.99	0.0155925

Preparation of the ligands. Partial Gasteiger charges [95] added in AUTODOCK were used rather than charges calculated with semi-empiral methods. Peptides structures used for docking were the closest-to-mean NMR calculated structures of peptides 1-4 and Cilengitide was extracted from the PDB file 1L5G. The peptide backbones were kept rigid and only the side chains were allowed to rotate.

The size and the number of potential rotatable bonds in peptides 2, 3 and 4 are larger than the maximum number of rotatable bonds in AUTODOCK (max. 32 rotations). Therefore, peptides 2-4 were modified in order to reduce the size of the non-peptidic moieties. Two docking runs were performed using peptides 2 to 4 in order to assess their bound-state conformations. First, the docking made use of truncated peptides in order to improve the convergence and to reduce the number of rotatable bonds whereas the entire peptide was used in a second run.

In order to reduce the number of possible rotatable bonds (30, 69 and 106 for peptides 2, 3 and 4 respectively) and to ensure that the box was large enough to contain the ligands, Ncc residues were truncated after C^{β}. Similarly, the Lym residue in peptides 3 and 4 were truncated after C^{Q}. This had reduced the number of rotatable bonds to 15 in truncated peptides 2-4. The truncated structures were used to analyse the results of the docking. In order to perform the docking on entire peptides, in addition to the 13 rotatable bonds of the peptide core, the first 4 rotatable bonds of Ncc9 were allowed to rotate while others were kept rigid.

Docking. Docked conformations were sampled using the Lamarckian genetic algorithm (LGA) and a local search implemented in AUTODOCK. An initial population of 50 randomly placed individuals were used for the LGA with 1.5×10^6 maximum number of energy evaluations, 2.7×10^4 maximum number of generations, an elitism of 1, a mutation rate of 0.01 and a crossover rate of 0.8. The local search was performed using the pseudo-Solis and Wets algorithm for maximum of 300 iterations. The LGA-LS run was repeated 50 times and produced 50 conformations. The structures of truncated ligands were clustered according to an rmsd of 2.0 Å for calculations performed in a 61^3 points grid.

Part III

RESULTS AND DISCUSSION

Chapter 4

NMR Experiments

NMR parameters related to the conformation of peptides were measured and used in structure calculations and conformational analysis. This chapter presents the results of the NMR experiments and describes the procedures that were used to extracted the data from the NMR spectra. The peptides were studied in two solvents, H_2O and DMSO. The first solvent is more physiologically relevant but results obtained in DMSO can be compared with related peptides that have been solved in this solvent.

4.1 Resonance assignments

A method described by Wüthrich [79] was employed to assign ^1H resonances of the peptides. The assignment of the resonances and establishing sequential connectivities was carried out in XWINNMR and ANSIG. The known primary structure of the peptides was confirmed during the process of assignment of the ^1H, ^{13}C and ^{15}N resonances, particularly via the analysis of 2D NOESY and HMBC experiments.

4.1.1 Analysis of the homonuclear spectra

In peptides there are no scalar proton-proton couplings across peptide bonds; each spin system is confined to one amino acid residue. Individual spin systems were identified from 2D TOCSY experiments, more specifically their fingerprint regions. The separation of signals in this region allowed an easy identification of cross peaks (Fig. 4.1).

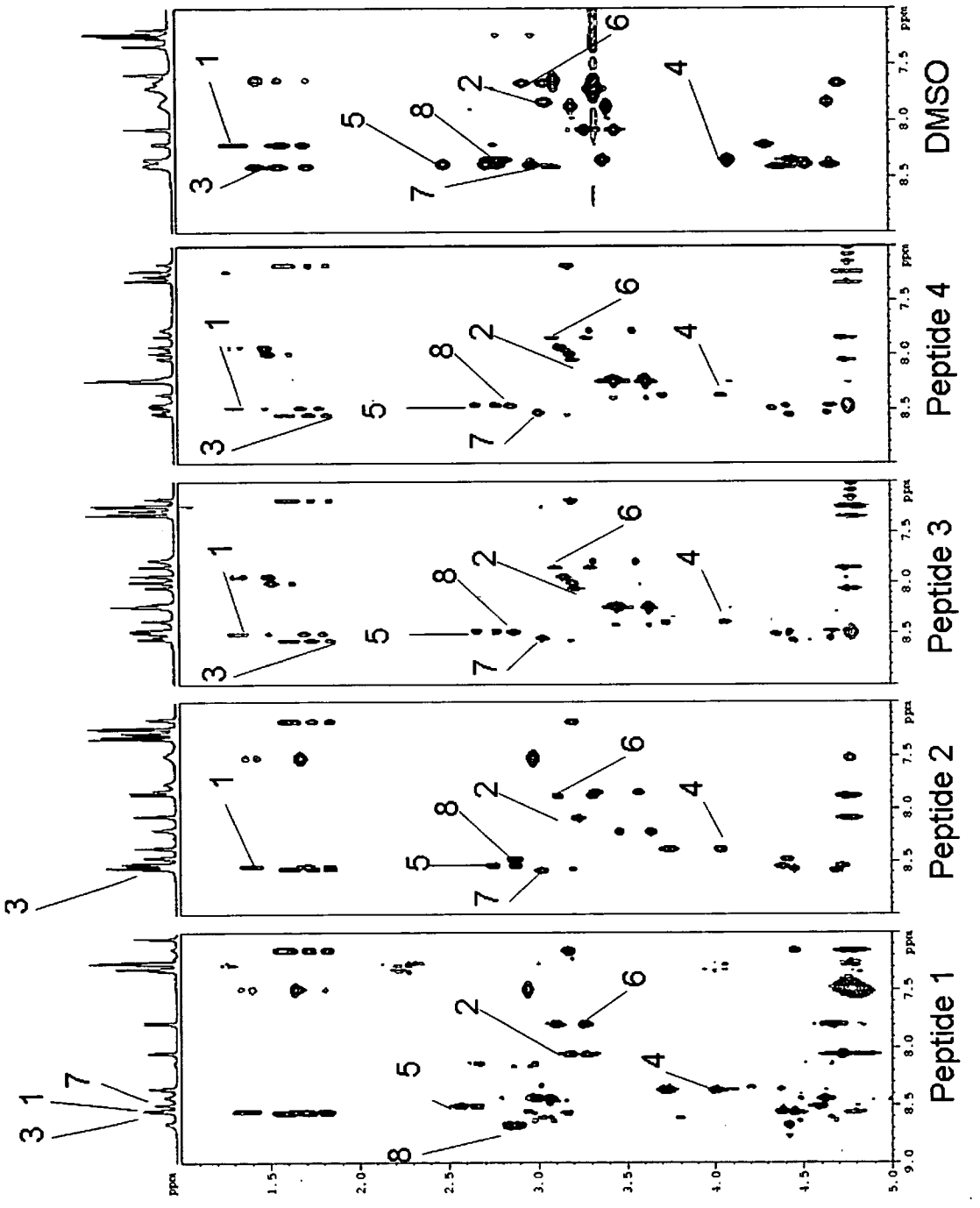


Figure 4.1: **Fingerprint region of the TOCSY spectra for peptides 1-4 in water and peptide 2 in DMSO:** From the bottom to top peptide 1, 2, 3, 4 in water and peptide 2 in DMSO. Residue numbers and H^N chemical shifts are indicated for each spectrum.

The sequential connectivities were established based on the analysis of 2D ROESY and 2D NOESY spectra. Starting from glycine, which is a unique residue, and shows a characteristic pattern of an AMX system, the sequential assignment was achieved by identifying the H^N (i)H^α (i+1) cross peaks. It was possible to complete the sequential assignment of residues 1 to 8 in all peptides unambiguously. The cross peaks observed between the amide proton of cysteines and the H^α of the other cysteine ($6H^N$ - $2H^\beta$ and $2H^N$ -$6H^\beta$) confirmed the presence of a disulfide bridge. Likewise, cross peaks between $1H^N$ and $8H^\delta$ confirmed the presence of the CLL link.

All protons of the peptide core were unambiguously assigned, apart from the degenerate protons of the ethylene groups and labile protons of the arginine and aspartic acid residues. The assignment of resonances of Ncc was only possible for protons at the beginning and end of the residue, but not for the ethylene glycol parts in peptides 2 and 3. In peptide 4, the signals of protons in residues Ncc10, Ncc11 and Ncc12 were strongly overlapped and no specific assignment of these groups could be made. In peptide 3 and 4, protons of residue Lym were assigned starting from the backbone untill H_{10} which is the junction between two symmetrical branches. Further resonances could not be specifically assigned.

Heteronuclear experiments were used to assign ^{13}C and ^{15}N resonances starting from the proton assignments obtained above. The ^{15}N resonances were assigned using ^{15}N coupled ^1H-^{15}N HSQC spectra, which were also used to extract $^3J_{H^N H^\alpha}$ coupling constants. ^1H-^{13}C HSQC and ^1H-^{13}C HMBC experiments were used to assign ^{13}C resonances. In peptides 3 and 4, the presence of the Lym residue increased the overlap in the aliphatic region of the spectra. The ambiguities were resolved by using a ^1H-^{13}C HSQC with multiplicity editing. In this experiment, the CH and CH_3 groups yield cross peaks with opposite phase to those of CH_2 groups. Heteronuclear couplings between two consecutive residues mapped by HMBC experiments confirmed the sequential assignments (via $^3J_{H_\alpha CO_{i-1}}$) and the Cya8-CLL-Lys1 connectivities (cross peak between $8H^\beta$ and $8H^\delta$ through the thioether bond). The HMBC experiments also provided the assignment of quaternary carbons, the carbonyl resonance of the backbone in particular.

The chemical shifts of ^1H, ^{13}C and ^{15}N are given in appendix F. The proton chem-

ical shifts were referenced using the water or DMSO signals, while the heteronuclear resonances were referenced using the ratio of gyromagnetic ratios (γ_H/γ_X) [83]. In the absence of stereospecific assignments of prochiral protons, indices 1 and 2 were used for down and high field signals, respectively.

4.2 Determination of distance restraints. NOESY, ROESY and T-ROESY experiments.

Several experiments utilizing dipolar interactions such as NOESY, ROESY and T-ROESY were carried out in order to determine the interatomic distances.

The set of distance restraints used for the initial, exploratory structure calculations were extracted from various two-dimensional experiments mapping the dipolar interactions. Fast correlation times of medium-size molecules, in particular peptide 1, are associated with weak nuclear Overhauser enhancements. ROESY experiments were therefore also acquired in order to increase the intensity of cross peaks that are important for structure determination of molecules by NMR.

The mixing times for the two-dimensional experiments were optimized using build-up curves (50 ms and 1000 ms) obtained by acquiring a series of 1D NOESY and 1D ROESY experiments. The amide region of the spectra was selectively inverted and the build up of the magnetization was observed on the remaining protons. For all peptides, the NOEs obtained were negative i.e, the cross peaks had the same sign as the inverted peaks. The choice of the mixing time for 2D experiments was based on the quality of the enhancements. This generally increases with increasing mixing times, however, for long mixing times the amount of spin diffusion in NOESY experiments increases as does the TOCSY transfer in ROESY experiments. Conversely, NOESY experiments using short mixing times contain only short distance information and often display artifacts (J-cross peaks from zero quantum coherence and t_1 ridge from subtraction artifacts) which causes problems when used in structure determination.

Figure 4.2 illustrates the dependency of the enhancements upon the mixing time for peptide 2 in DMSO. The intensity of the enhanced peaks grows linearly at short mixing times, reaches a plateau and eventually starts decreasing for longer mixing times. The

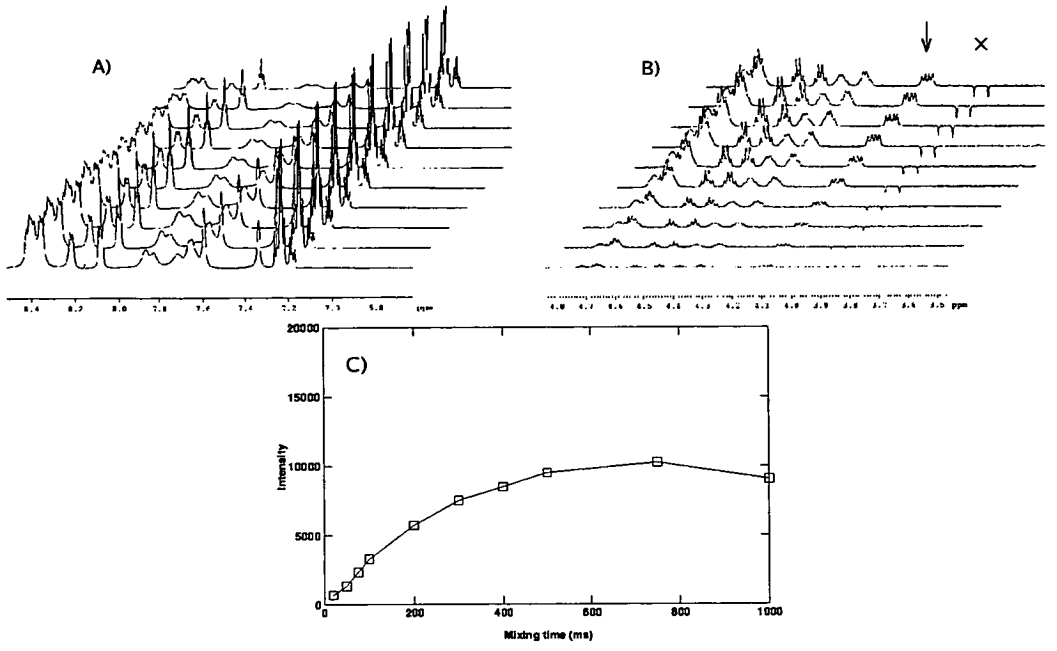


Figure 4.2: **NOE build-up curves followed by 1D NOESY experiments:** The enhancement of the aliphatic region (B) is followed for peptide 2 in DMSO after inversion of the amide region (A). The intensity of the $4H^{\alpha}$, indicated by an arrow in figure B, is plotted versus the mixing time of the experiment (C). Negative peaks arising from Ncc (indicated by "X") residues illustrate a tumbling regime different from the peptide core.

"reference" mixing time was chosen to be in the linear part of the NOE enhancement curves and the value of 100 ms was used for DMSO. In water the NOE enhancements increased until a mixing time of 1000 ms; initially linearly until 400 ms after which the rate of change decreased. Mixing times of 200-250 ms were therefore chosen to acquire spectra in H_2O, which represent a good compromise between intensity of the NOE peaks and limited spin diffusion. Earlier onset of maxima of NOE enhancements in DMSO (750 ms) compared with water (1000 ms) indicates faster relaxation in DMSO, and more efficient and earlier onset of spin diffusion in DMSO. This is a consequence of slower molecular tumbling in DMSO. This also implied that the spectra obtained at identical mixing times but in different solvents do not contain exactly the same information. In an attempt to alleviate this problem the mixing times chosen for DMSO were 50-100 ms shorter than those in H_2O [38, 96, 97].

The magnetic field at which NOESY spectra are acquired also contributes to the efficiency of NOE transfers. Therefore spectra were acquired and compared separately on

800 MHz and 600 MHz spectrometers for peptide 2 (water and DMSO), 3 and 4. Poor NOESY transfer for peptide 1 lead to acquire its NOESY spectra also on a 900 MHz spectrometer[1]. For comparison, a NOESY spectrum of peptide 3 was also acquired at this field. The 900 MHz spectra contained more cross peaks.

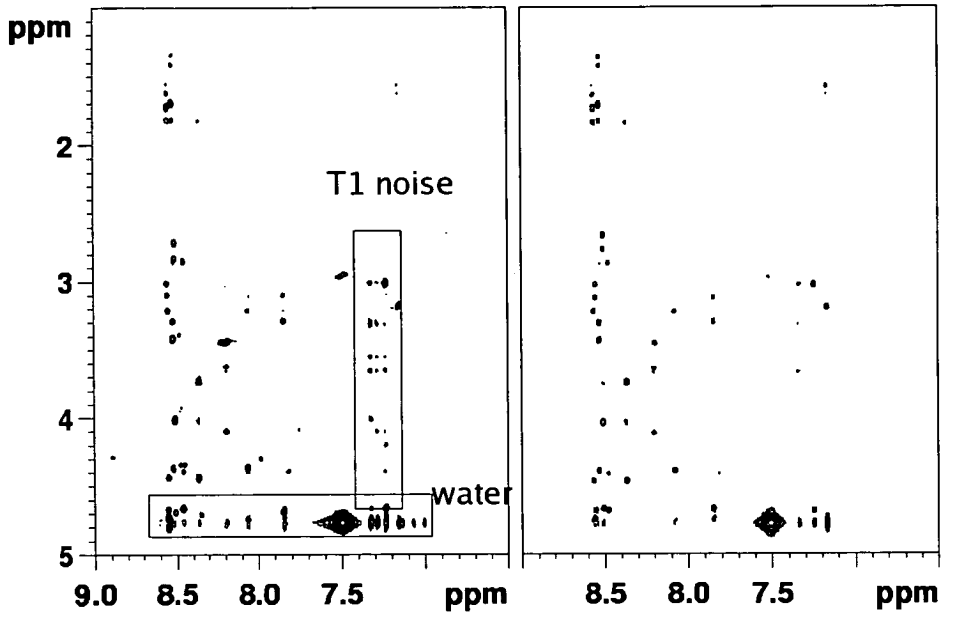


Figure 4.3: **Comparison of ROESY and T-ROESY spectra for peptide 2 in water.** Fingerprint regions of ROESY (left) and T-ROESY (right) spectra acquired at 800 MHz. The boxes indicate regions for which T-ROESY experiment improved the quality of the spectra.

A comparison of the fingerprint region of ROESY and T-ROESY spectra in Fig. 4.3 shows that T-ROESY greatly improves the intensity of H^N cross peaks. It also shows that the exchange with the water is limited which allow the assignment of the cross peaks near the water. It can also be noticed that the t_1 noise in the aromatic region is reduced in T-ROESY spectra.

A series of 2D NOESY experiments using five or six mixing times from 55 ms to 400 ms were acquired for peptides 2-4. For peptide 2, NOESY spectra were also acquired in D_2O and DMSO. When these experiments were collected, an efficient purging scheme for removal of zero-quantum artifacts from NOESY spectra had been published [65]. This scheme was incorporated into the NOESY pulse sequences. A purging element, consisting of an adiabatic inversion pulse and a low level pulsed field-gradient, inserted

[1] The Henry Welcome Building for Biomolecular NMR Spectroscopy, Birmingham

into the mixing time reduced zero-quantum contributions from J coupled protons. This reduction is particularly important at short mixing times, where it is also followed by a reduction of t_1 noise resulting in much higher quality of NOESY spectra. In particular, low intensity of the cross peaks at short mixing times suffers from t_1 noise and mixed phase of J-mediated contributions to the cross peaks. Although in principle, the latter have zero integrals, this could be difficult to achieve due to the long dispersive tails of these peaks. Obtaining "clean" spectra at short mixing times is essential for structure calculations as these are not contaminated by spin-diffusion.

4.3 Determination of scalar coupling constants

The following section presents NMR experiments and methods that were used to extract the three-bond coupling constants of the peptides. Three-bond coupling constants are related to dihedral angles via Karplus equations and thus the goal of these experiments is to obtain extra information about the local geometry of peptides. Coupling constants play an important role in the structure determination of peptides due to the limited number of dipolar interactions present in this class of compounds. Angles such as ϕ, ψ and χ^1 report on the conformation of the peptide bonds and side chains and are therefore valuable supplements to NOEs. There are two reasons why more than one coupling constant is required in order to restrict a dihedral angle within certain range. Firstly, Karplus curves are degenerate; a unique relationship between a coupling constant and a dihedral angle does not exist. Secondly, it is impossible to ascertain from one coupling constant whether its value reflects a rigid geometry or is a result of some conformational averaging. Only one heteronuclear coupling constant ($^3J_{H_i^\alpha H_{i+1}^N}$) can be measured that describes the ψ angle. Furthermore, this is a small coupling constant that varies little with the angle. Therefore, only coupling constants related to ϕ and χ^1 angles were measured.

In compounds with natural abundance of isotopes, ^{1}H-^{1}H homonuclear and ^{1}H-X (X = ^{13}C or ^{15}N) heteronuclear coupling constants can be measured provided mg quantities of material are available. On the contrary ^{13}C-^{13}C or ^{13}C-^{15}N coupling constants, accessible in labeled proteins, cannot be measured because of the low natural abundance of such pairs in unlabeled compounds. Numerous methods have been described

in the literature for the measurement of coupling constants in compounds with natural abundance of isotopes [98]. It was therefore decided to carry out preliminary studies in order to select experiments that are most appropriate for small peptides.

4.3.1 Direct methods

As described in chapter 3, the extraction of coupling constants from spectra by measuring the frequency difference between the individual lines of multiplets is the most straightforward method. The cases where this method can be used successfully are described next.

1D proton experiment with water presaturation: Occasionally, the dispersion of signals in 1D ^{1}H spectra was sufficient to allow extraction of proton-proton coupling constants. This was only possible for ^{12}C attached protons. The H^N protons were either overlapped or too broad to yield the coupling constants directly. In practice, few coupling constants were measured in this way and 1D TOCSY was usually the experiment of choice as here overlap-free spectra were obtained.

1D TOCSY experiments: A series of selective 1D TOCSY experiments was acquired to resolve the spin-system of each residue and to measure the $^3J_{H^\alpha H_2^\beta}$ and $^3J_{H^\alpha H_3^\beta}$ coupling constants. Overall, few proton-proton coupling constants could directly be measured from the 1D TOCSY spectra because of the degeneracy of the H^β protons. In principle also, the $^3J_{H^N H^\alpha}$ couplings could be measured from 1D TOCSY spectra, but because, as discussed below, there is a better method to achieve this, 1D TOCSY spectra were not used for this purpose. $^3J_{H^N H^\alpha}$ are the largest and best parameterized amongst the couplings related to the ϕ angle; accurate values of these coupling constants are therefore an important source of information about this dihedral angle.

^{15}N coupled ^{1}H^{15}N-HSQC: 1D cross sections were extracted from these 2D spectra and the ^3J coupling constants were measured using the peak picking routine of XWINNMR (Fig. 4.4).

All backbone $^3J_{H^N H^\alpha}$ coupling constants could be extracted for all peptides, except for

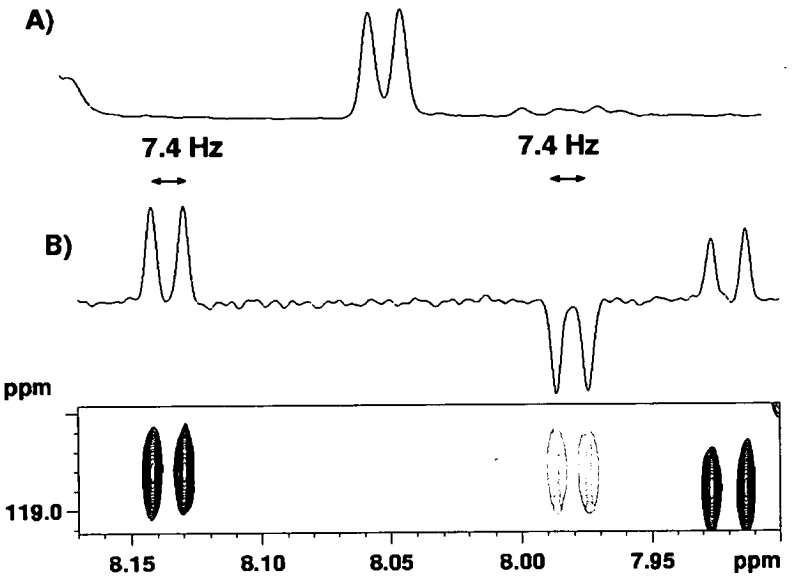


Figure 4.4: **Extraction of $^3J_{H^NH^\alpha}$ from $^1H^15N$-HSQC:** The NH doublet of Cys2H^N extracted from a 1D spectrum (A) and the cross section of the 2D spectrum (B) of peptide 2 in water. The line width of the Cys2H^N is narrower in the 2D spectrum and the coupling constant could be extracted more accurately. The spectrum shows the one-bond proton-nitrogen coupling in antiphase. The digital resolution was 0.2 Hz per point.

residues Asp5 and Cys2 of peptide 2 in DMSO which displayed large line widths (Fig. 4.5).

The values of $^3J_{H^NH^\alpha}$ extracted from 1D spectra were systematically lower compared with those extracted from the HSQC spectra. This is a well known consequence of measuring coupling constants from poorly-resolved multiplets caused by the quadrupolar relaxation of the ^{14}N [62]. Since the value $^3J_{H^NH^\alpha}$ coupling constants have the largest amplitude and are the best parametrized amongst the coupling constants related to the ϕ angle, accurate values of this coupling constants are important for dihedral angle determination.

HETLOC expèriments: Cross peaks with poor signal-to-noise ratio or overlapped cross peaks are unsuitable for computerized coupling constant determination. Coupling constants were manually extracted from such multiplets using the dual display capability of XWINNMR. In cases where even this approach was problematic, only a qualitative assessment of coupling constants was done and these were classified as

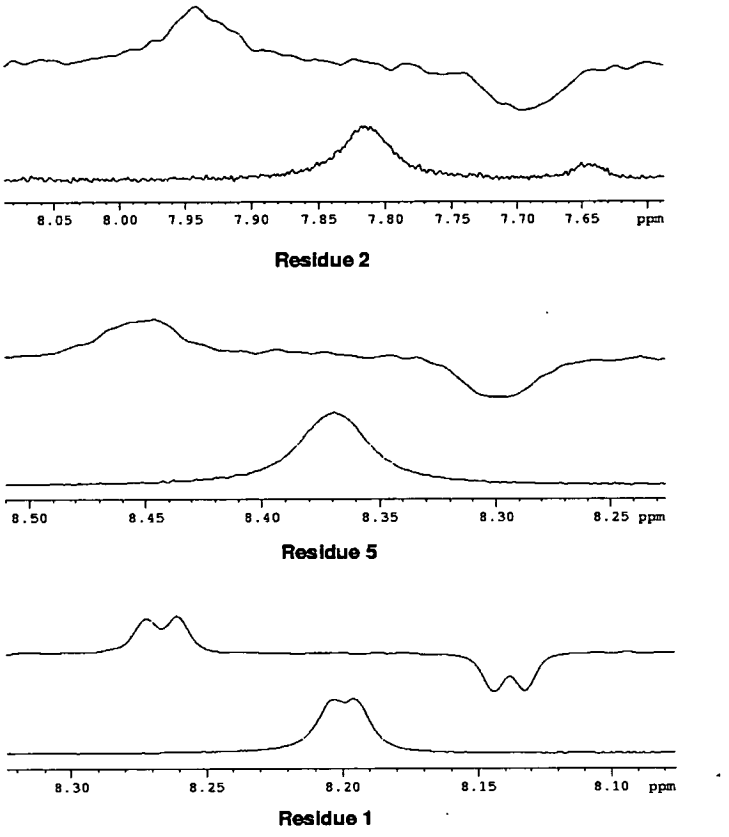


Figure 4.5: **Comparison of the line widths of ^{15}NH protons of residues 1, 2 and 5 for peptide 2 in DMSO extracted from a ^{15}N coupled ^{1}H-^{15}N HSQC:** The $^{3}J_{H^{N}H^{\alpha}}$ could be extracted only for residue 1. Figures show for each residue the 1D doublet from 1D-TOCSY and the cross-section extracted from 2D ^{15}N HSQC.

small or large.

The limitation of this technique is that proton-proton coupling constants must exist between directly- and long-range bonded protons, which are then used to spread the magnetization using proton-proton TOCSY transfer. No inter-residue couplings can therefore be measured. Also, only long-range coupling constants of protonated hetero atoms can be measured. This, for example, excludes the carbonyl backbone atom. Some of the coupling constants that in theory are accessible could not be measured either. For example, $^{3}J_{H^{\alpha}C^{\gamma}}$ coupling constants were not measured because of the overlaps with the other aliphatic protons and/or the proximity of the water signal to the resonances.

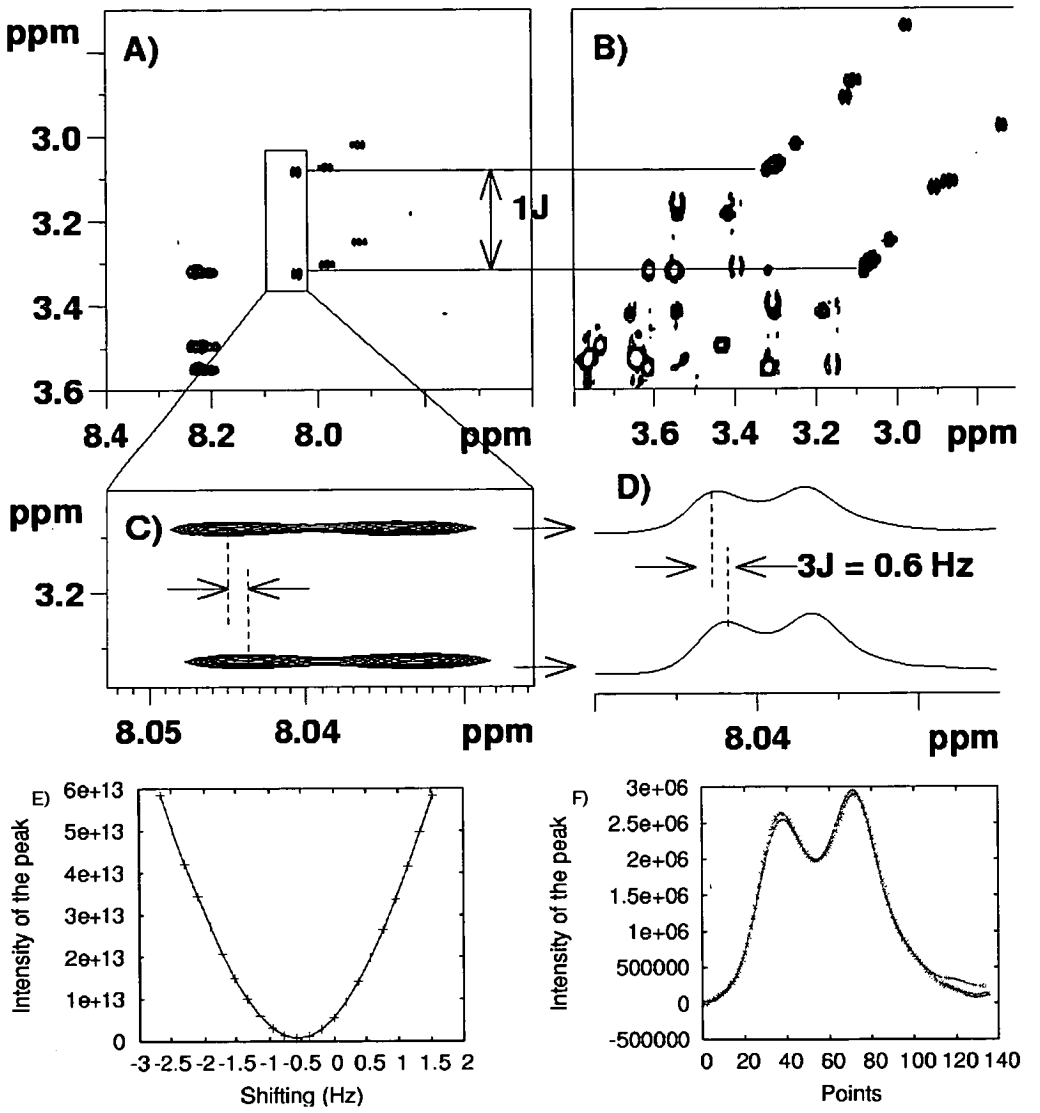


Figure 4.6: **Example of extraction of heteronuclear coupling constants from ^{13}C ω_1 half-filtered TOCSY spectra: Peptide 3, Cys2, $^3J_{H^N C^\beta}$ coupling constant:** The fingerprint (A) and the aliphatic (B) regions show the separation of cross peak components in F_1 by $^1J_{CH}$ coupling constants. The expanded regions show the $H^N H^\beta$ cross peaks (C) and the cross section (D). The area of the subtracted peaks as a function of the peaks displacement in Hz is shown in (E). The minimum at (0.6 Hz) corresponds to the three-bond coupling constant. The shifted overlaid peaks are visualized in (F)

4.3.2 Indirect methods

Determination of most of the coupling constants related to the peptide dihedral angles required a reconstruction or comparison of cross peaks. A simple measurement of the peak frequencies was not possible.

Gradient selected phase-sensitive HMBC experiments: This type of experiment yields complex multiplets with a mixed phase in the proton F_2 dimension as explained in chapter 2. Two techniques, described in the literature were tested [69, 70].

The feasibility of this approach was tested using a sample of $^{13}C_1$ D-glucopyranose for which the values of coupling constants are known and can also be determined from its 1D ^1H spectra (Fig. 4.7). Analysis of the coupled and decoupled ^1H spectra of glucose allowed the two coupling constants to be measured, independently from the HMBC experiment This also served as a test for another implementation of the HMBC experiment, discussed later, that was used for determination of long-range coupling constants. The coupling constants were determined by minimizing the following function:

$$\mathcal{A} = |S_{coupled} - 0.5(S_{decoupled_{leftshifted}} + S_{decoupled_{rightshifted}})| \tag{4.1}$$

A modified version of the home written program, "decoupantiphase", performed the fitting. Coupling constants determined in the frequency domain were $^2J_{C_1^\beta H_2^\beta}$=6.2 Hz and $^3J_{C_1^\beta H_5^\beta}$=0.4 Hz.

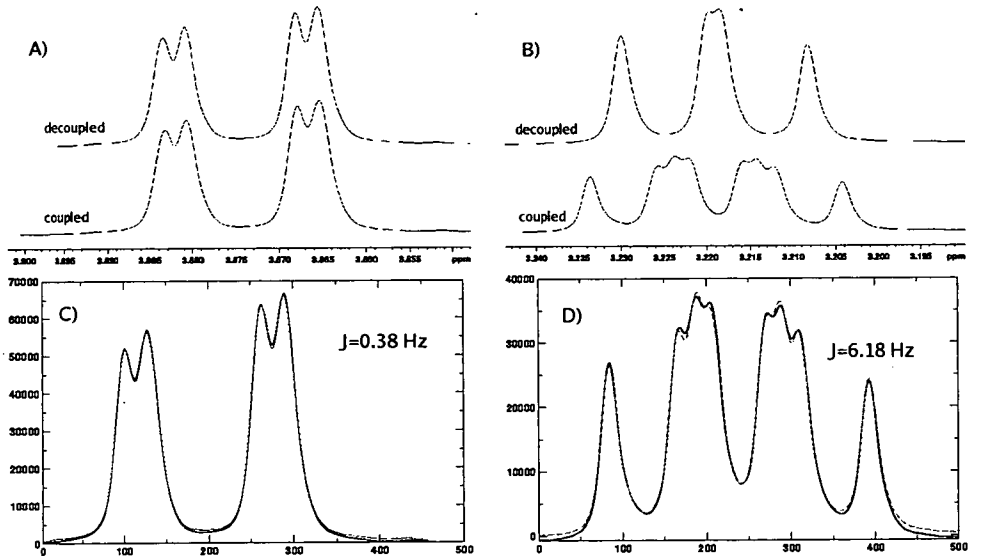


Figure 4.7: **Determination of the $^3J_{C_1^\beta H_5^\beta}$ (left) and $^2J_{C_1^\beta H_2^\beta}$ (right) coupling constants:** Coupling constants determined by reconstructing the coupled multiplet from two shifted decoupled multiplets (A and B). The experimental and reconstructed coupled multiplets are overlaid in C and D and indicated by solid and dashed lines, respectively.

A 2D HMBC spectrum was acquired along with 1D coupled and decoupled ^{1}H spectra of glucose focusing on the heteronuclear coupling constants ^{1}H -^{13}C$_1$. Two long-range coupling constants of C_1^β were analyzed, one small and one large. The delay in the HMBC experiment was set to 82.410 ms and the acquisition time in the directly-detected dimension of 0.85 s was used. The data were zero filled to 32 K point increasing the digital resolution to 0.076 Hz per point. Two long-range cross peaks mediated via $^2J_{C_1^\beta H_2^\beta}$ and $^3J_{C_1^\beta H_5^\beta}$ were analyzed using a home written program "decoupantiphase" and the values of 0.3 and 6.1 Hz were obtained, i.e. identical to those determined from the analysis of 1D spectra.

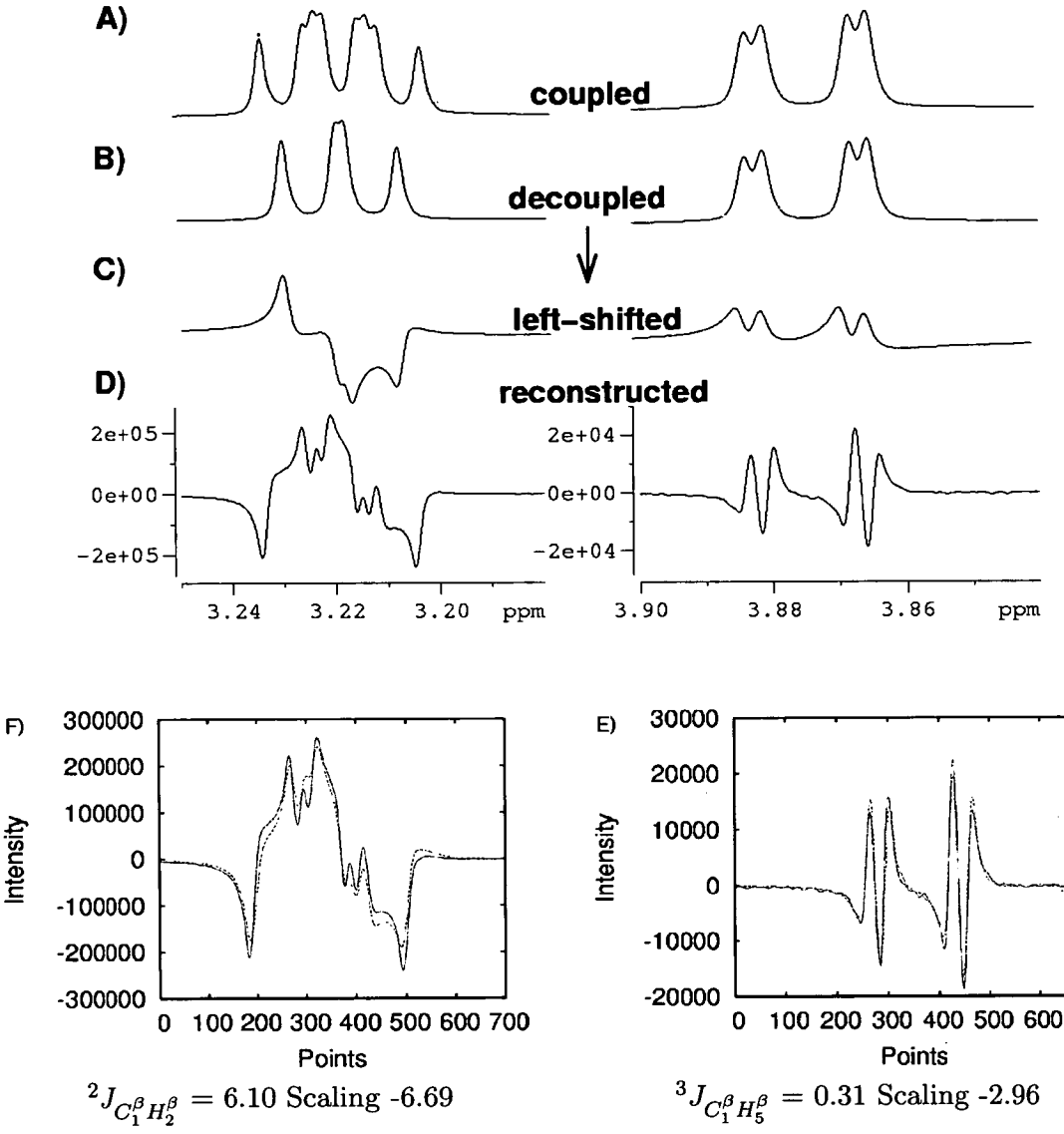
The examined protons had two very different coupling constants and the reconstruction of the coupled multiplet using two decoupled multiplets showed that this approach works successfully also for very small coupling constants.

The XWINNMR macros written by Edden *et al.* [69] were also tested and returned values $^2J_{C_1^\beta H_2^\beta}$=5.5 Hz and $^3J_{C_1^\beta H_5^\beta}$=1.6 Hz. Comparing the results of the two methods revealed that the Powell minimization used in "decoupantiphase" combined with a restriction of the range of possible coupling constants gave better results. This method was chosen to extract the coupling constants. Simple modifications of the minimized function made this program applicable also to the analysis of in-phase multiplets obtained in refocused/decoupled HMBC or X-filtered experiments.

The intensity of the HMBC multiplets is related to the amplitude of the coupling constants. The amplitude of multiplets depends on factor $sin(\pi J_{XH}\Delta)$, where Δ delay was calculated as $\Delta = 1/2J$ and the values of J between 4 and 6 Hz were used. For smaller coupling constants, lower intensity of coupling constants is to be expected. Smaller scaling factors are therefore associated with smaller coupling constants and the results could therefore be verified by comparing the two variables between related cross peaks.

Following preliminary studies on glucose, the HMBC experiments were acquired for peptides 1, 3 and 4. Shorter relaxation times of peptides meant that optimal acquisition times of 330 ms were used in the directly-detected dimension.

Fig. 4.9 shows the coupling constant determination of peptide 3 using 1D TOCSY



$^2J_{C_1^\beta H_2^\beta}$ = 6.10 Scaling -6.69

$^3J_{C_1^\beta H_5^\beta}$ = 0.31 Scaling -2.96

Figure 4.8: **Extraction of $^2J_{C_1^\beta H_2^\beta}$ and $^3J_{C_1^\beta H_5^\beta}$ coupling constants from HMBC spectra of ^{13}C labeled glucose:** (A) ^{13}C coupled and (B) decoupled ^{1}H spectra of glucose. (C) Mixed phase multiplets obtained by inverse FT, left shifting the FID and FT that simulate chemical shift and J evolution of HMBC spectra. (D) Reconstructed multiplets containing the heteronuclear coupling in antiphase best fitting the HMBC multiplets. E and F show overlay of multiplets from D with the experimental HMBC multiplets.

as a templates. A visual comparison shows that the achieved reconstruction is not always ideal. In other instances the reconstruction is satisfactory, but the value of the coupling constant was suspiciously small. This is due to the fact that the scaling and

the shifting, the two parameters of the search, are not independent. Very different combinations of scaling factors and coupling constants could reproduce the experimental multiplet equally well. A simple inspection of the intensity of the cross peaks in the HMBC spectra could therefore, in some cases, confirm or invalidate the values of coupling constants returned by the program. For example, multiplets in Fig. 4.9 E,F although showing visually a good fit have a large discrepancy between the returned and estimated scaling factors based on the cross peak intensity indicating that the returned coupling constant is in error.The signal-to-noise ratio is also a limiting factor in these reconstructions ; in particular, for small coupling constants the increased noise could significantly distort the shape of a cross peak.

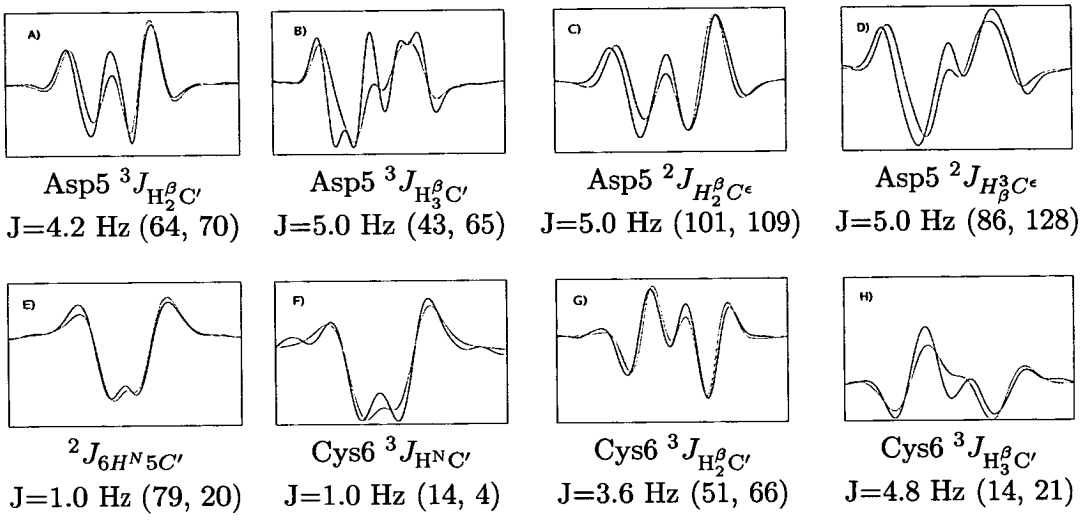


Figure 4.9: **Examples of coupling constants extracted for peptide 3 using ^{1}H - ^{13}C HMBC and 1D TOCSY spectra:** The cross peaks extracted from the HMBC experiment (black) are compared to the reconstructed multiplets (red). The scaling factors (calculated and estimated) are indicated in parenthesis. The values of coupling constants in multiplets E and F are inconsistent with the intensities returned by the fitting program and those measured from a dual display.

This same approach was also tested for the determination of ^{1}H-^{15}N long-range coupling constants using peptide 3. Due to the lower natural abundance and the sensitivity of ^{15}N, 1856 scans were accumulated in each of 72 increments. From the experiments, two couplings could be extracted for Asp5 ($^3J_{NH_2^\beta}$ =1.9 Hz and $^3J_{NH_3^\beta}$ = 2.7 Hz). The values of these coupling constants are very similar to those extracted from the ^{15}N HETLOC experiment. However, the HMBC experiment is less sensitive and the in-

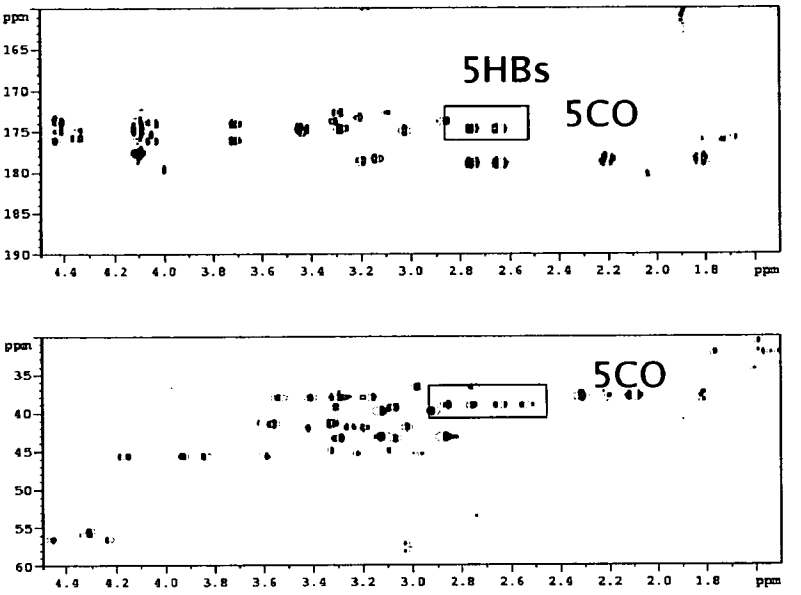
tensity of cross peaks mediated by small coupling constants is low. These cross peaks are easily compromised by low signal-to-noise ratios. The time spent in obtaining the values of two coupling constant is considerable and ^{15}N HETLOC experiments is a better option regarding the extraction of ^{1}H-^{15}N coupling constants related to ϕ and χ^1 than HMBC methods. $^{3}J_{H_i^{\alpha}N_{i+1}}$ coupling constants related to the ψ angle could, in theory, also be measured from HMBC spectra but their amplitude of less than 1.75 Hz makes this task difficult.

Reconstructing the HMBC multiplets from one-bond templates: Figure 4.10 illustrates extraction of Asp5 $^{3}J_{H_2^{\beta}C'}$ and Asp5 $^{3}J_{H_3^{\beta}C'}$ coupling constants from HMBC spectra of peptide 3 using one-bond cross peaks as templates.

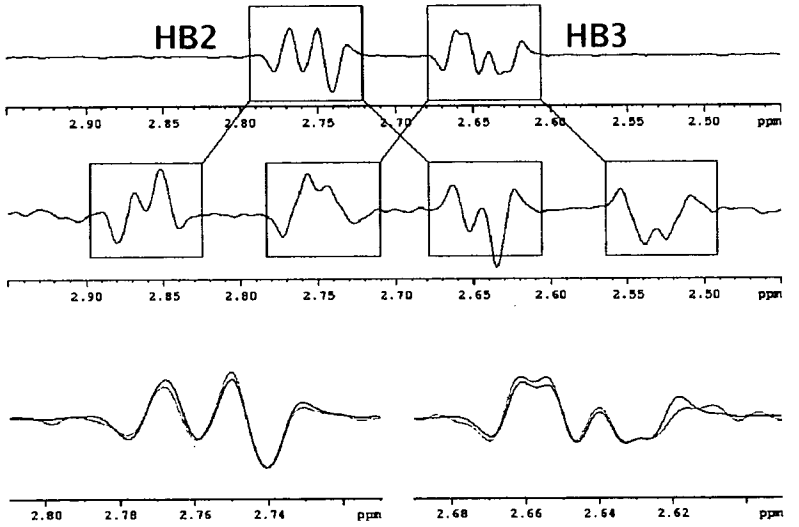
Decoupled HMBC and refocused HMBC: Two HMBC spectra are acquired, in which the heteronuclear couplings are refocused. The only difference between the two is that the X-nucleus decoupling is either switched on or off. This yields multiplets which either contain the long range coupling or not. The decoupled experiment provides a template that is used to reconstruct the coupled multiplet in the same manner as the scalar coupling constant was extracted from coupled/decoupled 1D spectra of ^{13}C glucose (Fig. 4.7).

The advantage of this experiment, in comparison with the antiphase HMBC, is that identical phase modulation is achieved for both the template and the coupled multiplet without any need for additional manipulations. Even more importantly, the scaling factor for the reconstruction of the multiplets is known and is 2. This simply reflects the fact that the template multiplet contains one less coupling constant than the coupled multiplet.

Larger number of $^{n}J_{CH}$ scalar coupling constants were be accurately determined by using this method (Fig. 4.11). For example, a coupling constant $^{2}J_{6H^{N}5C'}$ of 4.0 Hz was obtained by this method while antiphase HMBC experiments yielded 1.0 Hz. Coupling constants Cys6 $^{3}J_{H^{N}C'}$ and Asp5 $^{3}J_{H_2^{\beta}C'}$ illustrate this point. The coupling constants determined by reconstructed multiplets are Cys6 $^{3}J_{H^{N}C'}$=1.4 Hz and Asp5 $^{3}J_{H_2^{\beta}C'}$=3.8 Hz which differed dramatically from those obtained by the reconstruction



(a) 2D HMBC containing long-range (top) and one-bond (bottom) couplings. Boxes indicate the cross peaks of interest.



(b) Traces extracted from 2D HMBC spectra in one-bond HMBC (top) and multiple-bond HMBC (center). Overlay of experimental HMBC cross peaks modulated by Asp5 $^{3}J_{H_2^{\beta}C'}$ and Asp5 $^{3}J_{H_3^{\beta}C'}$ coupling constants and reconstructed multiplets using one-bond cross peaks (bottom)

Figure 4.10: Extraction of long-range coupling constants from OBMB HMBC experiments on peptide 3.

of antiphase HMBC multiplets.

The same coupling constants were determined by the intensity-based method using cross peaks mediated by $^{2}J_{6H^{N}5C'}$ and $^{2}J_{5H_2^{\beta}C^{\gamma}}$ coupling constants. Based on the ratio of intensities between the two-bond and three-bond cross peaks the ratio $^{2}J/^{3}J$ was

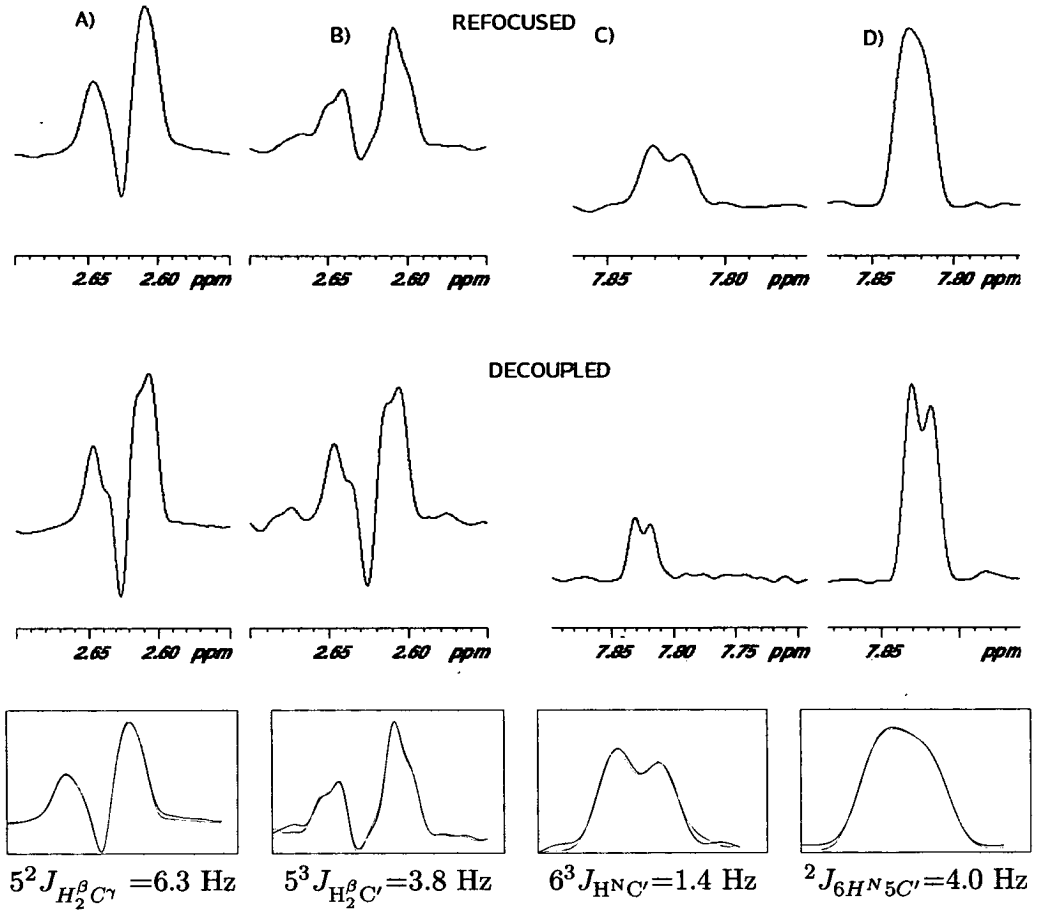


Figure 4.11: **Extraction of long-range coupling constants from refocused and decoupled HMBC experiments on peptide 3:** Corresponding decoupled and coupled multiplets are shown in the top and middle panels respectively. The coupling constants Asp5 $^3J_{H_2^\beta C'}$, $^2J_{5H_2^\beta C^\gamma}$, Cys6 $^3J_{H^N C'}$ and $^2J_{6H^N 5C'}$ correspond to columns A to D, respectively. The bottom panel shows non-decoupled cross peaks (black) are overlaid with the reconstructed cross peak obtained using the decoupled HMBC spectra (red).

determined as 3.33 and 0.41 for $6H^N$ and $5H_2^\beta$. This yielded the coupling constants Cys6 $^3J_{H^N C'}$=1.5 Hz and Aps5 $^3J_{H_2^\beta C'}$=3.6 Hz, i.e. close to the values obtained by multiplet reconstruction.

These experiments were particularly useful in determination of coupling constants of quaternary carbons. The presence of the water near the H^α region prevents reliable coupling constants determination of these protons. Accurate values could not be extracted for the $^3J_{H^N C^\beta}$ coupling constants but qualitative information, obtained via intensity-based estimate was helpful when analyzing dihedral angles. A small change

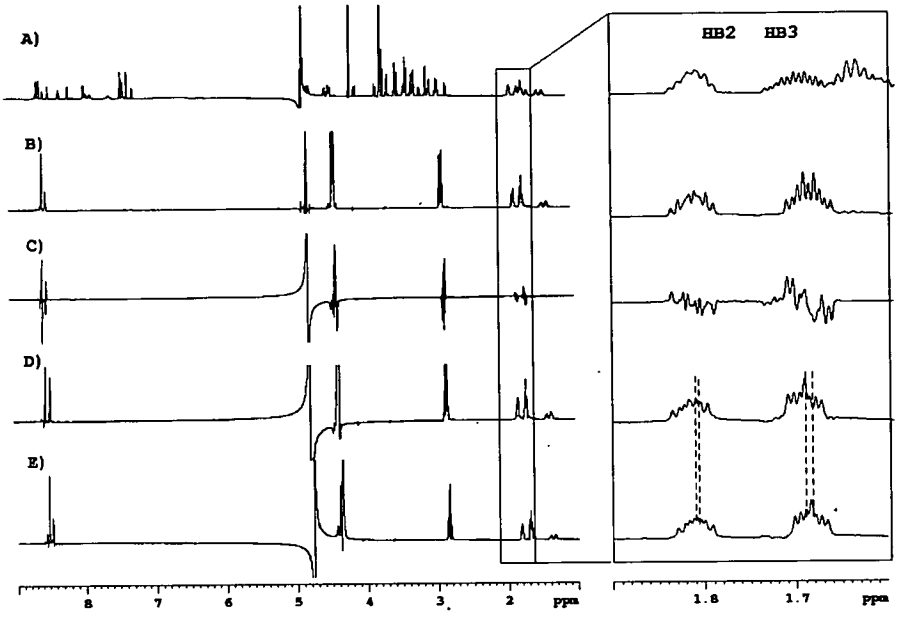


Figure 4.12: **Extraction of $^3J_{H^\alpha H^\beta_2}$ and $^3J_{H^\alpha H^\beta_3}$ from complicated multiplets using 1D COSY and 1D TOCSY spectra:** Spectra were acquired at 800 MHz for peptide 2 in DMSO at 298 K using 1024 scans and acquisition time of 853 ms. 1D ^{1}H spectrum (A) shows overlap in the region of the H^β of the lysine residue. The 1D TOCSY and 1D COSY are shown in (B) and (C). Expansion of H^β multiplets in (B) and (C). The reconstructed "decoupled" multiplets in (D) and (E). The chemical shift difference (Hz) between these identical multiplets corresponds to the coupling constant between H^α and H^β protons is indicated by dashed lines.

for the intensity ratio for $^2J$and ^3J cross peaks can return a very different value for the estimated coupling. Assuming an error of ± 0.5 Hz would therefore prevent any misinterpretation of these couplings in terms of structural restraints. The absence of a cross peak is also an indication of a small coupling constant and values $<$ 1Hz were used.

Determination of $^3J_{H^\alpha H^\beta}$ coupling constants via 1D TOCSY and 1D COSY: Using the principles of DIfference Sum with COsy (DISCO) analysis [99], addition of inphase and antiphase multiplets yielded simplified multiplets in which the antiphase splitting was removed, but was preserved as the chemical shift difference between the two resulting multiplets (Fig. 4.12).

Corresponding TOCSY and COSY multiplets were manually added and subtracted and the resulting peaks were shifted and added while the integral of the final signal

Table 4.1: Experiment used to extract the coupling constants.

Coupling constants	Experiments
ϕ dihedral angle	
$^3J_{H^N H^\alpha}$	^{15}N HSQC
$^3J_{H^N C'}$	^{13}C HMBC decoupled and refocused (multiplet reconstruction or intensity-based method)
$^3J_{H^\alpha C'_{i-1}}$	^{13}C HMBC decoupled and refocused (multiplet reconstruction or intensity-based method)
$^3J_{H^N C^\beta}$	^{13}C HETLOC and HMBC using the signal intensities
χ^1 dihedral angle	
$^3J_{H^\alpha H_2^\beta}$, $^3J_{H^\alpha H_3^\beta}$	1D TOCSY for resolved multiplets, 1D TOCSY/1D COSY for unresolved multiplets
$^3J_{H_3^\beta C'}$, $^3J_{H_2^\beta C'}$	^{13}C HMBC decoupled and refocused (multiplet reconstruction or intensity-based method)
$^3J_{H^\alpha C^\gamma}$	^{13}C HMBC decoupled and refocused (multiplet reconstruction or intensity-based method)
$^3J_{NH_2^\beta}$, $^3J_{NH_3^\beta}$	^{15}N HETLOC

was minimized by a computer. The displacements of the peaks corresponded to the three-bonds coupling $^3J_{H^\alpha H_2^\beta}$ and $^3J_{H^\alpha H_3^\beta}$. Crucial to success of these experiments was the use of a z-filter in the TOCSY sequence yielding pure absorption peaks.

This method was successfully employed to extract the homonuclear couplings constants $^3J_{H^\alpha H_3^\beta}$ and $^3J_{H^\alpha H_2^\beta}$ from all peptides with non degenerate H^β protons. This is a straightforward method offering good signal-to-noise ratios and a simple, reliable protocol for extraction of coupling constants. Its limitation is the required initial separation of resonances used for selected excitation (i.e. the H^α protons)

Table 4.1 summarizes the experiments that were used to determine coupling constants for all peptides. The coupling constants presented in Tables 4.2 and 4.3 do not include those obtained in antiphase HMBC experiments due to the insufficient accuracy of this method. The estimated error of the coupling: peak picking ±0.2 Hz ; computerized methods ±0.3 Hz (1.5 times the digital resolution) ; manually or intensity based ±0.5 Hz.

Several coupling constants related to χ^1 angles could not be extracted because of overlapping of ^{15}N or ^{13}C resonances in the 2D spectra or the proximity of the water

Table 4.2: **Summary of extracted coupling constants related to the ϕ dihedral angle:** The values in italics were measured using the intensity-based methods. The values in parenthesis were determined manually. "s" stands for small (< 1Hz) coupling constants.

Residue	P1w	P2w	P3w	P4w	P2d
	$^3J_{H^N H^\alpha}$ coupling constants (Hz)				
Lys1	7.4	7.7	7.5	7.5	8.2
Cys2	6.9	7.4	7.4	6.8	-
Arg3	7.5	8.1	7.9	8.0	8.6
Gly4	5.9[a]	6.0[a]	6.2[a]	6.1[a]	7.4[a]
	5.7[b]	5.4[b]	5.4[b]	5.3[b]	3.0[b]
Asp5	7.6	7.7	7.6	7.5	-
Cys6	7.4	7.9	7.9	7.9	9.7
Phe7	6.9	8.0	7.6	7.7	8.6
Cya8	7.8	7.9	7.6	7.4	8.6
	$^3J_{H^N C'}$ coupling constants (Hz)				
Lys1	s	-	-	-	-
Cys2	1.2 *1.7*	1.2 *1.9*	2.8 *1.7*	1.6 *1.6*	-
Arg3	*1.2*	s	*1.0*	-	-
Gly4	*1.1*	1.3	-	-	-
Asp5	-	s	-	-	-
Cys6	2.4 *1.5*	1.8 *1.5*	1.4 *1.6*	1.8 *2.2*	-
	$^3J_{H^\alpha C'_{i-1}}$ coupling constants (Hz)				
Lys1	-	-	-	2.0	2.6
Arg3	(1.0)	-	-	-	-
Gly4	2.8[a]	2.2[a] *2.5*[a]	3.2[a]	2.0[a]	2.2[a]
	5.4[b]	5.2[b] *5.4*[b]	5.4[b]	5.2[b]	4.4[b]
Asp5	2.2	-	-	-	-
Phe7	(5.2)	-	2.8	*2.3*	-
Cya8	-	-	-	-	2.2
	$^3J_{H^N C^\beta}$ coupling constants (Hz)				
Lys1	1.8	1.7	-	1.2 *1.6*	s
Cys2	0.4	0.4 *1.03*	0.6	1.0	s
Arg3	-	1.9	*1.6*	*1.9*	s
Asp5	1.3	1.3 *1.3*	1.3 *2.3*	1.9 2.0	-
Cys6	0.6	0.4 *0.8*	0.6 *1.0*	1.0 *1.0*	0.7
Phe7	1.1	0.9	0.9	0.9	s
Cya8	1.1	0.9	-	0.2 (0.85)	0.5

[a] and [b] refer to down and high field protons

signal for $^3J_{H^\alpha C^\gamma}$ coupling constants. In these cases, coupling constant could be neither measured or given an approximate value. Similarly, degenerate H^β protons prevented $^3J_{H^\beta C'}$ and $^3J_{NH^\beta}$ from being measured.

Table 4.3: **Summary of the coupling constants related to the χ^1 dihedral angle:** The use of parenthesis have identical meaning as in table 4.2.

Residue	P1w	P2w	P3w	P4w	P2d
$^3J_{H^\alpha H^\beta}$ coupling constants (Hz)					
Lys1	5.6	5.6	-	5.7	5.4
	9.1	9.1	-	9.1	9.3
Cys2	-	-	-	-	-
	5.0	-	-	-	-
Arg3	5.7	5.6	-	5.7	7.0
	9.0	9.1	-	8.9	8.7
Asp5	6.0	6.3	6.1	6.0	5.4
	8.0	7.4	7.5	7.4	6.8
Cys6	5.1	4.9	4.4	4.2	-
	5.1	5.1	5.5	5.6	-
Phe7	7.2	-	-	-	5.2
	9.0	-	-	-	8.2
Cya8	6.6	-	-	-	6.3
	6.6	-	-	-	6.8
$^3J_{H^\beta C'}$ coupling constants (Hz)					
Lys1	2.4	-	-	-	-
	3.4	-	-	-	-
Arg3	3.0	-	-	-	-
	3.4	-	-	-	-
Asp5	3.6	3.4	3.2	3.4	-
	3.8	4.0	3.8	3.8	*3.8*
Cys6	5.6	5.6	4.6	4.8	-
	2.2	2.0	2.4	(1.2)	-
Phe7	6.5	-	-	-	2.4
	2.2	-	-	-	3.8
Cya8	3.6	-	-	-	-
	3.4	-	-	-	-
$^3J_{H^\alpha C^\gamma}$ coupling constants (Hz)					
Lys1	2.8	1.9	2.8	3.0	2.6
Arg3	-	-	-	-	(2.9)
Asp5	4.8	-	2.2	2.0	3.6
Phe7	<3	-	2.4	-	2.4
$^3J_{NH^\beta}$ coupling constants (Hz)					
Lys1	-	-3.6	(-1.4)	-	-2.3
	-	-1.7	(-0.7)	-	-1.3
Cys2	-1.7	-	-	-	-
	-3.0	-	-	-	-
Arg3	-	-3.4	-2.4	-	-9.0
	-	-1.7	-6.3	-	-1.7
Asp5	-3.0	-2.9	-3.0	-3.0	-4.6
	-2.5	-2.1	-2.1	-1.9	(-0.2)
Cys6	-1.5	-1.7	-2.1 (-4.2)	-2.1	-0.9
	-3.2	-3.0	-2.9 (-3.8)	-2.9	-3.6
Phe7	-	-	-	-	-0.9
	-	-	-	-	-3.0
Cya8	-	-	-	-	-1.7
	-	-	-	-	-2.9

4.4 From scalar coupling constants to dihedral angle restraints

NOE enhancements are the principal experimental data used in NMR structure determination. However, for small peptides, the ratio number of distance-restraints/surface is low compared to proteins. Three-bond coupling constants are therefore a valuable supplement to NOEs. They are capable of restraining the local geometry of molecules as they can be converted into dihedral angles via Karplus-type equations. Interpretation of coupling constants in terms of dihedral angles is, however, not trivial and several points must be considered. The degeneracy of the Karplus equations implies a non-unique relationship between the dihedral angles and ^3J coupling constants. The Karplus-type equations have a general form of $(f \circ g)(x)$ with $g(x) = cos(x)$ and $f(x) = Ax^2 + Bx + C$. Hence, one coupling constant could lead to two or four solutions - dihedral angles. Moreover, the measured coupling constant can be time averaged, reflecting a flexibility of peptides.

Both of these problems can be addressed by combining information contained in several coupling constants related to the same dihedral angle [100]. As the sizes of coupling constants depend on the substitutions, Karplus curves have to be parameterized individually for different couplings. Several sets of coefficients have been described/used for Karplus type equations in peptides/proteins [56, 57, 62, 101]. Although very similar, the coefficients vary slightly depending on the authors and techniques employed to parametrize the Karplus equations. We have decided to use angles those employed by Kessler's group, who have been working on RGD peptides [100] for ϕ, angle those proposed by Polshakov et al. [56] and parameters of Wasylishen et al. [102] for the analysis of $^3J_{H^\alpha C^\gamma}$ coupling constants for χ^1.

The Karplus coefficients presented in Table 4.4 were parametrized experimentally using rigid molecules and are the results of the best fitting for an ensemble of data or were derived from *ab initio* calculations [57, 101].

A case-by-case interpretation of the measured coupling constants was undertaken. The reason for this prudence is that the slopes of the Karplus curves vary and when the slope is close to zero a given value of a coupling constant corresponds to a range of

Table 4.4: Coefficients A, B and C of Karplus-type equations and the phase shift (phase) that relates the dihedral angle to certain coupling constants.

Coupling constant and substituents	A	B	C	phase
Related to the ϕ dihedral angles				
$^3J_{H^NC^\beta}$	4.7	-1.5	-0.2	60
$^3J_{H^NC'}$	5.7	-2.7	0.1	-180
$^3J_{H^\alpha C'_{i-1}}$	9.0	-4.4	-0.8	120
$^3J_{H^NH^\alpha}$	6.98	-1.38	1.72	-60
Related to the ψ dihedral angles				
$^3J_{H_\alpha N_{i+1}}$	-0.88	-0.61	-0.27	-120
Related to the χ^1 dihedral angles				
$^3J_{H^\alpha H_2^\beta}$	9.5	-1.6	1.8	-120
$^3J_{H^\alpha H_3^\beta}$	9.5	-1.6	1.8	0
$^3J_{H^\alpha C^\gamma}$	7.1	-1	0.7	120
$^3J_{NH_2^\beta}$	-4.4	1.2	0.1	120
$^3J_{NH_3^\beta}$	-4.4	1.2	0.1	-120
$^3J_{H_2^\beta C'}$	7.2	-2.04	0.6	0
$^3J_{H_3^\beta C'}$	7.2	-2.04	0.6	120

dihedral angles. Furthermore, one has to take into account the uncertainty in the parameterization of Karplus curves that can also lead to increased error bounds that have to be set when interpreting the values of coupling constants.

When all of the coupling constants were converging onto one dihedral angle, this angle was restrained according to the value returned for the $^3J_{H^NH^\alpha}$ coupling constant. This coupling constant was chosen because of the high accuracy of its measurement and reliable parametrization of the Karplus curve for proton-proton coupling constants. Error bounds of ± 30 degrees were set for dihedral angles converging towards a unique angle. This error is large enough to compensate for possible inaccuracies of the Karplus-type equations and some limited flexibility that is supported by unrestrained molecular dynamics. In cases where the lack of data led to two solutions (usually between -180 degrees and 0 degrees) the error bounds were set to ± 90 degrees. Finally, if coupling constants were not converging to a particular dihedral angles, the angle was not restrained. The different cases described above are illustrated in Fig. 4.13, using an example of ϕ angles for selected residues of peptide 2 in DMSO. The dihedral angles were calculated from Karplus equation-types using a home written program for the

analysis of Karplus equations. The glycine was not included because of the lack of parameters for Karplus equation.

Starting with residue 8 of peptide 2 dihedral angles were calculated for all coupling constants [100, 103]. The inspection of dihedral angles allowed inconsistent values to be excluded as illustrated in Table 4.5.

Table 4.5: **Determination of dihedral angles using the exclusion method.** The coupling constants of residue 8 of peptide 2 in DMSO. Bold values correspond to the dihedral angles remaining after the exclusion of the inconsistent values.

Coupling constants	Value (Hz)	Angles (degrees)			
$^3J_{H^NH^\alpha}$	8.56	**-146.36**	-93.64		
$^3J_{H^\alpha C'_{i-1}}$	2.22	-90.83	**-149.17**	-7.40	127.40
$^3J_{H^NC^\beta}$	0.55	-6.14	-113.86	45.70	**-165.70**

Interpretation of coupling constants for this residue 8 also illustrates that the quantitative approach should be supplemented by a qualitative inspection of Karplus curves. For this residue three coupling constants were extracted; this should be enough to restrain the dihedral angle, unless signs of flexibility are uncovered during the analysis of these coupling constants. The values of $^3J_{H^NH^\alpha}$ and $^3J_{H^\alpha C'_{i-1}}$ indicated solutions for angle at -146 or -94 degrees. However, the value of $^3J_{H^NC^\beta}$ was slightly out of this range (the solutions were -166 and -114 degrees). Two assumptions guided the choice of the restraint. (1) The coupling constants $^3J_{H^NH^\alpha}$ indicates a relatively rigid conformation as the value is above 8 Hz, which cannot be an average of smaller values. (2) Visually it can be seen that the dihedral angle of -166 as implied by the $^3J_{H^NC^\beta}$ coupling constant, is closer to the -146 corresponding to the $^3J_{H^NH^\alpha}$ of 8.6 Hz.

In the absence of data for residue Asp5 the ϕ angle could not be restrained. Except for residues 2 and 5, the $^3J_{H^NH^\alpha}$ coupling constants were always larger than 8 Hz, consequently possible solutions were restricted to the 0 to -180 degrees region. Only one coupling constant was obtained for residues Phe7 and Arg3; consequently their dihedral angles were weakly restrained. The dihedral restraint[2] was set -120 $\pm 40^o$ and

[2] Refers to the statement CDIH in CNS which constrains the dihedral angle to a value and associates an error. This is expressed in degrees

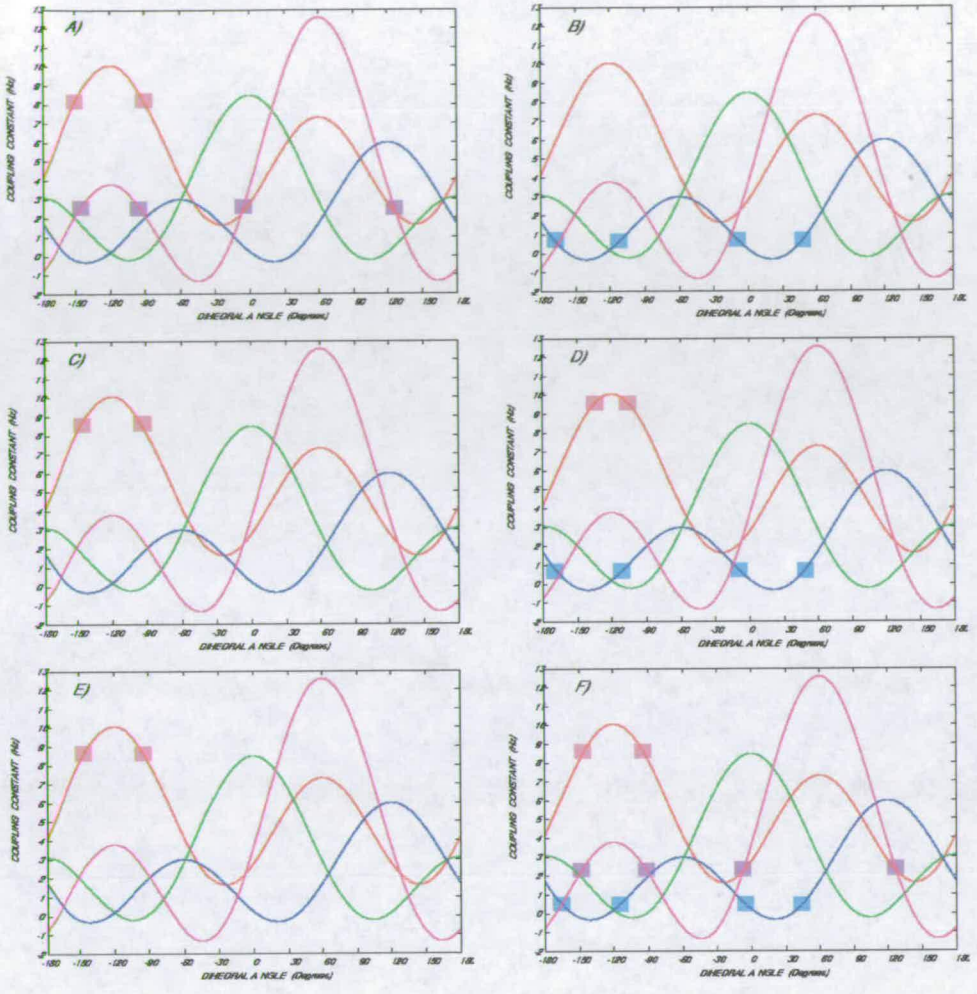


Figure 4.13: **Interpretation of coupling constants related to the ϕ dihedral angle. An example of peptide 2 in DMSO:** Figures A to G refer to residues 1, 2, 3, 6, 7 and 8, respectively. The Karplus curves obtained using coefficients given in Table 4.4 for $^3J_{\mathrm{H^N H^\alpha}}$ (red), $^3J_{\mathrm{H^\alpha C'_{i-1}}}$ (violet), $^3J_{\mathrm{H^N C^\beta}}$ (blue) and $^3J_{\mathrm{H^N C'}}$ (green). Possible solutions are for the measured coupling constants (see Table 4.2) are indicated by squares.

no coupling restraint[3] was used. Coupling constants of Lys1 ($^3J_{\mathrm{H^N H^\alpha}}$ and $^3J_{\mathrm{H^\alpha C'_{i-1}}}$) showed that two solutions are compatible with both coupling constants (-150 and -90) and the dihedral was set as -120 $\pm 60^o$, the coupling restraints were not set. The value of $^3J_{\mathrm{H^N C^\beta}}$ coupling constant (less than 1 Hz) of Cys2 is compatible with four values of the ϕ angle. In the absence of any other coupling constants, the $^3J_{\mathrm{H^N C^\beta}}$ coupling constant could not be interpreted and the ϕ angle was left unrestrained for this residue.

[3] Refers to COUP statement in CNS and is a strict interpretation of the coupling constants using Karplus-type equations. It is expressed in Hz.

Qualitative interpretation of coupling constants can also provide valuable information about the local geometry. The lack of long-range cross peaks in certain experiments can be interpreted as a sign of very small values of coupling constants. This information can also be interpreted in terms of dihedral angles [100].

In general, the values of the dihedral restraints were based on the value of the $^3J_{H^N H^\alpha}$ coupling constant which shows a high amplitude and could be measured for most of the residues. When all the coupling constants converged to one dihedral angle, the error was set to $\pm 30^o$. For coupling constants converging to similar angles but with two solutions, the error was set to $\pm 60^o$. In the case where no convergence could be found but certain angles could be excluded, the error was $\pm 90^o$. In the latter case, these exclusions were based on the examination of two ambiguous solutions arising within the same region of the curves, e.g. the coupling constants of Phe7 of peptide 2 indicate two solutions for ϕ angles at -150 and -90 degrees. This ambiguity was reduced to -90 ± 90^o because none of the solutions converged to the other allowed ϕ angle (according to Ramachandran plots and steric effects) which is near $+60^o$. High values of the $^3J_{H^N H^\alpha}$ coupling constants (above 8.5 Hz) unambiguously indicated that the angles were located around -120^o for which an error of $\pm 40^o$ was given. This was reduced to $\pm 20^o$ for the particular case of Cys6 in DMSO where $^3J_{H^N H^\alpha}$ (9.7 Hz) and $^3J_{H^N C^\beta}$ converged toward a unique solution. Finally, if the analysis did not yield consistent dihedral angles, the dihedral angle was not restrained.

The set of restraints used in the structure calculation of peptides 1-4 is summarized in the Table 4.6. It can be seen that the error bounds vary and in general are on the large side. Such weak restraining of dihedral angles takes into account possible inaccuracies in the measured coupling constants arising primarily from limited signal-to-noise ratio of spectra. This is also taking into account limited flexibility and inaccurate parameterization of Karplus curves.

Using the Newman projections, the three staggered conformations (Fig. 2.5) can be analyzed and the conformation of the χ^1 dihedral angles can be established by comparing the coupling constants related to the this angle (Table 4.3) and the distances $d(_{N\beta})$ and $d(_{\alpha\beta})$ [56, 104–106]. This analysis also allows the prochiral assignment of the H^β protons of the side chain. The analysis of coupling constants related to the χ^1 failed to

Table 4.6: Dihedral restraints deduced from the coupling constants.

Residue	1 (H_2O)	2 (H_2O)	3 (H_2O)	4 (H_2O)	2 (DMSO)
Dihedral angle ϕ (degrees)					
1 Lys	-84±30	-87±30	-	-84±30	-120±60
2 Cys	-160±30	-150±30	-156±30	-120±60	-
3 Arg	-120±60	-89±30	-86±30	-92±30	-120±40
4 Gly[a]	90±90	90±90	90±90	90±90	90±90
5 Asp	-86±30	-90±90	-86±30	-85±30	-
6 Cys	-156±30	-154±30	-152±30	-150±30	-111±20
7 Phe	-	-90±90	-90±90	-120±60	-120±40
8 Cya	-87±30	-90±90	-155±30	-156±30	-150±20
Dihedral angle ψ (degrees)					
4 GLY[a]	-90±90	-90±90	-90±90	-90±90	-90±90

[a] according the glycine filter discussed page 112

converge to any of the staggered rotamers (-60, 60 and 180) in all peptides indicating a conformation averaging of side chains. The measured values of coupling constants correspond to angles of 0, 120 or -120 degrees which are the average angles observed on the NMR time scale. These energetically unfavorable dihedral angles clearly indicate flexibility of side chains.

The stereospecific assignments can be achieved by using a statistical approach imbedded in the Pacshler's equations. This approach has been extensively used to perform stereospecific assignments of the methylene prochiral protons [106–111]. The population of each rotamer can be assessed by combining the coupling constants ${}^3J_{H^\alpha H^\beta}$ and ${}^3J_{H^\beta C'}$, which show large variations as a function of χ^1. This model assumes relative populations PI, PII and PIII of the three staggered conformers with χ^1=-60, 180 and 60^o yielding an average observed coupling constants which is a weighted average of the three corresponding coupling constants [112].

$$J_{obs} = \sum_i P_i J_i \tag{4.2}$$

Three equations can be written for ${}^3J_{\mathrm{H}^\beta\mathrm{H}^\alpha}$ coupling constants:

$$\begin{aligned} PI &= \frac{J_{H^\alpha\beta_{pro-R}} - Jsc}{Jap - Jsc} \\ PII &= \frac{J_{H^\alpha\beta_{pro-S}} - Jsc}{Jap - Jsc} \\ PIII &= 1 - PI - PII \end{aligned} \tag{4.3}$$

and for $^3J_{H^\beta C'}$ coupling constants:

$$PII = \frac{J_{H\beta pro-RC'} - J_{sc'}}{J_{ap'} - Jsc'}$$

$$PIII = \frac{J_{H\beta pro-SC'} - J_{sc'}}{J_{ap'} - Jsc'} \tag{4.4}$$

$$PI = 1 - PII - PII$$

The coupling constants for the antiperiplanar (ap) and synclinal (sc) conformations are: J_{ap}=13.6 Hz, J_{sc}=2.6 Hz, $J_{ap'}$=8.5 Hz and $J_{sc'}$=1.4 Hz [106] where apostrophes indicate the heteronuclear coupling constants. Two sets of calculations were performed assigning the down field H^α alternatively as a Pro-R or a Pro-S proton for both $^3J_{H^\alpha H^\beta}$ and $^3J_{H^\beta CO}$ coupling constants. The combination which yielded the best match between the Pi values obtained with both types of coupling constant represented most likely the correct stereospecific assignment. Because of the degeneracy or strong coupling of most of the H^β protons, this analysis could only be carried out for Asp5, Cys6 and Phe7 at most. Using these values, the stereospecific assignments of the H^β protons was carried out and is reported in Table 4.7. The stereospecific assignment of the protons of residues Cys6 and Phe7 agreed with the assignment of these protons returned by CNS in structure calculations without the specific assignments.

Calculated populations of rotamers show that no single side chain conformation is preferred. Based on the values of the coupling constants and established averaging of the side chains, the conformation of the χ^1 angles were not restrained during the calculations.

4.5 Temperature coefficients of amide protons

Measurements of temperature coefficients are performed routinely on peptides in order to assess the presence of intra-molecular hydrogen bonds. This is an attractive method for small peptides because of their thermal stability and small number of amide protons which allows a large range of temperature range to be covered, while acquiring simple 1D spectra.

The chemical shifts of the backbone amide protons were measured over a range of 40°C from 275 to 320 K. A series of 1D TOCSY, with selective excitation of the aliphatic

Table 4.7: **Rotamer populations of $\chi 1$ angles and stereospecific assignments:** "d" indicates degenerate β protons. Only the population based on the correct stereospecific assignment are presented. For each peptide, the left and right columns indicate rotamer populations based on homo- and heteronuclear couplings, respectively.

Peptide	1 (H_2O)	2 (H_2O)	3 (H_2O)	4 (H_2O)	2 (DMSO)
			Cys6		
PI	23 ; 29	23 ; 31	26 ; 40	28 ; 54	
PII	23 ; 11	21 ; 9	17 ; 14	15 ; -3	
PIII	54 ; 60	56 ; 60	57 ; 46	58 ; 48	
$H^{\beta 2}$ →Pro-R ; $H^{\beta 1}$ →Pro-S					
			Phe7		
PI	42 ; 17	d	d		57 ; 51
PII	58 ; 71	d	d	d	29 ; 24
PIII	0 ; 12	d	d	d	14 ; 25
water: $H^{\beta 1}$ →Pro-R ; $H^{\beta 2}$ →Pro-S					
DMSO: $H^{\beta 2}$ →Pro-R ; $H^{\beta 1}$ →Pro-S					
			Asp5		
PI			33 for all		
PII			33 for all		
PIII			33 for all		
No prochiral assignment possible					

protons, was carried out to remove the overlap in the amide region. The duration of the selective pulse and the mixing time were optimized individually. A typical mixing time of 60 ms was used for excitation of H^{β} protons while 40 ms was used for for H^{α} protons. The spectra were referenced using the water signal according to the pH, salt concentration and the actual temperature of the sample using the parameters described in 3.8. The number of scans varied depending on the efficiency of the TOCSY transfer. At low temperature, the number of scans was increased to 2 K, 128 scans were sufficient for higher temperatures. Each residue had at least one isolated peak that could be selectively excited and from which magnetization of its amide proton could be built.

Figure 4.14 shows an example of the determination of temperature coefficients using peptide 3 in H_2O.

The temperature coefficients were determined for peptides 1 to 4 in water and for peptide 2 dissolved in DMSO. Only the amide protons of the backbone were considered (residues 1 to 8). The chemical shifts of the amide protons varied linearly with the

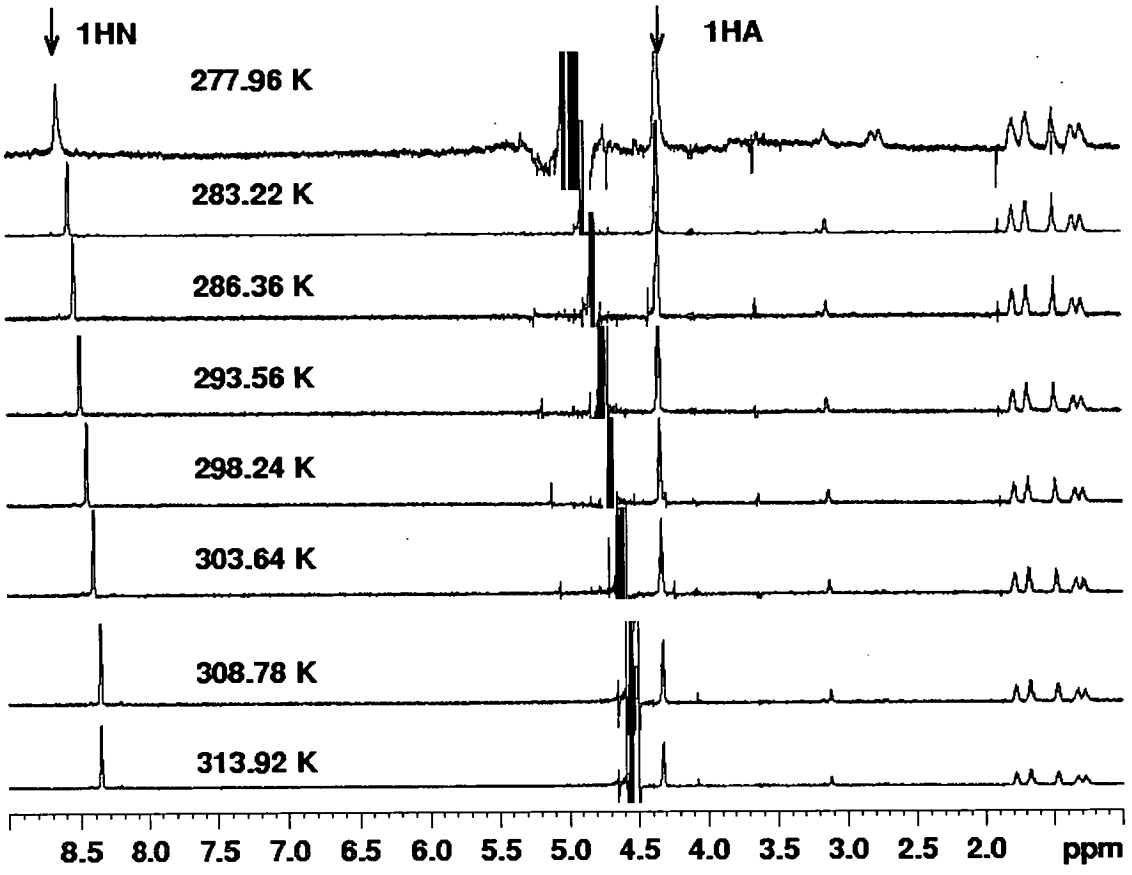


Figure 4.14: Series of 1D TOCSY spectra of Lys 1 in peptide 3 where the H^{α} proton is selectively excited and the magnetization is transferred to the aliphatic protons and H^{N} proton.

temperature (Appendix G).

The temperature coefficients were calculated from the slope of a linear regression of proton chemical shifts vs. temperature and are summarized in Table 4.8.

Overall the temperature coefficients were very similar for all peptides in H_2O. Large absolute values of the temperature coefficient of residue 8 in peptide 1 could be explained by the presence of the PEG tail attached at the end of this residue in peptides 2-4. This has protected the amide proton from the solvent and slowed down the rate of exchange in other peptides. Temperature coefficients for residues 1-8 in all peptides were in the range of -5 to -9 ppb/K with an exception of the already mentioned H^{N} of residue 8 in compound 1 and H^{N} of Cys6 in all compounds. The latter is indicative of a hydrogen bond formed with this H^{N} proton.

Table 4.8: **Temperature coefficients (ppb/K) for peptides 1-4:** In water, at least five temperatures where used to calculate the coefficients - from 275 to 320 K. In DMSO, the range 293 to 333 K was used.

Residue	P1w	P2w	P3w	P4w	P2d
Lys1	-9.2	-7.8	-7.8	-8.1	-5.4
Cys2	-5.2	-5.9	-5.0	-5.2	-0.1
Arg3	-7.5	-7.6	-7.6	-7.7	-5.3
Gly4	-9.3	-8.6	-8.3	-8.2	-6.0
Aps5	-8.8	-7.8	-8.0	-8.2	-7.6
Cys6	-2.7	-3.5	-3.1	-3.5	1.3
Phe7	-5.1	-7.1	-6.3	-6.4	-3.4
Cya8	-12.5	-8.6	-8.9	-9.0	-7.1

These observations contrast somewhat with the behavior of amide protons in DMSO. Generally, the temperature dependence of H^N chemical shifts is lower than these observed in water with an exception of Asp5 H^N which is comparable to the coefficient calculated in water for P2w.

The temperature coefficients are disscussed further in chapter 5.

Chapter 5

Structure calculations

The structures of the cyclic peptides were calculated by means of restrained molecular dynamics where the restraints were derived from NMR data. This chapter describes the design of templates for the unusual parts of the molecule and the implementation of the NMR experimental restraints for the structure calculations.

5.1 Parametrization of the unusual peptides for CNS and XPLOR

Structures calculations are based on molecular mechanics and empirical parameters defined for all parts of the molecule. These empirical parameters are used to calculate energies (Eqn. 2.31). Any residue undefined in the force field must therefore be designed first. In general, molecules are described by their topology (connectivity, nomenclature), properties of the atoms (charge and volume) and covalent geometry (bonds, angles, torsion angles). For a particular molecule, these parameters are gathered in a structure file, connecting the force field and the molecular topology.

The geometry of the unusual parts of the cyclic peptides were created from existing residues or designed *ex nihilo* using molecular the modeling program HYPERCHEM [113]. The first step consisted of creating the topology for each residue followed by adding covalent geometries, as explained in the following parts. New residues were kept short to ease the assignment of NMR resonances and to create structure files that resembles those of proteins (i.e. residues separated by amide bonds).

5.1.1 Parameters for the modified cysteine (Cya)

Cya residue was created starting from CYsteine by adding an Acetyl (Cya) moiety to the sulfur atom (Fig. 5.1.1). The side chain of this modified cysteine was attached to the N terminus of the lysine and formed the second bridge in the structures of these cyclic peptides. This peptide-like construct was advantageous because it allowed Cya to be connected to other residues via peptide-like bonds. The force field parameters for the new residue Cya were based on the existing cysteine parameters. The parameters for the new parts of the modified cysteine could be copied from existing parameters of other amino acids (methionine and aspartic acid) meaning that no new atom types had to be created for this residue.

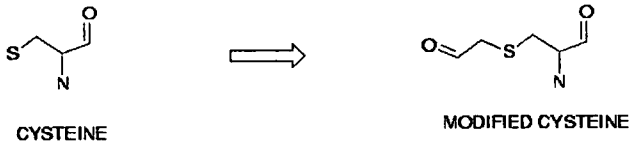


Figure 5.1: Design of the modified cysteine residue, Cya.

The atom types and atom names of the residue are presented in appendices C and D. The nomenclature for the side chain follows the Greek alphabet ($\alpha, \beta, \gamma, \delta, \epsilon$) in line with the convention used for amino acids.

5.1.2 Parameters for the backbone to side chain link (CLL)

The new Cya residue had to be bonded to the N-terminal Lys residue. This side-chain/ backbone link is not defined in the standard setup and had to be created (see appendix E). The new Cya Lys Link (CLL) was designed as a peptide bond. The parameters and the geometry of this unusual peptide bond were copied from the standard peptide bond defined in the force field. This link was implemented in the file which generates the structure file[1] as a "patch", in the same manner as the disulfide bonds or the peptide bonds are generated.

[1] psf file for XPLOR and mtf file for CNS

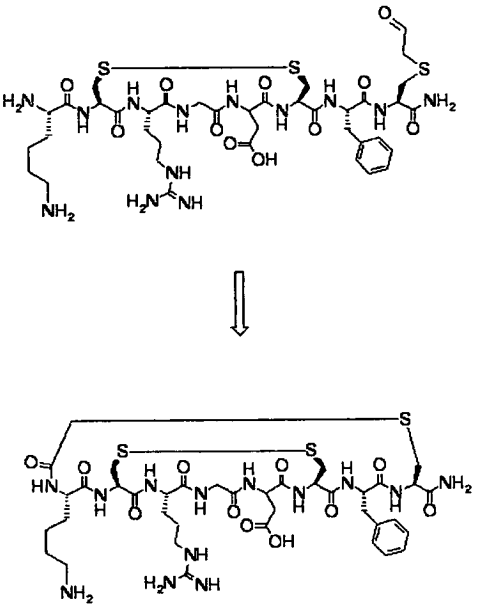


Figure 5.2: **Formation of the second cycle via CLL link:** The atoms and bonds involved are represented in red.

5.1.3 Parameters for the polyethylene glycol tail (Ncc)

The parameters of the central part of the Ncc residue (named from the first three atoms of the N-terminus, NCC) were derived from the parameters of the tetra ethylene glycol. These parameters were obtained from "Hic-up" database [87]. The terminal parts of the residue could be found in the force field and the N and C-terminus and did not require the creation of new atom types. Only one atom type, at position α, was created (atom type was called "CH2N") to avoid any conflict with the carbon alpha of the glycine (called "CH2G" in the parameter file). Similarly to Cya, the Greek alphabet was used to to name the atoms (figure 5.3).

This residue has N and C termini as any other amino acid. The creation of repeated Ncc residues in peptides 3 and 4 was naturally accomplished by a peptide bond.

5.1.4 Amide in C-terminal

In all the peptides either Cya or Ncc residues ended with an amide group. This group is defined in the force field but also had to be added to the unusual residues. Describing these residues as amino-acids, the implementation of terminal NH_2 groups

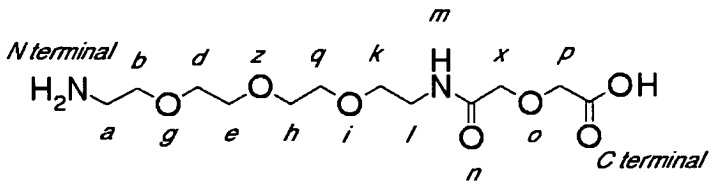


Figure 5.3: Structure and nomenclature of residue Ncc.

required the addition of the name of these unusual residues to the list of amino acids. The modifications are shown in appendix E where only the peptides relevant to the structures are reported.

5.1.5 Parameters of a modified lysine residue (Lym)

It was decided to include the metal bind site (MBS) within a modified lysine for two reasons. Firstly, in order to compare directly the different peptides, it was desirable to keep the same numbering, i.e. from residue number 1 for the lysine to residue number 8 for the modified cysteine. Second, it became evident from the analysis of NMR data, that most of the resonances belonging to the MBS were degenerate reflecting the symmetry of this moiety. Thus, resonances could not be individually assigned and the three-dimensional structure of the MBS could not easily be solved by NMR. Indeed, the situation is very different from the degenerate methylene protons of amino acids where a pseudo-atom correction can be added to account for the ambiguity of the assignment. In MBS case, protons of both branches are degenerate which makes geometrical interpretation impossible from either NOEs or coupling constants. Based on these observations, the modified lysine (Lym) was sub-divided into three parts creating a sequence Lys-link-MBS. The parameters for the lysine are already defined and no further atom types were needed for the "link". The use of Greek letter notation was not practical beyond the "link" (Fig. 5.4) and individual atoms of MBS were annotated with numbers.

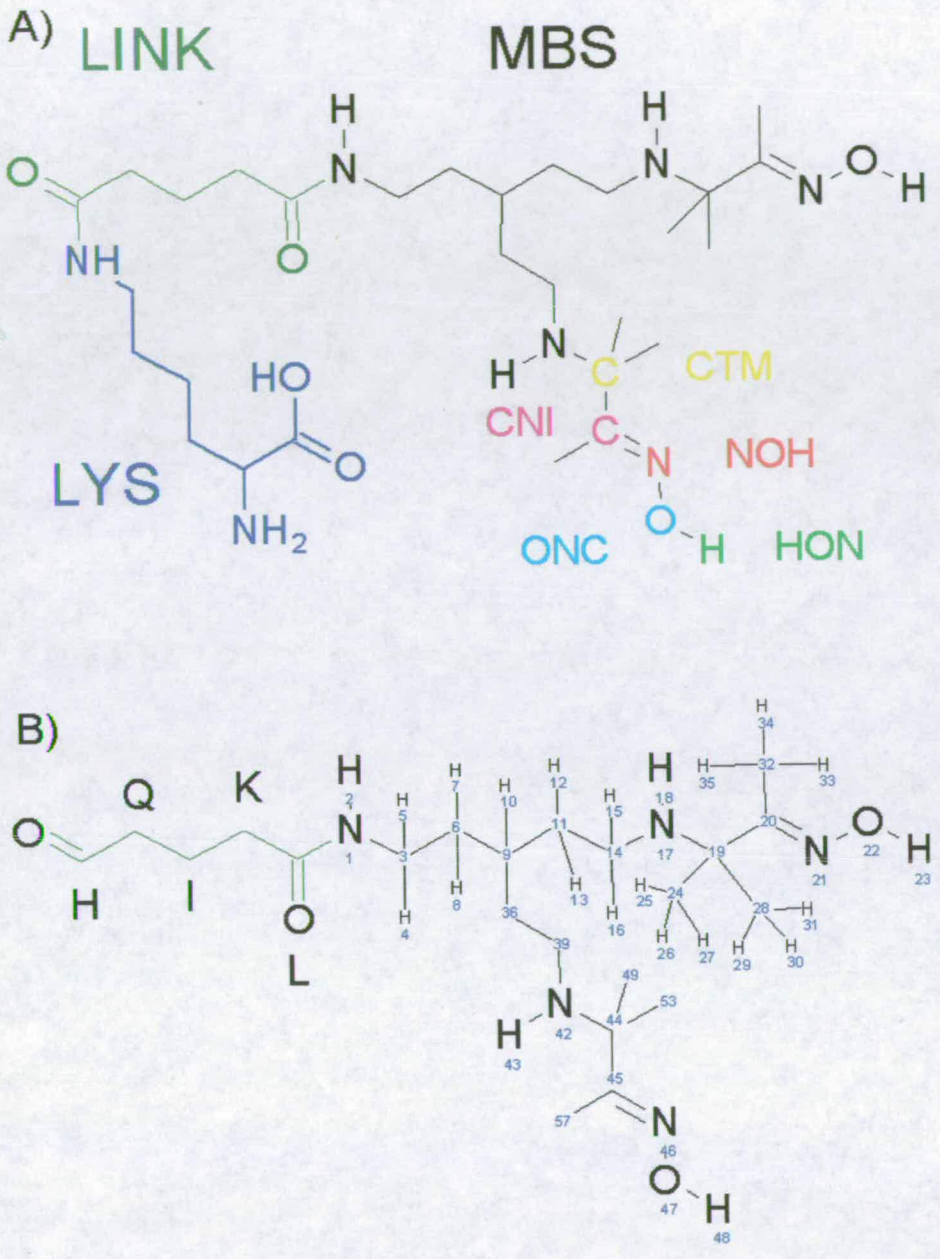


Figure 5.4: **The structure of a new residue Lym:** (A) Different parts of the new residue Lym; common part with lysine in blue, link in green and MBS in black. The atoms newly defined in the force field are indicated in one of the the branches of the MBS. (B) Numbering of the link and MBS parts of Lym. One branch of MBS is fully numbered and the other shows only the numbering of the heteroatoms and labile hydrogens.

The topology of the Lym residue was created and added to the overall topology file. At this point, it was necessary to describe bonds, angles, improper angles for the newly created atom types. Overall, five new atom types were defined for Lym. They were

named CTM, HON, NOH, ONC and CNI where the first letter corresponds to the element described. The non-bonded interactions of these atoms were based on similar atoms of the original force field. To start with a good geometry, residue Lym was drawn in the molecular modeling program HYPERCHEM. The conformational space was searched via molecular dynamics in vacuum using the OPLS force field, followed by energy minimization in order to optimize the geometry of Lym (Table 5.1). The molecular dynamics starts at high temperature to allow large conformational changes and ended at 300 K, the temperature at which the parameters are defined in the force field.

Table 5.1: Parameters of the optimization of Lym residue in HYPERCHEM.

Conformational search	**Parameters**
Molecular Dynamics	Initial temperature = 1000 K ; Final temperature = 300 K ; Temperature step 10 K Step size = 1 fs ; Total time of the run = 10 ps ; Cooling time = 1 ps
Minimization	Fletcher-Reeves conjugate gradient algorithm with a convergence to 0.1 kcal.Å^{-1}

Equilibrium parameters were extracted from the minimized structures using "learn.inp" in CNS (Appendix B) from Cartesian coordinates and appended to the parameter file. The new atom-type based parameters appended to the original parameters required checking for any potential errors. Also the bond length, torsion angles or the angles generated by HYPERCHEM can differ slightly from those defined in CNS. Any potential errors could be corrected by calculating the structures of cyclic peptides using the empirical parameters only. This analysis was carried out using peptide 4 which contains all the unusual residues. Template structures were calculated after each modification. These modifications concern the angles and impropers that contribute most to the total energy. First, the energy and the geometry of the template structure were examined to check the correctness of the implemented topologies. This was particularly true for the improper angles whose parameters were not intuitive. Subsequently, the thresholds of the violations for the empirical parameters were lowered in order to examine the major contributors to the total energy and the parameters were iteratively modified to optimize the structure for CNS. The violations of the equilibrium

parameters, small but numerous, occurred mainly for the angular parameters of the unusual residues. This observation was interpreted as a mismatch between the values used for the original force field and those set for the unusual parts of the molecule. The modifications dramatically reduced the energy of the structure. Table 5.2 shows the result of the optimization of empirical parameters for the unusual residues of peptide 4. After iterative adjustments of parameters, the angular energy decreased dramatically while the other energies were kept low.

Table 5.2: **Optimization of the empirical parameters of peptide 4:** The table shows the result of the optimization of the empirical parameters after iterative adjustments.

Energy (kcal.mol^{-1})	Total	Bonds	Angles	Impropers	VDW
Initial parameter set	75.183	0.586	69.141	3.005	2.450
Final parameter set	35.392	0.735	25.608	1.654	7.395

The energies of the empirical parameters were compared between the peptides. Being the smallest compound and having most of its structure defined by usual amino acids, compound 1 was expected to present the lowest total energy. The increase of energy occurring in the other compounds can be attributed to the unusual parts of the peptides.

Table 5.3: Energies of the template structures of the cyclic peptides 1-4.

Energy (kcal.mol^{-1})	Total	Bonds	Angles	Impropers	VDW
Peptide 1	3.100	0.058	2.532	0.164	0.346
Peptide 2	8.164	0.136	6.653	0.732	0.643
Peptide 3	19.400	0.324	14.735	1.225	3.115
Peptide 4	35.392	0.735	25.608	1.654	7.395

5.2 NOE restraints

The intensity of the cross peaks in NOESY/ROESY spectra is related to the interatomic distance. However, the distances derived from the NOE intensities are imprecise since the relationship between the two variables is distorted by the use of the

two-proton approximation, internal dynamics and spin diffusion. The process of calibration of NOE intensities into distance restraints normally starts with calibration of cross peaks volumes. The most convenient way of calibrating distances is to use the CH_2 proton as the distance between the two H^{β} protons or the H^{α} protons of the glycine is known (1.78 Å). Other possibilities include distances from methyl to methyl (valine, leucine) or aromatic protons in tyrosine [61]. Using CH_2 protons suffers from spin diffusion artifacts and tends to miscalibrate the long distances which can severely affect the outcome of the calculations. Moreover, the intensities of cross peaks considered for the calibration would have to be taken at short mixing times, when the spin diffusion is negligible. At these mixing times the zero-quantum effects are significant and for protons with small a chemical shift separation, the volume of the cross peaks cannot be completely measured, due to the dispersive pattern arising from zero-quantum effects, even by using the latest z-filters [65]. Preliminary structure calculations using cross peak volumes calibrated on methylene protons yielded large violations of distances for both NOESY and ROESY data. Hence, further calculations were not performed using calibrated cross peaks. Instead, the cross peak volumes were converted into distances more loosely and arranged into the usual four classes corresponding to strong, medium, weak and very weak intensities [114].

The intensities extracted from two dimensional NOESY or ROESY experiments were converted into distances using the program "ansig2rdb", written by A. Raine, which generates normalized intensities as:

$$N_i = \frac{I_i}{\sum_{i=1}^{n} I} \tag{5.1}$$

where N_i is the normalized intensity and I the intensity.

The distance restraints were generated by the program "Connect" from the package AZARA. This program converts the normalized intensities into inter-proton distances with lower and upper limits. The choice of these limits had a notable effect on the calculated structures and several iterations were performed in order to adjust their values. The criteria during this process were the number of distance restraints violations and the NOE energy relative to the total energy. A 250 ms mixing time 2D NOESY spectrum of peptide 3 was chosen to set up the distance limits. A set of structures was initially calculated using the standard ranges of inter-proton distances as usually used

in protein structure calculation (Table 5.4).

Table 5.4: Conversion of normalized cross peak intensities into interatomic distances.

Original values		**Modified values**	
Normalized intensity	Interatomic distance (Å)	Normalized intensity	Interatomic distance (Å)
0 - 0.3	0 - 6	0 - 0.2	2.7 - 5.2
0.3 - 1.5	0 - 5	0.2 - 1.4	2.5 - 4.5
1.5 - 2.9	0 - 3.3	1.4 - 2.9	1.8 - 3.3
more than 2.9	0 - 2.7	more than 2.9	1.8 - 2.7

The absence of lower limits and setting the upper limit to 6 Å did not restrain the structures at all. Although no violations were observed, the NOE energy was close to zero and the rmsds of the structures were high. Several modifications of the original Connect script were tested. The main premise was that, operating close to the crossover enhancement on the NOE curve, the maximum distance of 6.0 Å was too long and this was subsequently lowered to 5.2 Å. This was accompanied by an adjustment of the distance classes and normalized intensities.

The modified distance classes are presented in table 5.4. This set was used for the structure determination of all peptides in water using a single mixing time. The classes in DMSO were left to the original setting as setting the upper distance to 5.2 Å using the modifed class in Table 5.4 led to many violations. This reflects the higher intensity of the NOEs in DMSO. Concerning the full relaxation matrix, intensities from the NOESY spectra were directly used as a restraint. In this case, the distances are based on the build-up curve and the motional model during structure calculations. Hence no distance classes have to be generated.

Another modification included the addition of 0.88 Å to all distances that involved degenerated CH_2 protons. These modifications dramatically improved the quality of the structures without bringing further violations.

Structures calculated using the parameters outlined above were partially unrefined or adopted several conformations with similar energies. Further improvements were brought about to the structures by filtering the generated conformations rather than

restraining them further.

The glycine filter. First calculated structures yielded glycine residues adopting several conformations. The back-calculated distances showed discrepancies between the experimental data and the structures. In particular, the intensity of ROESY or NOESY cross peaks, between $4H^N$-$4H^\alpha$ and $5H^N$-4^α proton pairs were not reproduced. In order to produce only structures consistent with NMR data, a loop was implemented in CNS to reject any structure that would not satisfy the pattern of cross peak intensities shown in Figure 5.5. The strategy used to filter the structures did not require any stereospecific assignments..

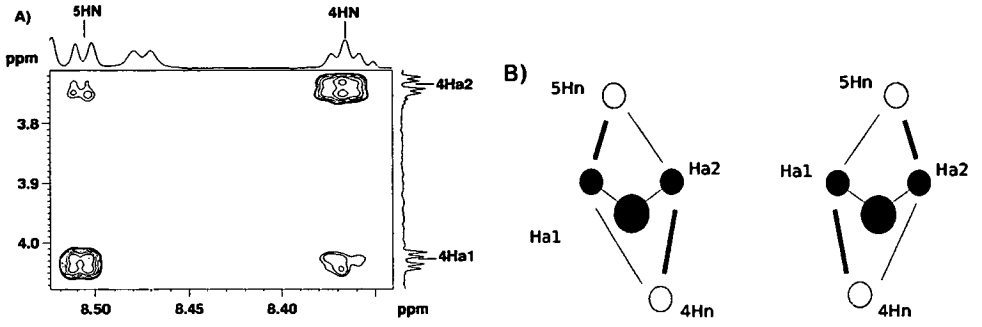


Figure 5.5: **Analysis of Gly4-Asp5 conformation:** (A) Expansion of a 200 ms 800 MHz NOESY spectrum of peptide 2 showing $4H^N$-$4H^\alpha$ and $5H^N$-$4H^\alpha$ cross peaks. (B) Schematic representation of the two possible arrangements that satisfy the experimental restraints.

At the heart of the glycine filter, is a comparison of the cross peak volumes as this can be interpreted in term of distances. According to the cross peak pattern observed in Figure 5.5, the relationship between their volumes is:

$$\begin{aligned} Vol_{Gly4HN-Gly4\alpha2} &> Vol_{Gly4HN-Gly4\alpha1} \\ &\text{and} \\ Vol_{Asp5HN-Gly4\alpha2} &< Vol_{Asp5HN-Gly4\alpha1} \end{aligned} \tag{5.2}$$

Which when translated into distances gives;

$$\begin{aligned} Dist_{(Gly4N-Gly4\alpha2)} &< Dist_{(Gly4N-Gly4\alpha1)} \\ &\text{and} \\ Dist_{(Asp5N-Gly4\alpha2)} &> Dist_{(Asp5N-Gly4\alpha1)} \end{aligned} \tag{5.3}$$

In essence, if $5H^N$ is closer to $4H^\alpha_1$ than $4H^\alpha_2$, $4H^N$ must be closer to $4H^\alpha_2$ than $4H^\alpha_1$. Because the stereospecific assignment is unknown, two solutions are possible. A loop

was implemented in CNS by defining the parameter Δ as follows:

$$(Dist_{(G4N-G4\alpha 2)} - Dist_{(G4N-Gly4\alpha 1)})(Dist_{(D5N-G4\alpha 2)} - Dist_{(D5N-G4\alpha 1)}) = \Delta \quad (5.4)$$

Only negative values of Δ satisfy the experimental NOESY intensities. The value of Δ was evaluated in regularized structures and the calculations were repeated when positive values of Δ were obtained. The filtering process is simple and effective but increases the computational time 5 to 10 times. All filtered structures can only have ϕ and ψ angles with opposite signs for the glycine, unlike the unfiltered structure which yielded all combinations. Combined use of the distance restraints and the glycine filter yielded positive values for ϕ_4 and negative for ψ_4 in all peptides.

Based on this result, a similar outcome was achieved by setting the dihedral restraints for Gly4 with ϕ and ψ angles of 90 ± 90^o and -90 ± 90^o respectively. This reduced the computational time considerably.

Anti-Distance Restraints. Anti-distance restraints (ADR) were also added to the restraint files. These correspond to non-observable cross peaks in the spectra and were implemented only when the lack of a cross peak could be positively identified, i.e. using the dispersed regions of the spectra such as the fingerprint region. Anti-restraints were implemented when it was noticed that some protons in some structures which were close in space did not show cross peaks in 2D NOESY or 2D ROESY spectra. For such instances the inter-proton distance were restrained to fall into the range of 4.5 Å to 15 Å. Assuming that the longest observable distance in the calculations was set to 5.2 Å, the lower limit of 4.5 Å was considered to be a safe choice.

For peptides 1 and 2, the NOEs were weak and 2D ROESY were acquired to overcome the problem of weak enhancement. A ROESY spectrum of peptide 3 was acquired for a comparison. When the ROESY spectra were used for the structure calculations using the modified distances classes (Table 5.4) the structures showed many violations. The cross peaks originated from the ROESY spectra and non-observed in the 2D NOESY, were therefore added to the distance restraints if the cross peaks belonged to two different spin-systems. Their intensity being small, they were set to a distance class of 0-6 Å.

Modification of the original protocol described here and the extra distance restraints, i.e. the anti restraints and additional ROESY restraints, produced high quality structures.

5.3 Structure calculations

The first round of calculations made use of single sets of distance restraints and the relationship between the normalized cross peaks intensities and the inter-atomic distances given in Table 5.4. The energy profiles, both the total and the NOE energies, were rather flat most of the time, without significant jumps. Arbitrarily, the first twenty structures, out of one hundred, were analyzed. This cut-off was chosen as none of the structures showed any violations of experimental restraints and an equal number of structures was required for the comparison between various structure calculation protocols. In this first stage, the conformation was searched via restrained MDSA starting from random coordinates and progressively increasing the weight of the experimental restraints (see chapter 3 for details). Backbone rmsds of 20 structures were calculated in MOLMOL [115] for residues Cys2 to Cya8, which are common to all cyclic peptides in this study.

Unlike water-refined structures, the set of restraints used to calculate the structures of peptide 2 in DMSO was obtained from the 100 ms mixing time 2D NOESY spectrum. The first attempts using a 200 ms mixing time NOESY spectrum led to too many violations and energy profiles where the evolution of the total energy and NOE energies were not consistent.

The structures calculated using dihedral restraints in vacuum are used as the starting structures for the FRM refinement in vacuum for both FRM only and FRM + dihedral restraints calculations. The iterative calculations of the FRM protocol adjust the distances and reduces the R-factor by comparing the theoretical and measured distances using 5 or 6 sets of distance restraints obtained at different mixing times. In XPLOR, the internal correlation time is assumed to be much shorter than the overall correlation time. This approximation reduces the description of the spectral density to two adjustable parameters: the order parameter S and the overall correlation time τ_c.

These parameters are set for particular proton-proton vectors. This approximation, which works well for proteins did not yield therefore satisfactory results for small peptides like peptides 1-4. Positive NOE cross peaks, observed in NOESY the spectra for Ncc residues and side chain-side chain NOE of long residues (Arg3 and Lys1/Lym1), and negative NOE for the peptide backbone and short side chains could not be reproduced at the same time. This clearly indicates that the tumbling is not uniform within the molecule. It was therefore decided not to include positive cross peaks in the refinement. A complete set of cross peaks was used during the calculations i.e. if certain cross peaks appeared only for longer mixing times these were included. Only unambiguously assigned cross peaks were used in calculations (unlike the calculations using a single set of restraints where the ambiguous assignments can be treated and analyzed with ARIA). These included the prochiral H^{β} and $GlyH^{\alpha}$ protons, the assignment of which was swapped during the calculations. Calculations were only performed for peptides 2-4 for which sets of NOESY experiments were acquired.

5.3.1 Structures of all the peptides

The following sections present the structures of all peptides calculated using protocol A-D. For all peptides, ensembles of structures calculated with and without dihedral restraints were represented in black and red, respectively. For each of these ensembles of structures, the corresponding Ramachandran plot showed thedispersion of the backbone dihedral angles. For simplicity, all peptides were represented with a similar view and only peptide 1 was numbered. Finally, final structures being discussed in section 5.7, the discussion concerned only the intermediate calculations.

5.3.1.1 Peptide 1

Protocol A: A comparison of the energies of the structures calculated with or without dihedral restraints did not show any significant variation in either the experimental or the empirical component. On the other hand, the rmsd of the latter structures has dropped dramatically (from 0.74 to 0.24 for the backbone). This is because the unrestrained structures mainly adopted two different conformations with approximately equivalent energies while the restrained structures converged to one of them. The

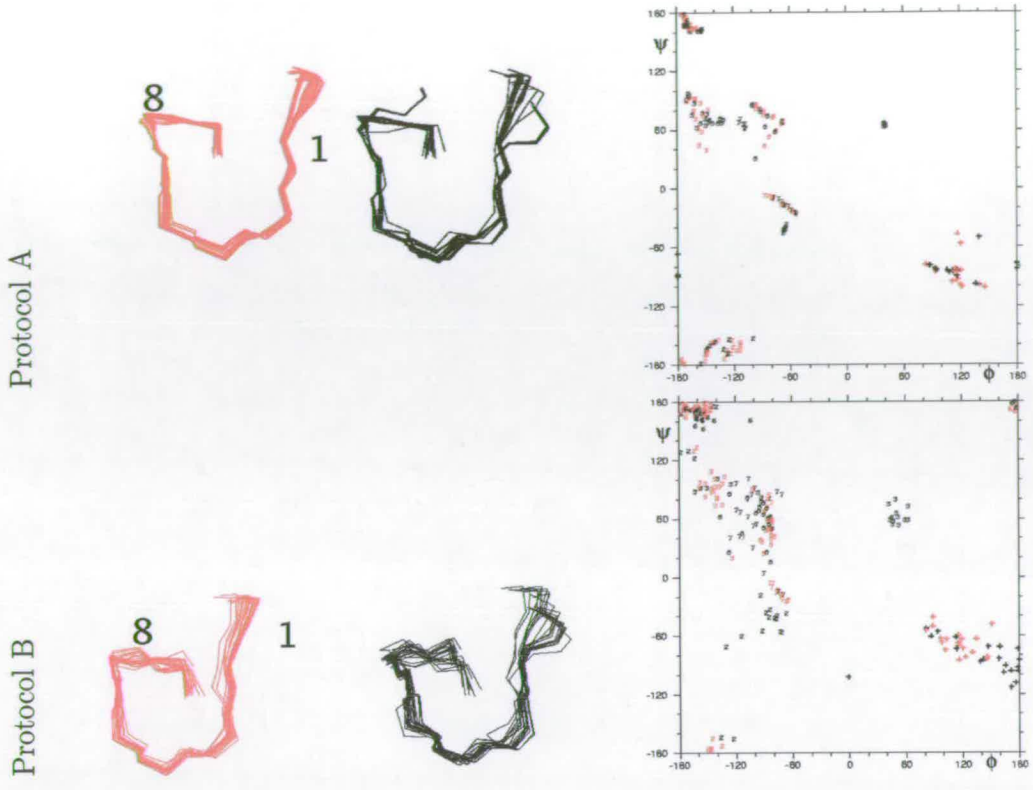


Figure 5.6: P1w. Ensembles of 20 lowest-energy structures calculated with (red) and without (black) dihedral restraints and corresponding Ramachandran plot.

conformation of the NOE only structures diverged for residues 2 and 3 in particular. The Ramachandran plot shows the following combinations of ϕ and ψ angles Arg3 (60, 60) / Cys2 (-90, -60) and Arg3 (-160, 60) / Cys2 (-160, -160). This shows that the conformation of one residue directly influences that of the next. Inclusion of dihedral restraint act as a filter which directs the conformation search toward the later conformer. The dihedral energy of the restrained structures stayed at zero. This means that all the low energy structures satisfy dihedral restraints. These restraints were set rather loosely and therefore can be completely satisfied without providing any energy penalty.

Protocol B: For peptide 1, structures calculated after inclusion of dihedral restraints adopt conformations that exist in structures where dihedral restraints were not used. These dihedral-restrained structures are calculated without violations of experimental restraints. Moreover, no increase of potential energy increase was observed. The convergence of the structures is also improved by the dihedral restraints (0.75→0.42

Å for backbone atoms of the peptide ring). Only remaining ambiguity is observed for residues Phe7 and Cya8 for which two conformations concerning the dihedral angles ψ_7 and ϕ_8 are compatible with experimental restraints (76±15 ,-73±10) or (-16±7 , -116±5). In water, a hydrogen bond between $6H^N$ and 4CO is formed in most of the structures. The existence of a hydrogen bond involving $6H^N$ is supported by the high temperature coefficient of $6H^N$ but this was not used as an explicit restraint during the structure calculations. As a consequence, some residues change their conformation compared with calculation in vacuum. This was the case for Asp5 in which the ψ angle changed from 70^o to 50^o. Similarly the ϕ angle of Gly4 is rotated from 120^o to 100^o. In vacuum the absence of dihedral restraints leads to appearance of two conformations for Arg3 and Cys2 with a similar energies, while the water-refined structures converged to one conformation only with ϕ and ψ angle at (-151,172) and (-138,88) for Cys2 and Arg3, respectively .

5.3.1.2 Peptide 2

Protocol A: Similarly to P1w, the rmsds for P2w decreased after inclusion of dihedral restraints. This time, however, the effect was much smaller. A characteristic feature of this structure calculation is very low NOE energies. This small number of medium range restraints, only 6, as compared with 14,m 21 and 18 in other peptides in water. Analysis of the structures calculated using dihedral restraints showed that their inclusion led to dihedral angles being pushed to their boundary values. This could be seen for example for ϕ angle of Gly4 and ψ angles of residues 3, 5 and 6. Such distribution of the dihedral angles along a single value illustrates the weak influence of the distance restraints during the process of calculation.

Protocol B: Similarly to the structure of peptide 1 in vacuum, the rmsds did not change significantly between the water structures refined with or without the dihedral restraints. In absolute terms, the rmsds of peptide 2 structures calculated in vacuum and water are the largest amongst all the peptides. As discussed previously, this is caused by the low number of medium range NOEs.

Protocol C: Incidentally, the rmsd values for structure calculated with a single set of restraints obtained from 200-250 ms mixing time NOESY spectrum are similar to

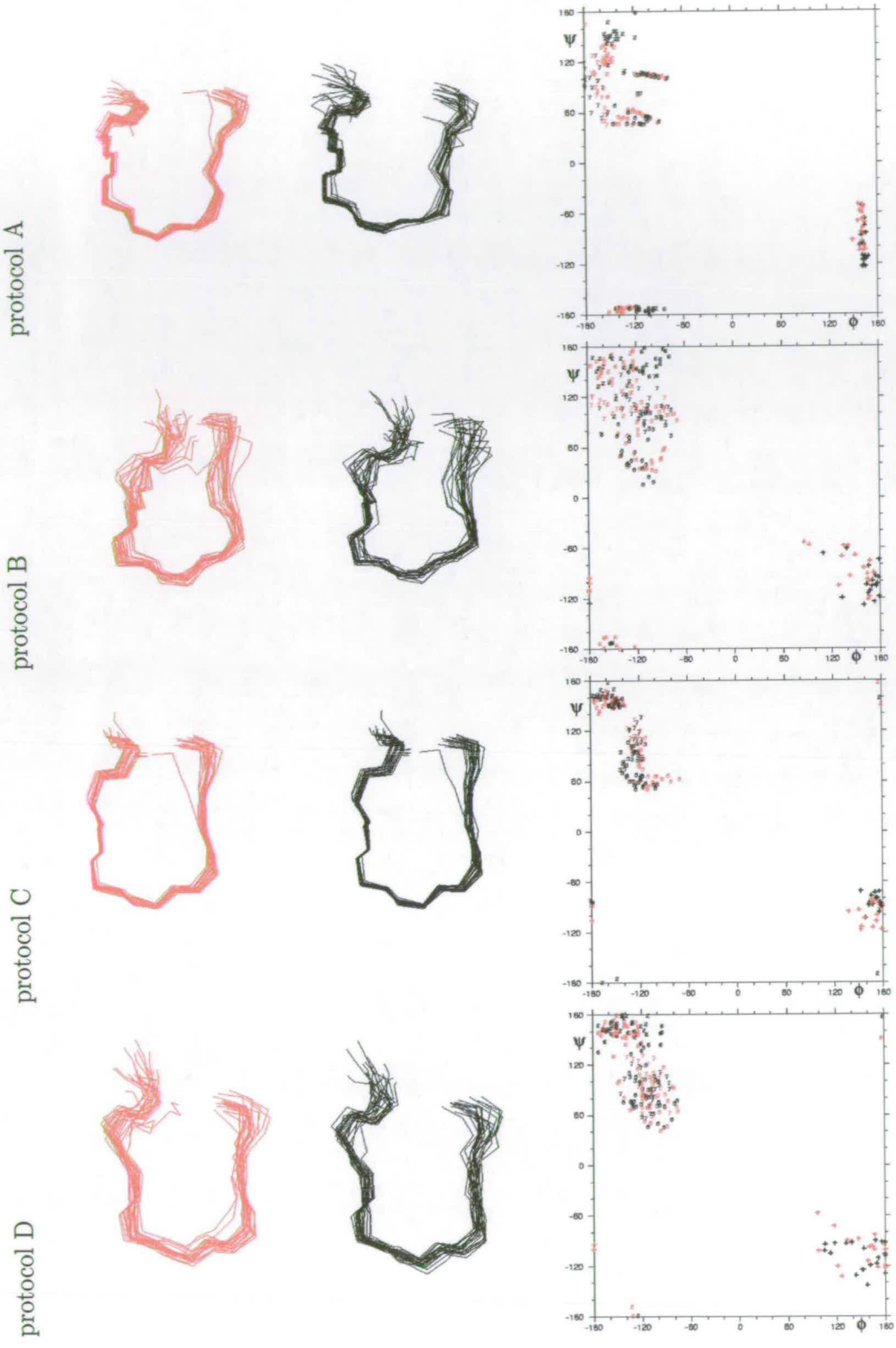


Figure 5.7: P2w. Ensembles of structures calculated with and without dihedral restraints.

those observed after FRM refined. It was noted before that the the NOE energy was too low in the former structures implying that these were not adequately restrained. The dihedral angles observed in FRM structures fall into the allowed regions of the Ramachandran plot shows which was not always the case with the previous set of structures (e.g. Phe7). This resulted in a relatively large changes in the backbone dihedral angles of residues Arg3 (-102, 103) to (-128, 87), Cys6 (-113,-174) to (-148, 153) and Phe7 (-166, 90) to (-124, 107). Inclusion of dihedral restraints did cause only minor changes, with exception of dihedral angles of Arg3 for which changed from (-102, 103) to (-119, 89).

Protocol D: For peptide 2 refined with protocol D, structures calculated with and without dihedral restraints are similar as, shown by the structure statistics (Table 5.9) and the Ramachandran plot (Fig. 5.7).

5.3.1.3 Peptide 3

Protocol A: Similarly to P1w, Cys2 and Arg3 in Pw3 structures adopt two conformations in structures calculated with distance restraints only, while the structures calculated with the dihedral restraints converge toward a single conformation. This is accompanied by a significant lowering of rmsds without any substantial increase in the total energy. Major changes are observed for residues 2 and 3. The dihedral angle ϕ and ψ of these residues changes from (80,120) / (-140,160) to (-100,150) for Arg3 and from (-120,-60) / (-160,160) to (-160,160) for Cys2. The structures calculated with dihedral restraints are similar to P1w. Once again changes of two dihedral angles affect the conformation of the rest of the molecule. Dihedral angle for Arg3 appears to be pushed towards the edge of the interval allowed by the dihedral restraints. However, unlike in peptides 1 and 2, these dihedral angles were not observed in structures calculated with distance restraints only.

Protocol B: Unlike peptide 1, the structure of peptide 3 is similar in vacuum and in water, i.e. the conformations which satisfied the presence of the hydrogen bond $6H^N \cdots 4CO$ seen in the water refined structures of peptide 1 is already present in the vacuum structures of peptide 3. As observed previously, the dihedral restraints orient peptide 3 into one of the conformations observed in NOE only structures. The

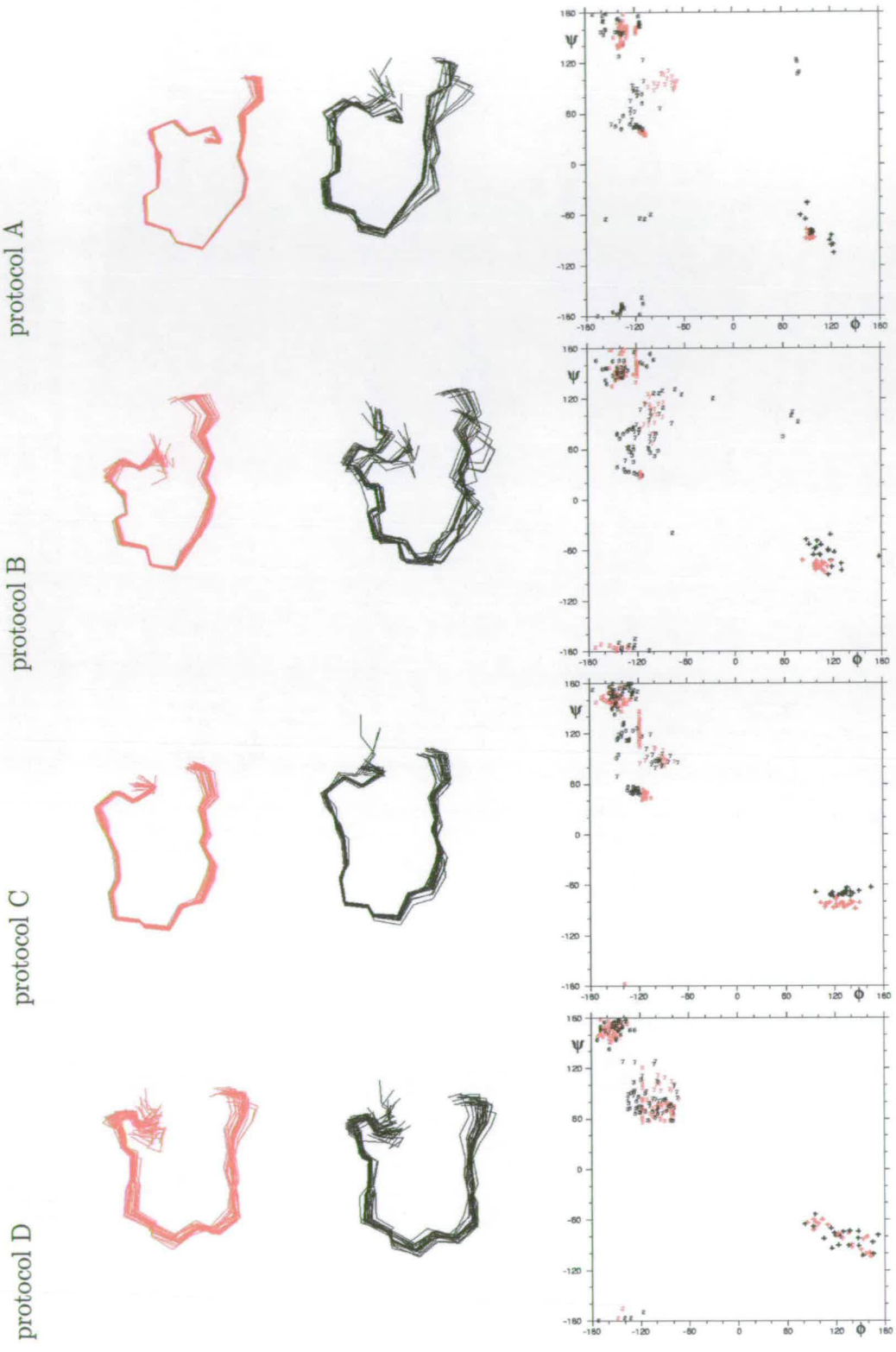


Figure 5.8: P3w. Ensembles of structures calculated with and without dihedral restraints.

Ramachandran plot shows that Arg3 does not satisfy both the distance restraints and the dihedral restraints and $\phi 3$ angle is responsible for increasing of the dihedral energy.

Protocol C: Similar distribution of dihedral angles is observed for the single mixing time structures and the FRM refined structures for peptide 3 with an exception of Arg3. After inclusion of dihedral restraints its ϕ angle rotates by -40^o. Compared to the starting structures this residue in the FRM refined structure this angle is pushed to the boarder value of the dihedral restraint.

5.3.1.4 Peptide 4

Protocol A: A comparison of P3w and P4w structures shows that both peptides have similar conformations. As for Pw3 the rmsd of the structures of P4w structures improved dramatically for the structures calculated with dihedral restrains. Similarly to P3w structure, the dihedral angle ϕ of Arg3 in structures with dihedral restraints was not observed in the structures calculated with distance restraints.

Protocol B: The observations made for peptide 3 are also valid for peptide 4. The values of the dihedral angles did not change significantly with water refinement and Arg3 is again pushed to the edge of the allowed region when dihedral restraints are used. The only noticeable change, compared with the vacuum structures, involved Phe7 ϕ angle which changed from -140^o to -120^o in both dihedral-restrained and unrestrained structures.

Protocol C: Similarly to peptide 3, larger differences between the dihedral angles in structure restrained only by NOEs and dihedral angles were observed for residue Arg3 which is displaced from (-148, 135) to (-117,134)

5.3.1.5 Peptide 2 using H_2O/D_2O data

Protocol A: The Ramachandran plot and the ensemble of structures shows that the structures calculated using H_2O and D_2O restraints are similar to those calculated using the H_2O restraints only. The energy profiles now show larger NOE energies comparable to those seen for other compounds in this series of structure calculations. The rmsds have improved and are now comparable to those obtained for P3w structures,

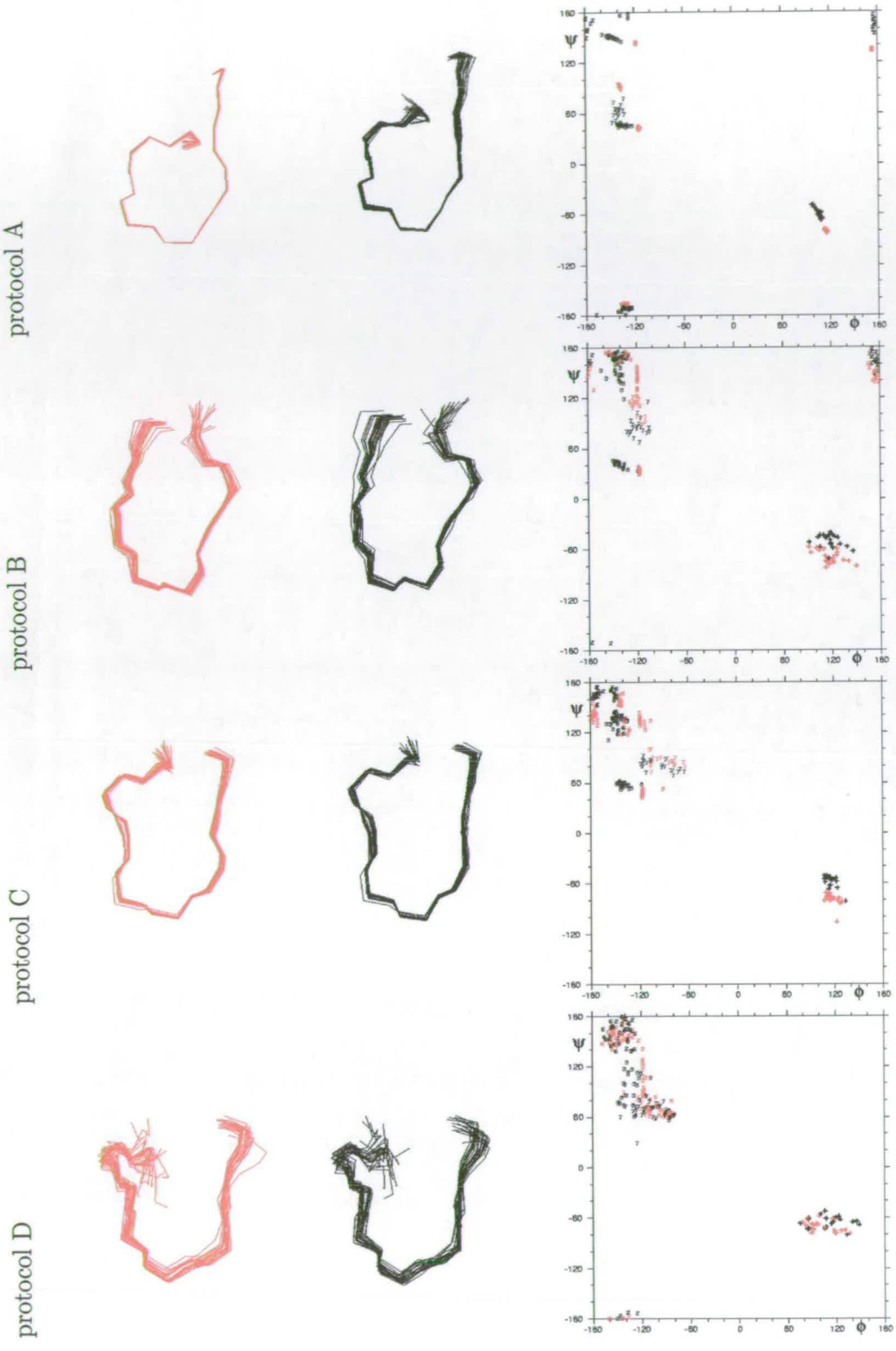


Figure 5.9: P4w. Ensembles of structures calculated with and without dihedral restraints.

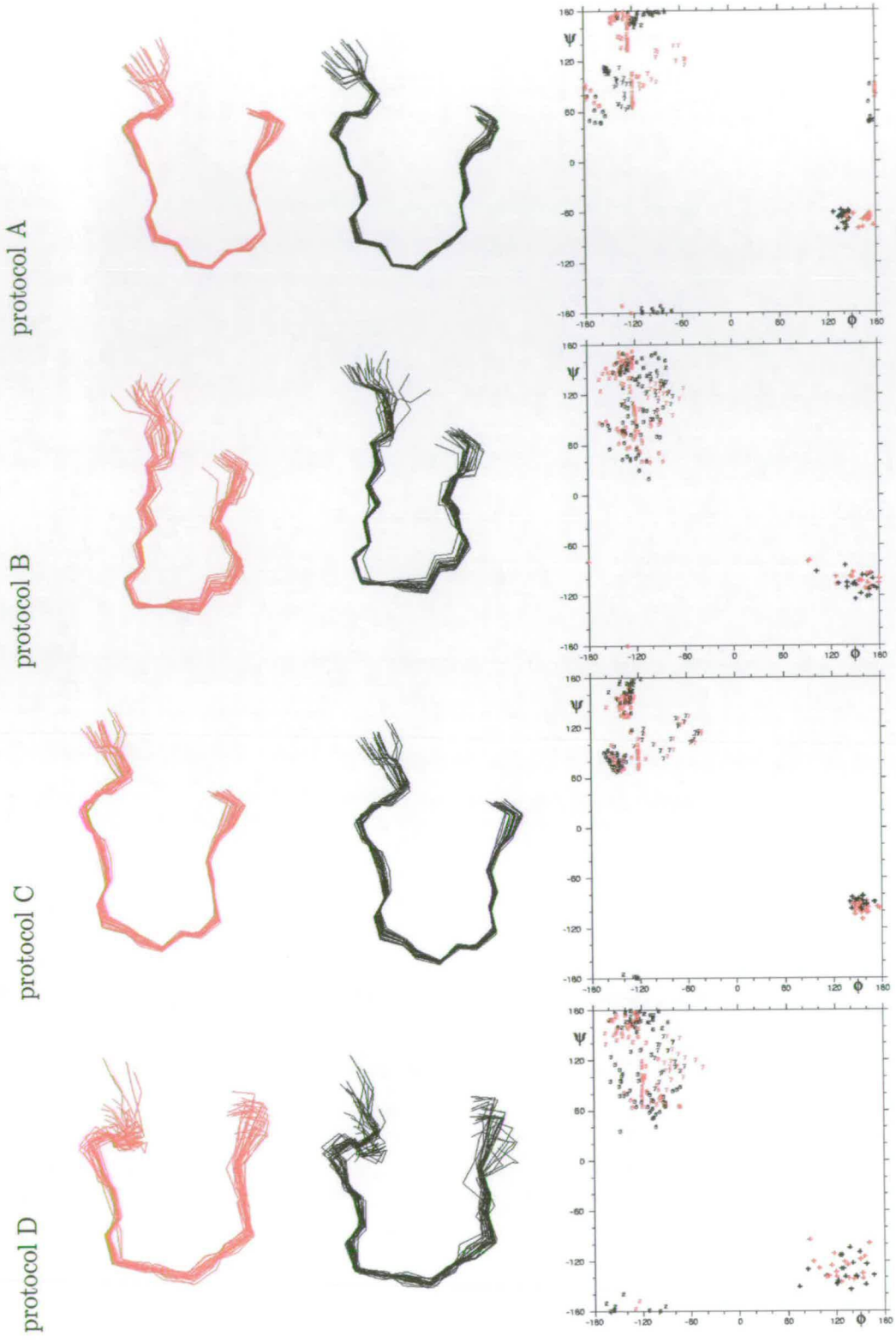


Figure 5.10: P2w using H_2O D_2O. Ensembles of structures calculated with and without dihedral restraints.

which were based on similar number of NOE restraints. As was the case for H_2O structures obtained by inclusion of dihedral restraints, the H_2O and D_2O structures show that dihedral angles for some residues are pushed towards the edge of the restrained region. this is observed for residues Cys2, Arg3 and Asp5 in structures calculated with H_2O and D_2O restraints.

Protocol B: The water refinement rotate the ϕ angle of Asp5 closer to the value observed in peptides 1, 3 and 4 which is more consistent with the formation of a hydrogen bond between $6H^N$ and 4CO. The ψ and ϕ angles of Phe7 shows a large spread of values particularly for structure without dihedral restraints and on average rotates with regard to the vacuum structures by about $+50^o$ in the NOE only structures. Arg3 comes also in a better agreement with the dihedral restraints. Unlike P3w and P4w, the ϕ angle is not constantly pushed towards the edge of the restraining region. These changes bring P2w structures closer to P3w and P4w structures. This seems reasonable comparing between various peptides the experimental data (section 5.9).

Protocol C: FRM protocol for structure calculations based on combined H_2O and D_2O data was set in such a way as to distinguish the labile protons from the aliphatic and aromatic ones. Two groups of restraints were created, the first for the labile protons (extracted from H_2O spectra) and the second containing non-labile protons (extracted from D_2O spectra). This separation was necessary as to account for the frequency at which the two sets of data were acquired (800 MHz for D_2O and 600 MHz for H_2O). Structures obtained using the full relaxation matrix from spectra measured in H_2O and D_2O brings the structures in a better agreement with the dihedral restraints compared to the structures calculated with protocol A. When dihedral restraints are included, residues Asp5 and Cys2, which were pushed towards the edge of the restraining region in protocol A, are in the regions defined by dihedral restraints. Similarly to the other structures calculated with protocol C, Arg3 remains the only residue pushed by the dihedral restraints. Overall, the structures of P2w calculated from H_2O and D_2O resemble more those calculated with H_2O restraints only, exception with residue Phe7 which angles are (-125, 110) in H_2O and (-75, 111) in H_2O + D_2O structures.

As expected, the precision of structures improved by the addition of more distance restraints (282 distance restraints in H_2O for FRM against 494 in H_2O / D_2O from

the series of NOESY spectra). The precision and the number of restraints are also comparable to those found for P3w and P4w.

5.3.1.6 Peptide 2 in DMSO

Protocol A: Large values of $^3J_{H^N H^\alpha}$ coupling constants together with a limed set of other coupling constant, were are used to restrain the dihedral angles. These experimental data (Table 4.2) indicate that ϕ angles of residues 3, 6, 7 and 8 should be around -120^o. For some residues, the inclusion of dihedral restraints cause a large rotation of ϕ angle (e.g. -40^o for Cys6).

Protocol B: The most important changes between the vacuum and DMSO structures occur for ψ angle of residue 5 and ϕ angle of residue 6. This is the case for both dihedral-restrained and unrestrained structures. The dihedral angle ψ_5 and ϕ_5 rotate on average by 60^o and -60^o. The rotation of ϕ_6 reduces the distance between the backbone 6H^N and the carboxylate of Asp5 making a hydrogen bond backbone to side chain possible. At the same time, the carbonyl of Asp5 become available for an hydrogen bond with 3H^N . Smaller adjustments of ϕ_7 and ψ_7 angles favor a formation of the hydrogen bond between 1C' and 7H^N but also between 7C' and 1H^N. Prior to these conformational changes, no hydrogen bonds were observed in the structures calculated in vacuum. After refinement in explicit DMSO all hydrogen bonds are observed in most of the structures.

Protocol C: The use of FRM brought about several rearrangements of dihedral angles. In particular, 6ψ and 7ψ undergo rotations of about +30^o and -20^o respectively.The conformation of the other residues is only slightly or not at all changed.

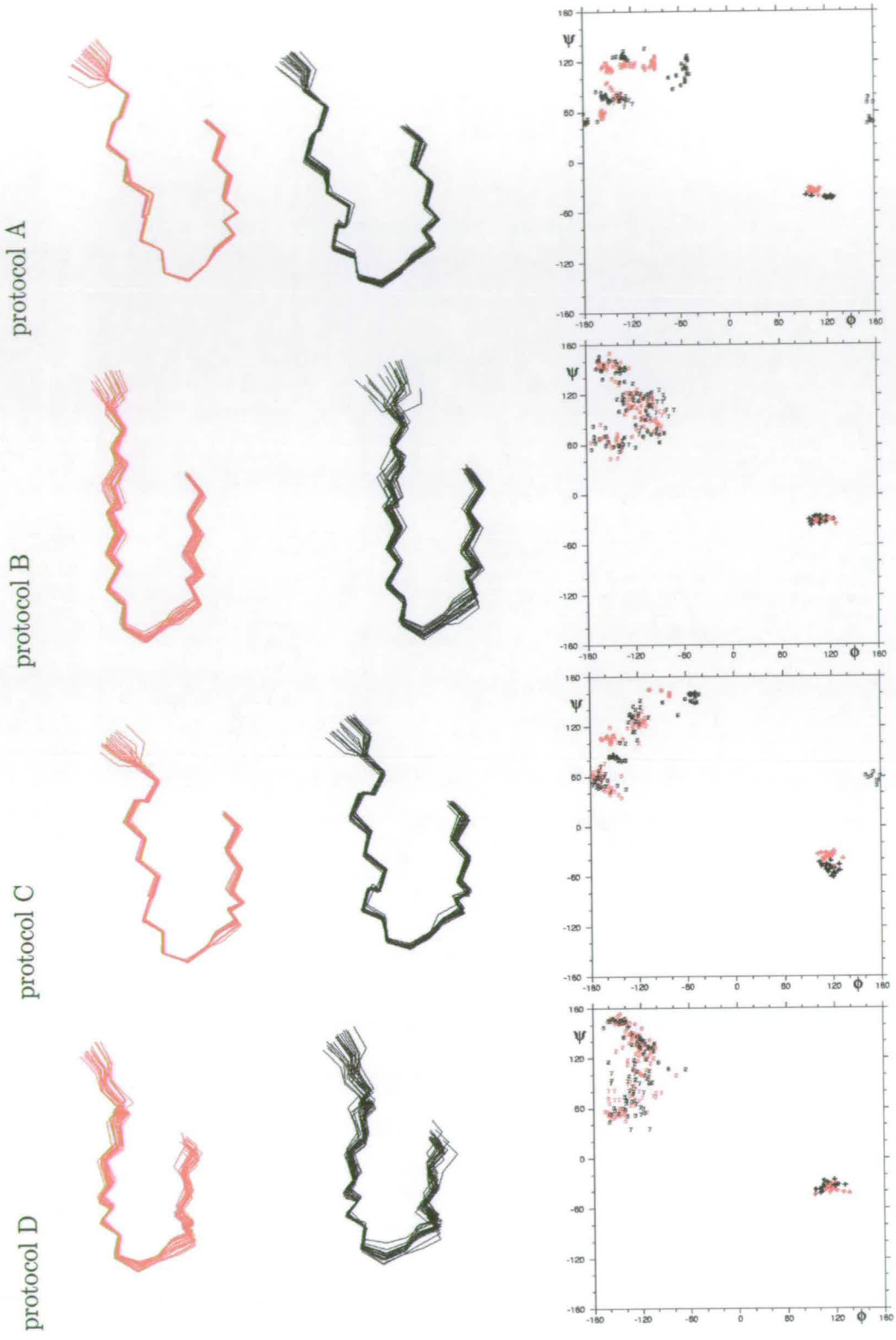


Figure 5.11: P2d. Ensembles of structures calculated with and without dihedral restraints.

5.3.2 Structure statistics for peptides calculated with protocols A-D

The calculations were carried out without stereospecific assignment of the prochiral H^{β} protons as these cannot be unambiguously determined for most of the residues. The distance restraints are classified as: intra-residue, sequential and medium range. The medium range category corresponds to cross peaks observed between residue i and i+2, i+3 or i+4. Due to the cyclic nature of these short peptide (8 residues) none of the amino acids can be more than 4 residues apart. This classification does not include ambiguous assignments and distances to the metal binding site (MBS) or polyethylene glycol tail (Ncc) moieties. Anti-distance restraints (ADR) based on the analysis of the finger print region were used. This section presents structure statistics for all peptides according to the protocol employed for structure calculations.

Table 5.5: **Structure statistics for peptides 1-4 calculated using protocol A:**

Peptides	P1w		P2w		P3w		p4w	
E_{pot}	15.98±0.69	16.09±0.47	15.84±3.02	14.78 ±0.17	25.98±0.34	26.57±0.20	32.02±0.18	32.94± 0.15
E_{bond}	1.87±0.07	1.88±0.06	1.66±0.09	1.62±0.01	1.85±0.03	1.90±0.02	1.72±0.01	1.76±0.01
E_{ang}	3.07±0.40	3.25±0.40	6.31±0.28	6.20±0.03	13.89±0.17	14.23±0.11	21.54±0.03	21.91±0.06
E_{impr}	0.33±0.21	0.46±0.13	0.19±0.04	0.18±0.01	0.73±0.05	0.75±0.02	0.96±0.00	1.07±0.01
E_{vdw}	7.00±0.21	7.07±0.26	6.58±0.34	6.47±0.10	6.80±0.15	6.97±0.07	6.40±0.02	6.39±0.02
E_{noe}	3.69±0.26	3.43±0.29	1.11±2.34	0.31±0.19	2.71±0.11	2.72±0.14	1.39±0.14	1.81±0.10
E_{cdih}	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00	0.00±0.00
Dist. rest.	112		118		177		164	
Intr.,Seq., Med., ADR.	57,32,14,9		51,27,6,8		74,56,21,12		73,53,18,8	
Viol. > 0.3 Å	0	0	0	0	0	0	0	0
Dih. rest.	2	8	2	9	2	8		
Viol. > 5^o	0	0	0	0	0	0	0	0
Pairwise rmsd (Å)	2-8	2-8	2-8	2-8	2-8	2-8	2-8	2-8
Backbone	0.7±0.5	0.2±0.1	0.48±0.26	0.37±0.14	0.56±0.36	0.13±0.05	0.19±0.08	0.04±0.03
Heavy atoms	2.2±0.7	1.6±0.4	1.49±0.49	1.31±0.30	1.40±0.46	1.06±0.29	1.17±0.32	1.14±0.33

Peptides	P2w D_2O+H_2O		P2d	
E_{pot}	23.00 ± 0.12	24.95 ± 0.28	16.45 ± 0.22	20.70 ± 1.21
E_{bond}	2.11 ± 0.03	2.14 ± 0.02	1.83 ± 0.02	2.09 ± 0.29
E_{ang}	6.97 ± 0.06	7.36 ± 0.11	6.34 ± 0.11	9.09 ± 0.77
E_{impr}	0.24 ± 0.02	0.26 ± 0.01	0.17 ± 0.01	0.81 ± 0.16
E_{vdw}	7.20 ± 0.12	7.55 ± 0.08	6.26 ± 0.03	6.31 ± 0.14
E_{noe}	6.47 ± 0.16	7.61 ± 0.23	1.85 ± 0.08	2.40 ± 0.09
E_{cdih}	0.00 ± 0.00	0.03 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Dist. rest.	113+64		159	
Intr.,Seq., Med., ADR.	104,48,10,8		75,51,5,0	
Viol. > 0.3 Å	0	0	0	0
Dih. rest.	2	9	2	9
Viol. > 5^o	0	0	0	0
Pairwise rmsd (Å)	2-8	2-8	2-8	2-8
Backbone	0.42 ± 0.29	0.14 ± 0.10	0.23 ± 0.10	0.15 ± 0.08
Heavy atoms	1.30 ± 0.41	1.08 ± 0.41	1.38 ± 0.36	1.16 ± 0.32

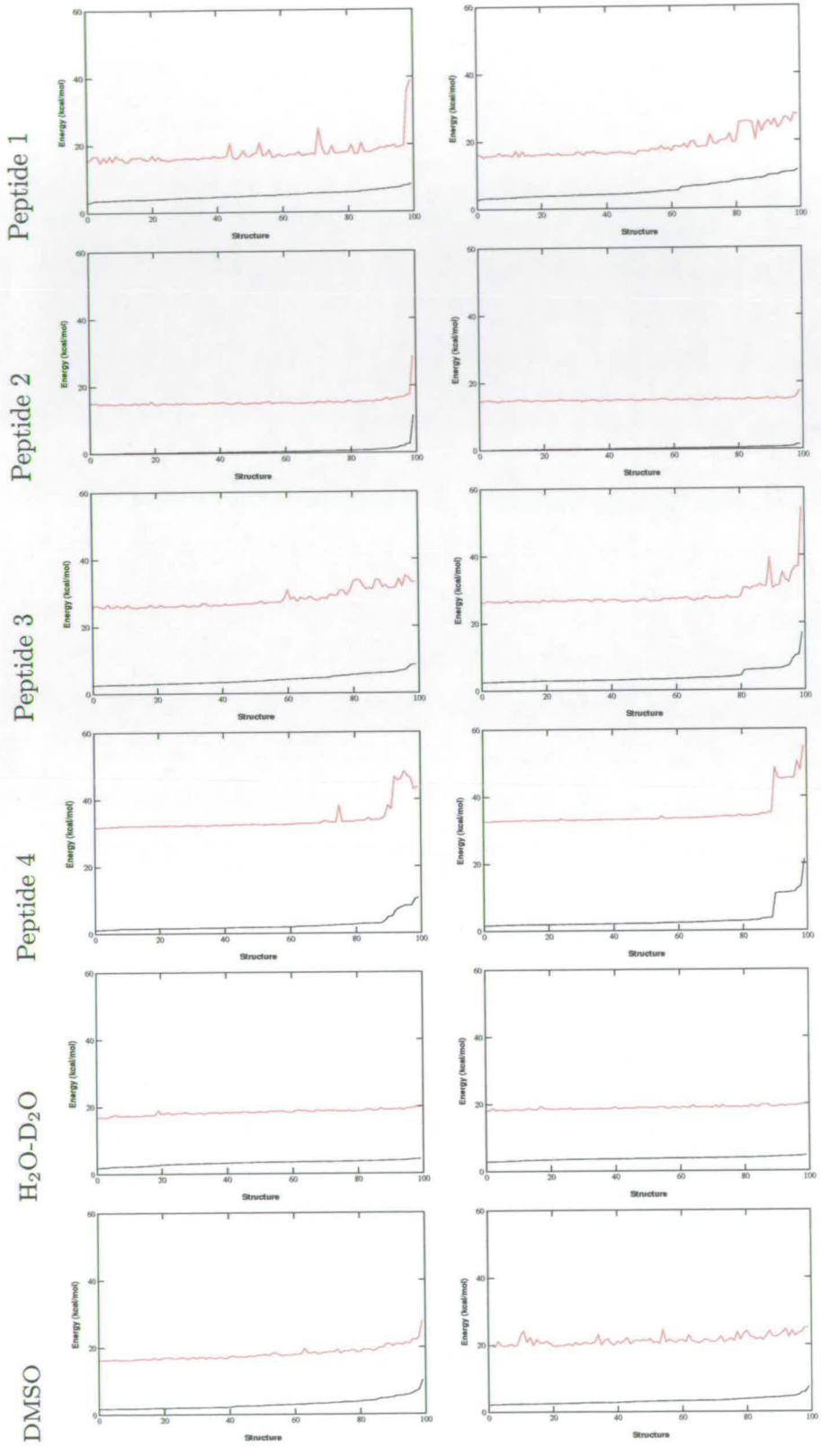


Figure 5.12: **Energy diagrams for structures of all the peptides using protocol A:** For each peptide, NOE and total energies are represented in black and red, respectively. Calculations performed with dihedral restraints and without are on the right and the left, respectively.

Table 5.6: **Structure statistics of peptide 1-3 calculated using protocol B:**

Peptides	P1w		P2w		P3w	
E_{pot}	-3291.98±266.33	-3327.24±117.09	-4422.84±225.40	-4434.90±281.76	-6709.11±291.30	-6638.53±446.57
E_{bond}	104.15 ± 9.09	105.99 ± 4.49	142.26 ± 9.51	142.92 ± 10.23	225.87 ± 10.42	223.03 ± 15.30
E_{ang}	22.94 ± 6.18	27.67 ± 3.97	24.63 ± 2.85	24.73 ± 1.53	49.27 ± 3.94	47.70 ± 3.54
E_{impr}	3.97 ± 1.31	5.26 ± 1.44	73 ± 0.78	3.27 ± 0.86	5.45 ± 1.75	5.27 ± 0.76
E_{vdw}	414.34 ± 39.68	425.62 ± 20.76	544.86 ± 45.82	553.71 ± 42.87	850.14 ± 48.11	830.04 ± 66.65
E_{noe}	15.69 ± 2.39	16.61 ± 1.70	4.59 ± 1.16	5.34 ± 0.97	18.61 ± 2.81	20.89 ± 1.29
E_{cdih}	0.01 ± 0.03	0.00 ± 0.00	0.00 ± 0.01	0.01 ± 0.02	0.00 ± 0.00	0.55 ± 0.18
Viol. dist.> 0.3 Å	0	0	0	0	0	0
Viol. dih.> 5^o	0	0	0	0	0	0
Pairwise rmsd (Å)	2-8	2-8	2-8	2-8	2-8	2-8
Backbone	0.75 ± 0.35	0.42 ± 0.17	0.75 ± 0.25	0.74 ± 0.26	1.07 ± 0.64	0.27 ± 0.10
Heavy atoms	1.97 ± 0.55	1.81 ± 0.40	2.11 ± 0.58	1.78 ± 0.39	2.57 ± 1.23	1.28 ± 0.29
Peptides	p4w		P2w D_2O+H_2O		P2d	
E_{pot}	-8685.28±517.98	-8929.41±570.58	-4241.59±204.20	-4240.84±223.20	-3241.07±142.23	-3212.99±79.52
E_{bond}	293.63 ± 18.24	305.09±21.42	141.38 ± 7.85	142.64 ± 7.87	6.51 ± 0.45	6.80 ± 0.40
E_{ang}	68.78 ± 2.46	72.08 ± 3.30	33.11 ± 5.05	34.95 ± 3.66	29.64 ± 3.77	29.26 ± 2.60
E_{impr}	5.94 ± 0.63	8.52 ± 1.35	3.91 ± 1.27	4.79 ± 1.32	4.86 ± 0.95	5.35 ± 1.02
E_{vdw}	1097.56±82.48	1136.82±106.33	537.90±40.41	544.28 ± 37.39	-1166.42± 64.89	-1148.82± 38.52
E_{noe}	17.86 ± 1.93	25.54 ± 1.56	29.61 ± 2.83	32.53 ± 2.35	9.96 ± 1.29	12.50 ± 2.70
E_{cdih}	0.00 ± 0.00	0.58 ± 0.28	0.00 ± 0.00	0.29 ± 0.22	0.00 ± 0.00	0.01 ± 0.02
Viol.dist. > 0.3 Å	0	0	0	0	0	0
Viol. dih.> 5^o	0	0	0	0	0	0
Pairwise rmsd (Å)	2-8	2-8	2-8	2-8	2-8	2-8
Backbone	0.35 ± 0.17	0.28 ± 0.10	0.39 ± 0.22	0.46 ± 0.22	0.55 ± 0.27	0.41 ± 0.14
Heavy atoms	1.18 ± 0.32	1.25 ± 0.35	1.44 ± 0.36	1.55 ± 0.40	1.67 ± 0.45	1.37 ± 0.35

Table 5.7: **Structure statistics of peptides 2-4 calculated using protocol C:**

Peptides	P2w		P3w		P4w	
E_{pot}	94.92 ± 6.83	99.69 ± 5.85	360.66 ± 15.77	358.68 ± 13.39	247.20 ± 13.07	259.07 ± 13.30
E_{bond}	1.56 ± 0.18	1.66 ± 0.20	6.66 ± 0.57	6.75 ± 0.52	4.31 ± 0.59	4.66 ± 0.62
E_{ang}	21.28 ± 1.03	21.95 ± 1.08	55.88 ± 2.52	55.99 ± 1.94	79.61 ± 2.10	79.14 ± 2.05
E_{impr}	1.68 ± 0.22	1.85 ± 0.24	3.10 ± 0.40	3.09 ± 0.32	4.29 ± 0.54	4.19 ± 0.34
E_{vdw}	1.10 ± 0.49	1.07 ± 0.54	3.10 ± 6.17	1.55 ± 0.61	1.85 ± 1.06	1.68 ± 1.07
E_{rela}	69.29 ± 6.08	72.96 ± 5.47	291.93 ± 17.14	291.09 ± 11.89	157.14 ± 10.32	169.11 ± 10.90
E_{cdih}	0.00 ± 0.01	0.19 ± 0.11	.00 ± 0.00	0.19 ± 0.08	0.00 ± 0.00	0.30 ± 0.13
Noesy τ_{mix} (ms)	Number of cross peaks					
55	38		42		51	
75	46		50		58	
100	50		66		75	
150			93		89	
200	71		102		94	
400	77		117		107	
Total	282		470		474	
Pairwise rmsd (Å)	2-8	2-8	2-8	2-8	2-8	2-8
Backbone	0.34 ± 0.14	0.37 ± 0.16	0.31 ± 0.24	0.19 ± 0.07	0.20 ± 0.09	0.19 ± 0.08
Heavy atoms	1.34 ± 0.31	1.34 ± 0.31	1.24 ± 0.39	1.14 ± 0.30	1.18 ± 0.35	1.20 ± 0.35

Table 5.8: **Structure statistics of peptide 2 in H_2OH_2O and DMSO calculated using protocol C:**

Peptides	p2w D2O		P2d	
E_{pot}	206.27 ± 8.96	217.89 ± 14.01	298.19 ± 14.13	297.93 ± 11.83
E_{bond}	4.11 ± 0.29	4.81 ± 0.41	4.52 ± 0.24	4.57 ± 0.31
E_{ang}	25.21 ± 1.14	26.67 ± 4.76	28.12 ± 1.84	30.84 ± 2.26
E_{impr}	1.87 ± 0.52	2.24 ± 0.60	2.37 ± 0.28	2.72 ± 0.27
E_{vdw}	0.91 ± 0.32	1.26 ± 1.43	1.15 ± 0.50	0.91 ± 0.26
E_{rela}	174.17 ± 7.86	182.73 ± 7.31	262.04 ± 13.26	258.85 ± 12.59
E_{cdih}	0.00 ± 0.00	0.18 ± 0.14	0.00 ± 0.00	0.05 ± 0.03
Noesy τ_{mix} (ms)	Number of cross peaks			
55	38+19		100	
75	46+29		132	
100	54+44		159	
200	71+56		188	
400	77+64		204	
Total	282+212		470	
Pairwise rmsd (Å)	2-8	2-8	2-8	2-8
Backbone	0.19 ± 0.07	0.18 ± 0.06	0.24 ± 0.09	0.18 ± 0.07
Heavy atoms	1.17 ± 0.43	1.19 ± 0.44	1.40 ± 0.39	1.13 ± 0.30

Table 5.9: **Structure statistics of peptides 2-4 calculated using protocol D:**

Peptides	P2w		P3w		P4w	
E_{pot}	-4321.01 ±303.34	-4303.09±312.74	-6506.74±369.17	-6534.51±355.43	-8774.48±711.57	-8804.15±685.62
E_{bond}	141.91 ± 11.38	140.14 ± 11.34	226.59 ± 13.30	228.51 ± 13.99	300.70 ± 25.71	304.10 ± 26.62
E_{ang}	25.51 ± 2.22	26.35 ± 2.62	51.06 ± 4.84	51.72 ± 5.48	63.42 ± 2.36	64.36 ± 2.70
E_{impr}	3.17 ± 1.16	3.53 ± 1.27	4.61 ± 0.80	4.61 ± 0.80	4.29 ± 1.35	5.44 ± 1.27
E_{vdw}	556.24 ± 55.85	545.45 ± 57.45	849.33 ± 57.85	851.75 ± 70.23	1130.47 ± 122.47	1136.97 ± 134.29
E_{rela}	54.18 ± 8.07	57.58 ± 5.51	180.21 ± 14.48	180.21 ± 14.48	108.17 ± 9.84	115.44 ± 10.32
E_{cdih}	0.00 ± 0.00	0.01 ± 0.02	0.00 ± 0.00	0.01 ± 0.02	0.00 ± 0.00	0.08 ± 0.13
Pairwise rmsd (Å)	2-8	2-8	2-8	2-8	2-8	2-8
Backbone	0.74 ± 023	0.77 ± 0.29	0.54 ± 0.18	0.54 ± 0.17	0.66 ± 0.23	0.62 ± 0.28
Heavy atoms	2.22 ± 05.7	1.96 ± 0.49	1.43 ± 0.33	1.52 ± 0.40	1.69 ± 0.40	1.57 ± 0.37

Peptides	p2w D2O		P2d	
E_{pot}	-4289.02 ± 221.07	-4267.66 ± 218.99	-3156.66 ± 132.97	-3231.45 ± 85.04
E_{bond}	144.89 ± 7.49	145.00 ± 9.12	9.52 ± 0.69	9.22 ± 0.68
E_{ang}	33.10 ± 4.11	31.30 ± 3.15	36.06 ± 4.20	36.11 ± 4.93
E_{impr}	4.02 ± 1.46	4.23 ± 1.77	5.29 ± 1.20	5.62 ± 1.16
E_{vdw}	554.94 ± 44.09	555.17 ± 44.00	-1271.66 ± 55.15	-1303.35 ± 44.56
E_{rela}	140.47 ± 14.09	137.13 ± 9.58	236.85 ± 10.26	232.43 ± 9.90
E_{cdih}	0.00 ± 0.00	0.02 ± 0.04	0.00 ± 0.00	0.01 ± 0.03
Pairwise rmsd (Å)	2-8	2-8	2-8	2-8
Backbone	0.62 ± 0.23	0.55 ± 0.19	0.46 ± 0.17	0.38 ± 0.15
Heavy atoms	1.74 ± 0.36	1.70 ± 0.36	1.55 ± 0.46	1.32 ± 0.31

5.4 Free molecular dynamics in explicit solvent

Free molecular dynamics simulations (free MD) were carried out for 1 ns in XPLOR to investigate conformational stability of the low energy structures obtained via NMR restrained molecular dynamics. The closest to the mean structures obtained using calculations in water were soaked in a solvent box (water or DMSO) as described in chapter 3. The system was equilibrated prior to free molecular dynamics simulation to avoid any temperature instabilities. Starting positions of the solvent molecules do not correspond to the equilibrium state and close contacts between the water molecules and the peptide increase the van der Waals energies. Initial equilibration is therefore required. The relaxation of the water molecules was achieved by two minimizations where the coordinates of the solute were successively fixed and restrained by a harmonic function holding the coordinates close to their initial state. Once the water molecules were relaxed, the system was brought to equilibrium at 300 K by means of two molecular dynamics runs. The coordinates of the solute were restrained using a harmonic function during the first 10 ps and released during the following 20 ps. These steps allowed the system to reach the desired temperature without dramatic fluctuations and to start the conformational analysis from an equilibrated system. The SHAKE [116] algorithm was used to fix the length of bonds involving hydrogen atoms (solvent and solute). The trajectory was saved every 0.2 ps during the free MD and analyzed in VMD-xplor [117].

The stability of the system (solute+solvent) during the simulation was monitored by computing the evolution of the available physical parameters [2]. Parameters such as temperature, Van Der Waals energy and the improper angle energy were constant during the 1 ns simulation, revealing the relative stability of the system over the entire simulation. This means that the volume of the thin layer surrounding the peptide was constant (otherwise the VDW energy would drop dramatically) keeping the solute soaked during the entire simulation. The stability of the improper energy, arising from the solute, shows that the structures were in relatively stable geometries during the calculations (Fig. 5.13).

[2] The pressure and the volume cannot be controlled in XPLOR.

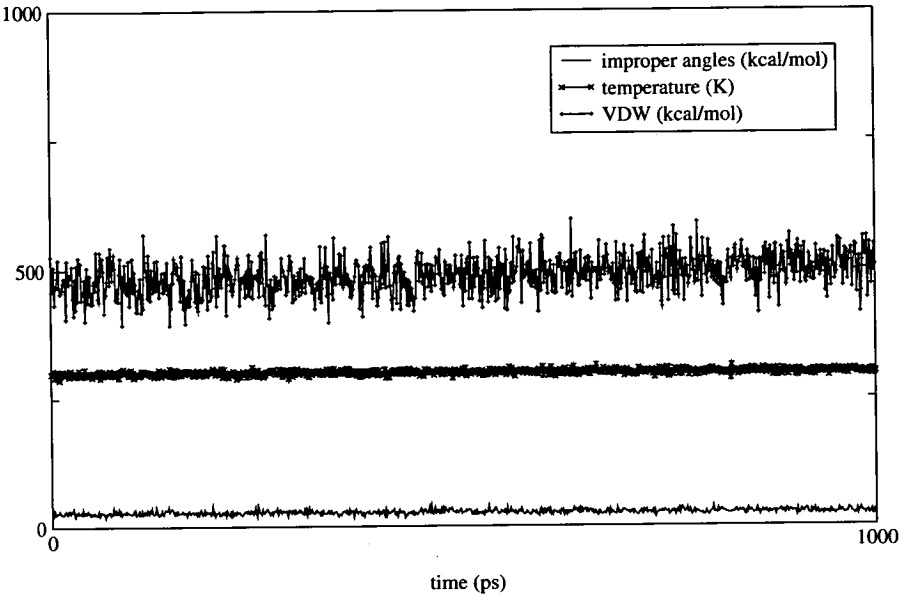


Figure 5.13: VDW energy, impropers energy and temperature monitored over the entire unrestrained molecular dynamics run.

In order to gain a more quantitative insight into the conformational flexibility of the compounds, the dihedral angles of the peptide core and the distances of the putative hydrogen bonds were followed over the course of the simulation. The values of the dihedral angles ψ_i and ϕ_{i+1} are interdependent and, because most of the experimental data available are related to these angles, they are the focal point of the following analysis. The rmsd of the entire peptide backbone that encompasses both rings and that of the CRGDC loop only were also computed. The comparison of these two types of rmsd is expected to give an indication about the flexibility of the RGD motif.

Free molecular dynamics in water

The standard deviations for ϕ angles during the simulation are summarized in Table 5.10. Previous studies showed that fluctuations of c.a. $\pm 40^o$ for the backbone dihedral angles are not to be interpreted as a conformational change for standard amino acids [118, 119]. According to Pfeiffer *et al.* [120] the standard deviations encountered in folded regions of a protein have an amplitude of less than 30^o. In the present work, except for the ϕ_4 angle in P4w and P2w, the standard deviation of the ϕ angles did not exceed 32^o (Table 5.10).

Table 5.10: Average values and standard deviation of the ϕ dihedral angles (in degrees) during the free molecular dynamics.

Residue	peptide1	peptide 2	peptide 3	peptide 4
Lys1 / Lym1	-104±15	-116±14	-85±13	-83±12
Cys2	-123±32	-97±21	-89±14	-89±16
Arg3	-109±19	-89±11	-72±17	-82±13
Gly4	101±19	-92±44	121±19	156±72
Asp5	-90±20	-102±24	-120±23	-113±27
Cys6	-85±15	-143±20	-100±20	-150±17
Phe7	-95±15	-87±14	-113±13	-89±24
Cya8	-116±26	-103±18	-99±21	-109±19

This suggests that the peptides in water fluctuate around one conformation, albeit to a varying degree of flexibility. This flexibility will be examined using the time series of the dihedral angles and the backbone rmsd (either those of the CRGDC motif or entire backbone).

In water, fast fluctuations of dihedral angles of order $\pm 20^o$ are observed throughout the course of the free molecular dynamics and appear as noise in the trajectories presented in Fig. 5.14. In addition, continuous changes or sudden jumps were also observed. For example, residues 2 and 3 in P1w clearly display two conformations whereas the ϕ_4 angle varies in a more progressive manner.

The trajectories for ϕ angles (Fig. 5.14) shows that, overall, no concerted motion occurs for the peptides in water. For example, despite a large rotation of ϕ_4 angle of glycine, ϕ angles of the neighboring residues in P4w and P2w are not affected. On the other hand some correlation can be found for peptide 1 for residues 1-3 or 4-6. Nevertheless, these simulations show that the studied peptides do not exist in several distinct states and that the molecules can accommodate local flexibility in number of ways, resulting in conformations that are close to the original one. The inspection of the putative hydrogen bonds shows that the hydrogen bonds $6H^N \cdots 4CO$ found in many experimentally-restrained structures is poorly conserved during the unrestrained simulations. The corresponding distance donor-acceptor is systematically above 2.5Å which is too large to be classified as a hydrogen bond (typically less than 2.4Å). This distance is well correlated with the fluctuation of the ϕ_5 angle (figure 5.15) which

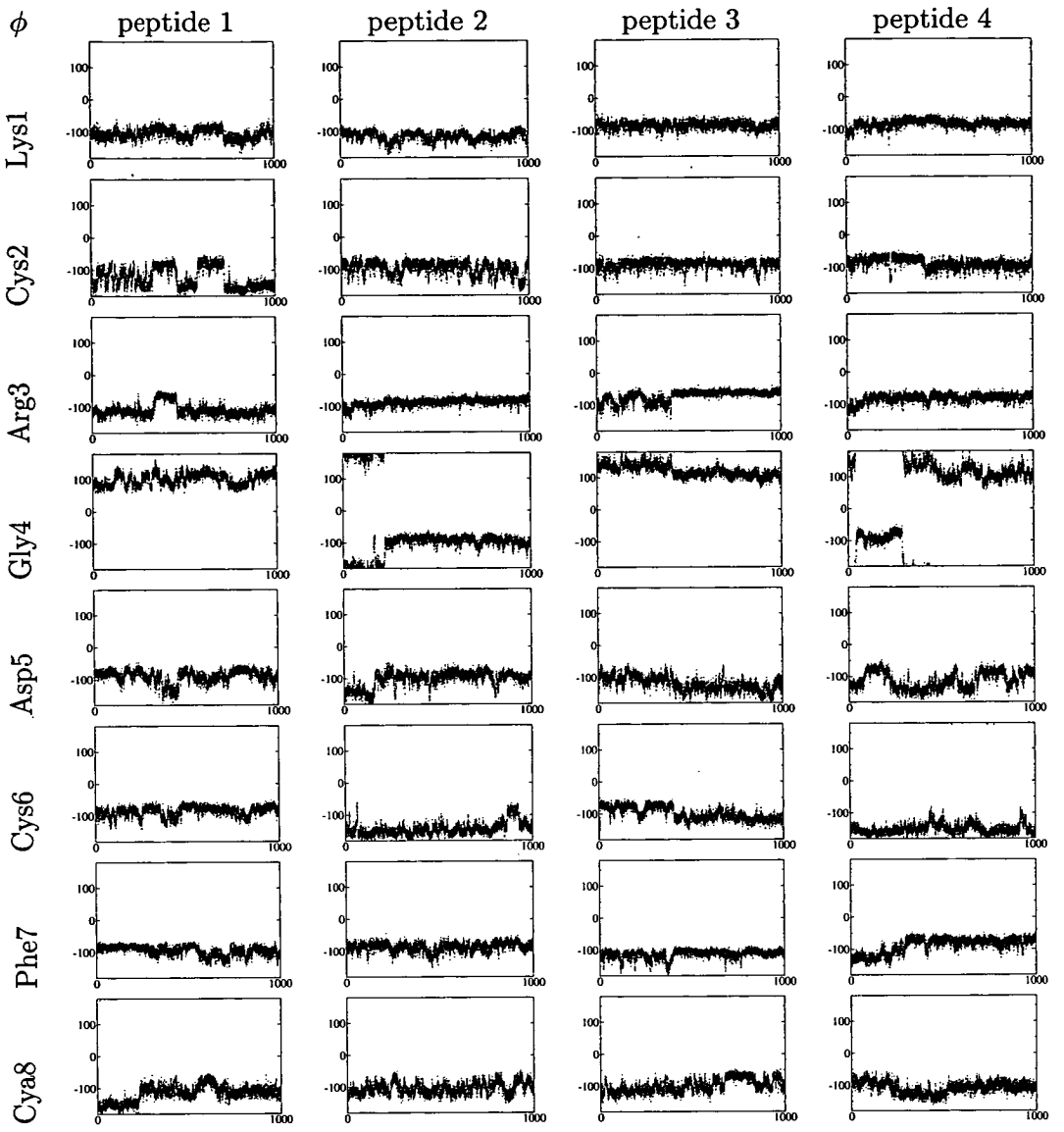


Figure 5.14: Dihedral angle variation of ϕ during free MD simulation with explicit water. The dihedral angle ϕ is plotted as a function of the simulation time (ps).

suggests that the long-range interactions influence the conformation of ϕ_5.

The rmsds were calculated for the backbone atoms (N,C^{α},C') of the CRGDC loop and also for all backbone atoms (Fig. 5.16). These plots shows that the rmsd of the peptide core varies in the same way as the loop does in peptides 2, 3 and 4. This contrasts with peptide 1 where the RGD loop seems to be less flexible than the rest of the molecule.

The fluctuations seen during the simulation in peptide 1 (0 to 200 ps ; 370 to 470 ps ; 590 to 720 ps ; 770 to 840 ps and 930 to 1000 ps) (Fig. 5.16) show that there

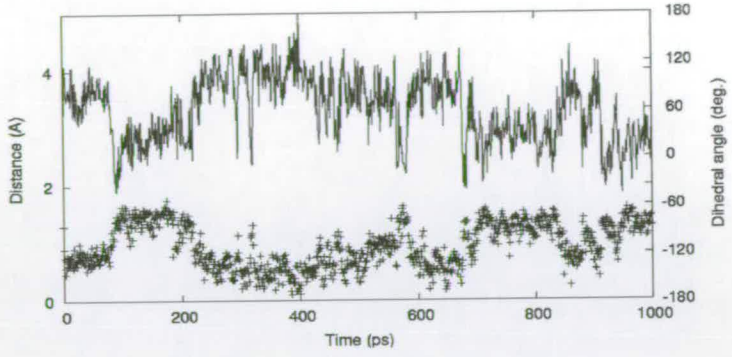


Figure 5.15: Correlation between the ϕ_5 (line) angle and the distance $6H^N \cdots 4CO$ (points) for peptide 4 in water.

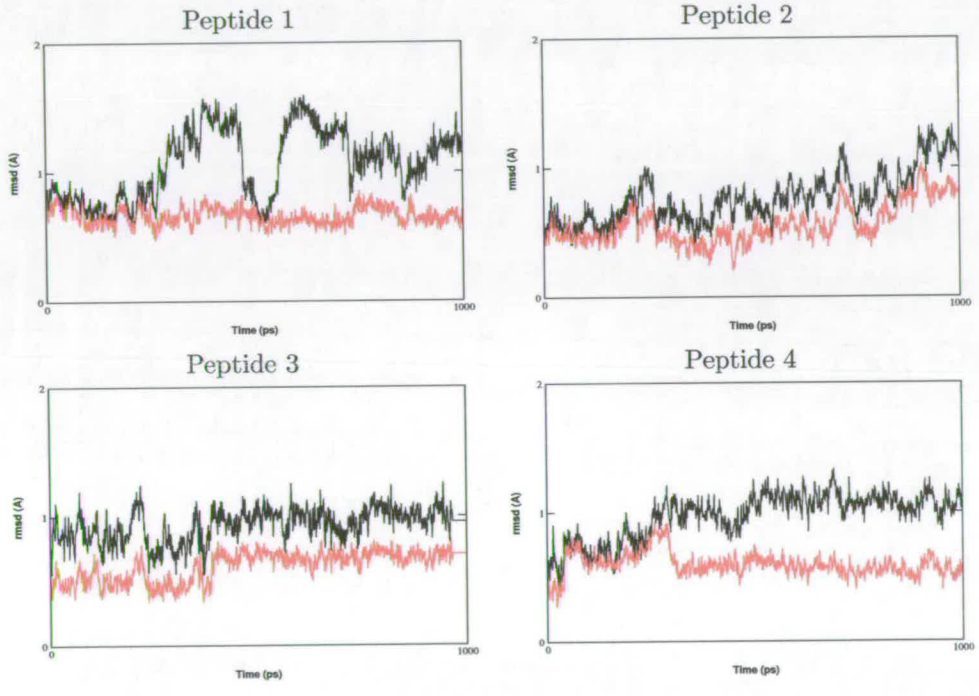


Figure 5.16: Backbone rmsd values with respect to a representative restrained structures. For comparison, the backbone rmsds of all residues (black) are plotted together with the backbone rmsds of the CRGDC loop (red).

is no direct correlation between the rmsds of the CRGDC loop and the conformation of the complete peptide. This indicates that peptide 1 can adopt several low energy conformations. According to the restrained structures of peptide 1 (see page 116), residues 7 and 8 can show two conformations. These are also observed during the free molecular dynamics. These conformational changes account for the large rmsd values of the complete peptide backbone observed from 370 ps and indicate that they occur

in the 12678 ring while the conformation of the CRGDC ring is largely stable. The starting structure can be characterized as having a closed conformation, stabilized by an electrostatic interaction between the C-terminal of Cya8 and the amide proton of Gly4.

No characteristic pattern emerged from the analysis of the simulations of peptide 2. The backbone rmsds do not seem to correlate with the fluctuations of any one particular angle. Unlike peptide 1, the conformation of the CRGDC loop fluctuates in the same manner as that of the complete peptide backbone (Fig. 5.16). The structures extracted from the trajectories confirm that the conformational changes mainly concern RDG residues and the mutual orientation of the 3-4 and 4-5 peptide planes. This rearrangement is due to rotations of the ϕ_4 and ϕ_5 angles.

The rmsds of the CRGDC loop and the time series of the dihedral angles show a stabilization of the peptide 3 after 410 ps. The analysis of the structures reveals that this coincides with the formation of two hydrogen bonds between $6H^N \cdots 2CO$ and $5H^N \cdots 2CO$ (Fig. 5.17) that were not seen in experimentally determined structures. The orientation of the NH bonds of Asp5 and Cys6 prevents the formation of the $6H^N \cdots 4CO$ bond observed in the experimental structures. As shown in figure 5.14 such stabilization is achieved by small but concerted variations of several dihedral angles (less than 30^o) after 410 ps.

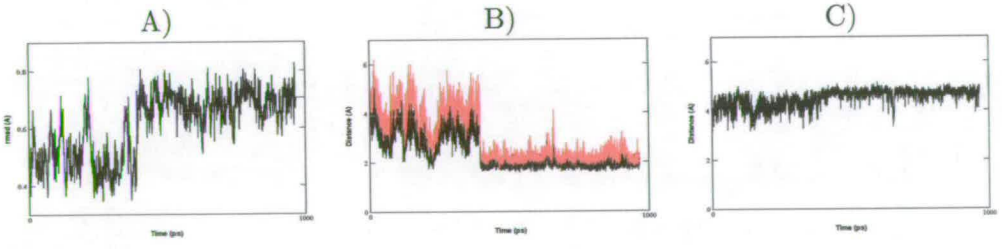


Figure 5.17: Comparison of rmsd of the CRGDC loop (A) and the distances $6H^N \cdots 2CO$ (black) and $5H^N \cdots 2CO$ (red) (B) and $6H^N \cdots 4CO$ hydrogen bond (C).

Unlike in peptide 3, no hydrogen bonds are seen throughout the simulation for peptide 4. This absence of hydrogen bond is accompanied by the flip of ϕ_4 angle by $+120^o$ and the rotation of ϕ_5 (-125→-75^o). After 330 ps, the rmsd of the CRGDC loop decreases, to 0.5 Å and is stable until the end of the 1 ns simulation although no hydrogen

bonds stabilize the structure from this point onwards (Fig. 5.16). The backbone rmsd increases on the other hand from 0.75 Å to 1 Å. This is particularly due to the rotation of ϕ_7 (-125→-65°).

Fluctuations of the ϕ angles indicate that the peptides are not rigid in solution. The backbone rmsds of the peptide core (up to 1.5 Å, over 1.0 Å for the CRGDC loop) indicate that peptide 1 behaves differently than the rest. As pointed above, this is mainly due to the fluctuations of the five member ring CFBKC of peptide 1 and can perhaps be attributed to the absence of substituents in this part of the molecule. The rmsds of peptides 2, 3 and 4 are in order of 1 Å and 0.75 Å for the backbone and the CRGDC loop respectively. These relatively small values indicate that the overall backbone conformation remains close to the starting structures. Despite starting the free molecular dynamics from very similar structures, the outcome of the simulations vary with the peptide concerned. The $6H^N \cdots 4CO$ bond observed in many NMR restrained structures appears to be too weak to conserve the conformation of the starting structures. The analysis of the trajectories show that once the coordinates of the restraint molecules are released, competing hydrogen bonds alter the original geometry of residues Arg3, Gly4 and Asp5. These hydrogen bonds are temporally formed during the free molecular dynamics to create γ-turns centered on Arg3, Gly4 and Asp5 (already observed in the NMR restrained structures).

Free molecular dynamics in DMSO

The time series of the ϕ angles (Fig. 5.18) show that the geometry of the backbone is much more conserved during the free molecular dynamics in DMSO than was the case in H_2O. The standard deviations of the ϕ angles do not exceed 24° (Cys6) and the values of the backbone rmsds are low (Fig. 5.19). These results indicate that the structures in DMSO are significantly more stable than the structures in water.

The most significant variations concern the ϕ_6 angle. This dihedral angle oscillates on the time scale of the simulation between two equally populated sites at -105° and -155°. The ϕ_3 dihedral angle also displays small fluctuations around 400 ps (-115 to -75°) but as its value settles at -115° for the rest of the simulation, this change cannot be correlated with the changes seen for ϕ_6 angle.

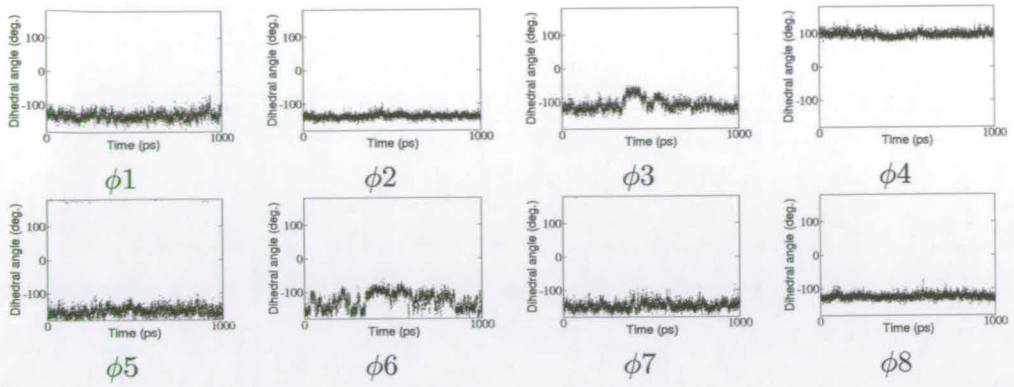


Figure 5.18: Dihedral angle variations during the free MD simulation with explicit DMSO.

The average rmsds of 0.34 Å and 0.77 Å were observed for the CRGDC loop and the entire backbone, respectively. These values indicate that the restrained structure is relatively rigid and in a low energy conformation. As already mentioned, only two angles rotate during the free molecular dynamics: ϕ_6 and ϕ_3. The rotation of the ϕ_6 angle does not correlate with backbone rmsd. This can be explained by the absence of a hydrogen bond between Cys6 and the backbone. On the contrary, an increase in rmsds (both the entire backbone and the CRGDC loop) at about 400 ps is accompanied by the formation of a hydrogen bond $5H^N \cdots 3CO$ and coincides with the rotation of ϕ_3 angle. This corresponds to the formation of a β-turn centered on the glycine residue which is however not seen in the experimental structures.

Residue.	ϕ (^{o})
1	-131±13
2	-137±7
3	-108±17
4	97±9
5	-151±18
6	-126±24
7	-146±13
8	-124±8

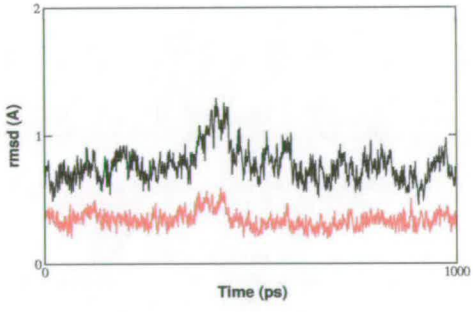


Figure 5.19: Average values and standard deviation of the ϕ dihedral angle in explicit DMSO for residues 1 to 8 (table). Backbone rmsd of the CRGDC loop (red) and those of peptide core (black) computed along the free molecular dynamics trajectory for 1 ns (graph).

The energy of the hydrogen bond (4 to 42 kJ/mol) is related to the distance between the

acceptor and the donor. These bonds maintain atoms close in space and stabilize the secondary structure. The hydrogen bonds seen in the NMR-restrained structures were followed over the simulated trajectory (Table 5.11). Three of these hydrogen bonds are well conserved during the simulation ($1H^N \cdots 7CO$, $7H^N \cdots 1CO$ and $6H^N \cdots 5O\delta$). However, the hydrogen bond $3H^N \cdots 5CO$ seen in 85% of the experimental structures disappears while the hydrogen bond $5H^N \cdots 3CO$ not observed in experimental structures is formed in 15 % of the structures during the free molecular dynamics. The absence of $3H^N \cdots 5CO$ and the formation of $5H^N \cdots 3CO$ indicates the only difference between the experimental (see Table 5.19) and the simulated structures, The changes of hydrogen bonds corresponds to the rotation of the ϕ_3 and ψ_2 angles which changes from an equatorial to an axial orientation with respect to the plane formed by the backbone.

Table 5.11: **Occurrence of hydrogen bonds seen in NMR restrained structures during unrestrained MD simulation of peptide 2 in DMSO:** The occurrence is calculated as the percentage of donor-acceptor distance less than 2.4 Å. The length of the hydrogen bond is time averaged over the entire time of the simulation

Donor	Acceptor	Occurrence (%)	Distance Å
$1H^N$	7C'	91	2.07
$7H^N$	1C'	89	2.10
$6H^N$	$O\delta_2$	96	1.93
$3H^N$	5C'	0	3.18

Compared to the peptides in water in general and P2w in particular, the stability of the restrained conformer is obvious. The rmsd values are significantly smaller in DMSO for both the backbone and the RGDC loop. Compared to the ϕ angle of the experimentally restrained structures, ϕ_3 is the only significant change which is presumably due to the formation of the hydrogen bond between $5H^N$ and 3CO as previously discussed.

5.5 Structure refinement protocols - an example of peptide 3

A detailed analysis of peptide 3 highlights the main structural changes occurring during the successive refinement steps. The mean structure was calculated using the command

"mean" in MOLMOL and the rmsds compared to the mean structure were computed with the "RMSD" command in MOLMOL. The lowest rmsd structure calculated using the backbone atoms (residues 2 to 8) was taken as the representative structure of the ensemble.

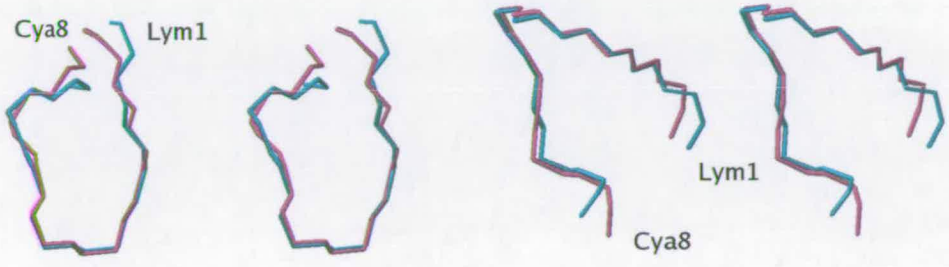


Figure 5.20: Representative peptide 3 structures calculated using protocols A (cyan) and B (magenta) overlaid using backbone atoms.

Examination of the dihedral angles of the representative structures of peptide 3 indicates that water refinement has a minor influence on the conformation of the backbone (Fig. 5.20). As reported by Linge *et. al* [91], the water refinement decreases the precision of the ensemble of structures. The backbone rmsds of peptide 3 are 0.13Å and 0.27Å for protocols A and B, respectively.

The structures calculated using protocol A were refined using the full relaxation matrix calculations of protocol C. Full relaxation matrix calculations use, as a input, the cross peak volumes extracted from a series of NOESY spectra. These are converted into inter proton distances which are iteratively adjusted during the course of the calculations. The criterion for convergence is the difference between the measured and calculated distances. By analogy with the R value commonly used in crystallography, a calculated R value measures the goodness of the fit. The R value is monitored during the full relaxation matrix calculations and is used to assess the progress of the refinement. A major advantage of the full relaxation matrix approach over the single mixing time calculations is that it accounts for the effects of spin diffusion and can therefore more accurately relate the cross peak volumes to the inter proton distances. This requires the spectra acquired at different mixing times to be processed identically in order to generate correct NOE build-up curves. Overall, this reduces somewhat the number of cross peaks that can be used, as, for example, overlapping cross peaks or flexible parts of

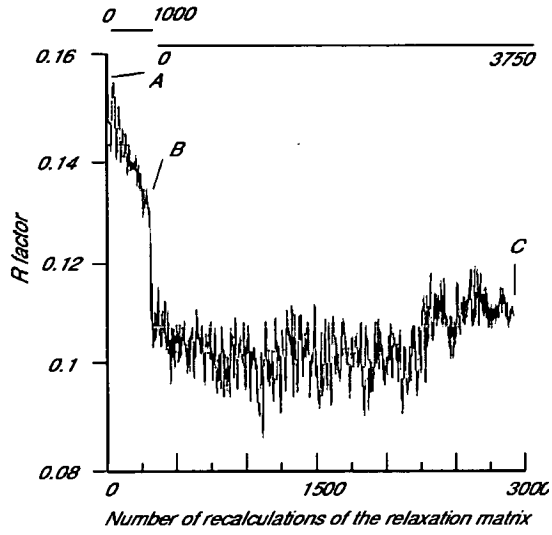


Figure 5.21: Evolution of the R-factor during the full relaxation matrix refinement. The first sequence (A to B) corresponds to the refinement in vacuum. The relaxation matrix is recalculated 301 times for 1000 steps of MD and 90 steps of minimization. The final refinement (B to C) of the structures is performed in water (B to C) using 3750 steps of MD and 300 steps of minimization. Overall the relaxation matrix is recalculated 2620 times.

the molecule with different tumbling regime. This contrasts with the single mixing time spectra where the processing was optimized (e.g. window function, baseline correction, noise level, individual integration) to maximize the number of cross peaks. For the full relaxation matrix calculations, a tolerance[3] was set to 0.04 Å and the relaxation matrix was updated once every 3 steps during the MD. The initial "R" value decreased during the course of the refinement (Fig. 5.21) and most of the conformational search is achieved by the molecular dynamics and simulated annealing (MDSA).

For peptide 3 the R value of the starting structure and the final structure were 0.149, and 0.126 respectively. The small difference between these two values shows that the structures calculated from a single set of distance restraints were already close to the refined structures. The overlay of the backbones of peptide 3 calculated with protocols A and C (Fig. 5.22) reflects this and shows that similar overall shape of the peptides is conserved. The differences are mainly due to rotations of the dihedral angles ϕ_2 (from -129 to -163°) and ψ_1 from -48 to -3°).

[3] The relaxation matrix is updated is the atoms move more than the distance tolerance. The distance set by the tolerance compares the calculated position and those at the last update of the relaxation matrix

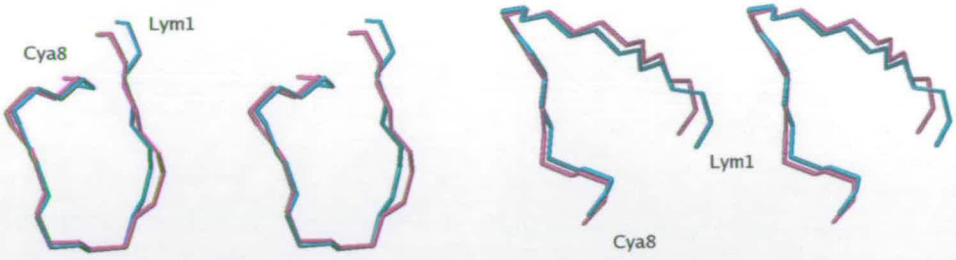


Figure 5.22: Representative peptide 3 structures calculated using protocols A (cyan) and C (magenta) overlaid using backbone atoms.

Representative structures calculated using the protocols A and C are shown in Fig. 5.23.

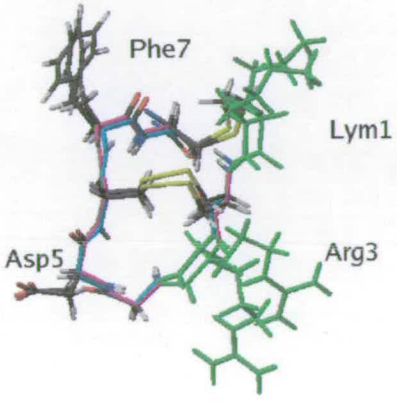


Figure 5.23: Comparison of the representative structures of peptide 3 calculated using protocol A (cyan) and C (magenta). Residues which underwent the largest conformational change are shown in green.

Similarly to protocol B, the conformational search in explicit solvent using the FRM (protocol D) is achieved at relatively low temperature while the empirical parameters of the force field (angles and impropers) are weakened and progressively brought back to the values defined in the force field. The values of the R factor follow this annealing procedure.

Starting from the FRM refined structures in vacuum, the R factor falls from 0.126 to the final value of 0.111 and the final structure (0.110) is similar to the starting one as shown in Fig. 5.21.

5.6 A comparison of structures calculated in vacuum and water

In the following, the structures calculated using the FRM in vacuum (protocol C) and water (protocol D), with exception of peptide 1 calculated in vacuum with protocol A and in water with protocol B, are compared. The conformations of the representative structures are showed in Fig. 5.26 for the structures calculated in vacuum and in water.

Table 5.12 shows the distance between the carbonyl of residue 4 to the amide proton of residue 6 where a hydrogen bond was found in several structures.

Table 5.12: Distances between $6H^N$ and 4CO in NMR restrained structures calculated in vacuum and water (Å).

Structure	P1w	P2w	P3w	P4w
Vacuum	2.68	3.04	2.99	3.11
Water	1.84	2.53	2.25	2.10

Although the conformational changes can best be visualized using a single, closest to the mean structure, analysis of an ensemble of structures (typically 20) is equally important in order to evaluate their convergence of structures. For peptide 1, water refined structures converged to two conformations that differed, in particular, in ψ_7 and ϕ_8 dihedral angles as indicated by their large standard deviations (Table 5.13). This is also the reason why these dihedral angles are so different from those of other peptides. During the refinement in water, changes of $<30^o$ occur for the ϕ_2, ϕ_3 and ψ_5 angles. These however, did not dramatically change the conformation of the backbone but rather increased the frequency of the occurrence of the hydrogen bond $6H^N \cdots 4CO$ (50% in vacuum vs. 95% in water).

The conformation of peptide 2 is essentially the same for the structures calculated in vacuum and in water although the precision of the latter set of structures decreased. This can be seen from the increased rmsd of the backbone atoms (Tables 5.7 and 5.9) and those of dihedral angles (Table 5.13). The only systematic change occured for ψ_8 which rotated from 70 to 105^o albeit without modifying the conformation of the peptide

5.6 A comparison of structures calculated in vacuum and water

In the following, the structures calculated using the FRM in vacuum (protocol C) and water (protocol D), with exception of peptide 1 calculated in vacuum with protocol A and in water with protocol B, are compared. The conformations of the representative structures are showed in Fig. 5.26 for the structures calculated in vacuum and in water.

Table 5.12 shows the distance between the carbonyl of residue 4 to the amide proton of residue 6 where a hydrogen bond was found in several structures.

Table 5.12: Distances between $6H^N$ and 4CO in NMR restrained structures calculated in vacuum and water (Å).

Structure	P1w	P2w	P3w	P4w
Vacuum	2.68	3.04	2.99	3.11
Water	1.84	2.53	2.25	2.10

Although the conformational changes can best be visualized using a single, closest to the mean structure, analysis of an ensemble of structures (typically 20) is equally important in order to evaluate their convergence of structures. For peptide 1, water refined structures converged to two conformations that differed, in particular, in ψ_7 and ϕ_8 dihedral angles as indicated by their large standard deviations (Table 5.13). This is also the reason why these dihedral angles are so different from those of other peptides. During the refinement in water, changes of $<30^o$ occur for the ϕ_2, ϕ_3 and ψ_5 angles. These however, did not dramatically change the conformation of the backbone but rather increased the frequency of the occurrence of the hydrogen bond $6H^N \cdots 4CO$ (50% in vacuum vs. 95% in water).

The conformation of peptide 2 is essentially the same for the structures calculated in vacuum and in water although the precision of the latter set of structures decreased. This can be seen from the increased rmsd of the backbone atoms (Tables 5.7 and 5.9) and those of dihedral angles (Table 5.13). The only systematic change occured for ψ_8 which rotated from 70 to 105^o albeit without modifying the conformation of the peptide

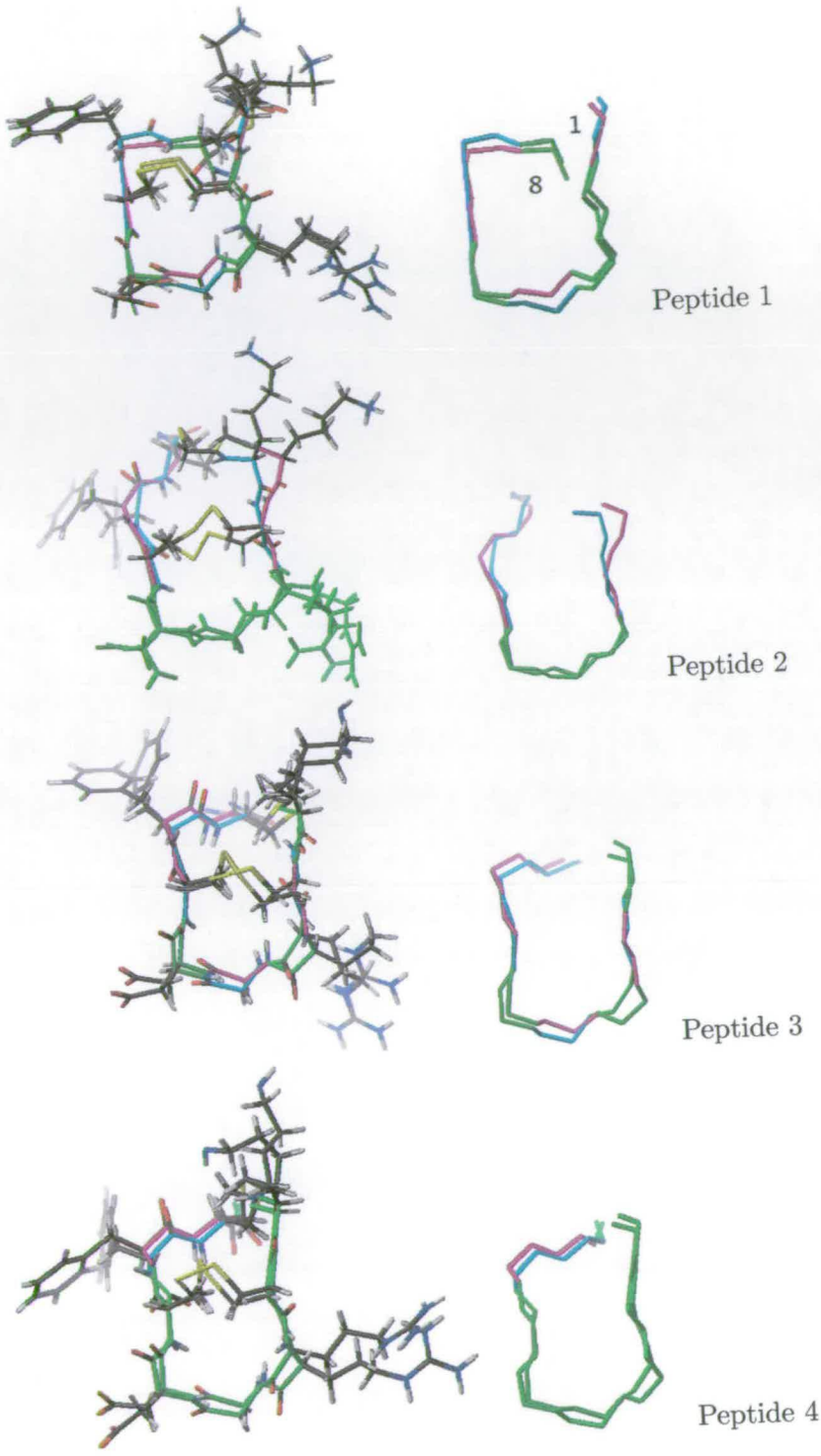


Figure 5.26: Vacuum (cyan) and water (magenta) refined structures of cyclic peptides 1-4. Significant differences between the two sets of structures are indicated in green.

core. Similarly to peptide 1, the frequency of $6H^N \cdots 4CO$ hydrogen bond (30% in water vs 20% in vacuum) accompanied by shortening of the interatomic distance (2.53

vs 3.05Å) occured.

The modification occurring in peptide 3 were discussed in section 5.5.

In peptide 4 the largest rotations are observed for angles ϕ_2, ψ_2, ψ_3 and ψ_8 (table 5.13). During the water refinement, the ψ_6 angle converges to 160^o for one of the two conformations (130^o and 160^o) present in vacuum. Also in this peptide, the refinement in explicit solvent brought the $6H^N$ and 4CO atoms closer in space (3.11 to 2.10 Å for the closest-to-mean structures) and favored the formation of a hydrogen bond (increase from 5% to 35% of the population of the NMR structures). Once again this distance shortening was accompanied by small changes in average backbone dihedral angles of Gly4 (117, -78 vs. 104, -70) and Asp5 (-115, 49 vs. -101, 68).

The agreement between the calculated structures and the dihedral restraints was improved in water refined structures as can be seen from the decreased dihedral energy in water and the formation of hydrogen bonds in between $6H^N$ and 4CO in most of the structures. This situation is illustrated by the ϕ_5 angle of peptide 4 (Ramachandran plots in Fig. 5.9)

Table 5.13: Summary of the backbone dihedral angles of peptides 1 to 4 calculated by the FRM in vacuum (protocol C) and in water (protocol D). The mean values and the standard deviations are calculated for each residue.

Residues	P1w		P2w		P3w		P4w	
			ϕ dihedral angles					
	vac.	wat.	vac.	wat.	vac.	wat.	vac.	wat.
1LYS or 1LYM	-95±2	-92±4	-120±2	-105±15	-73±8	-110±13	-119±2	-117±7
2CYS	-136±15	-151±6	-165±9	-147±15	-152 ± 11	-150 ± 7	-175±3	-146±9
3ARG	-160±6	-138±15	-119±2	-112±10	-119±0	-115±4	-118±3	-117±3
4GLY	115±14	115 ± 20	169±12	151±30	126±13	125±27	117±6	104±18
5ASP	-85±12	-85±4	-108±18	-113±18	-113±2	-105±22	-115±6	-101±12
6CYS	-167± 7	-171 ±10	-149±8	-144±18	-144±6	-160±7	-140±4	-150±10
7PHE	-70±11	-79±11	-125±8	-110±14	-93±9	-94±13	-90±15	-109±19
8CYA	-73±9	-97±39	-116 ±20	-117±20	-125 ±8	-112 ±18	-128±5	-133±10
			ψ dihedral angles					
1LYS or 1LYM	39±4	24±8	69±5	63±10	-16±13	36±15	48±5	59±8
2CYS	-164±6	172±4	158±8	166±10	160±4	169±10	141±6	162±10
3ARG	80±17	88±18	89±12	89±23	124±12	78±15	134±4	107±18
4GLY	-85±13	-67±12	-97±12	-101±21	-82±3	-78±14	-79±7	-69±7
5ASP	73±12	50±12	60±6	75±21	47±3	67±7	49±3	68±6
6CYS	170±8	172±4	154±5	158±8	162±9	166±7	144±17	160±11
7PHE	-15.4±7	12±46	110±15	102±20	90±6	89±15	86±6	72±8
8CYA	-83±8	-116±5	70±14	105±29	46±25	92±41	95±14	137±18

Consequences of water refinement

The structures of peptides 2, 3 and 4 obtained by a full relaxation matrix protocol in explicit water refinement adopt similar conformations, even more so than the full relaxation matrix structure in vacuum. The backbone rmsd for the closest-to-mean structures overlaid on residues 2 to 8, dropped from 0.45 to 0.36 Å in water refined structures. There is good experimental evidence to show that the conformations of these compounds are indeed similar: structure calculations carried out in explicit solvent indicate the presence of a hydrogen bond between $6H^N$ and 4CO in many water refined structures (Table 5.12). These are supported by low temperature coefficients observed for this H^N in all compounds. The analysis of chemical shifts presented in section 5.8.1 supports the notion of a similar conformation between the studied peptides.

5.7 A comparison of the final structures of peptides 1-4 in water

The structures refined in an explicit solvent are more realistic than their vacuum counterparts. As mentioned above, the peptide cores, after final refinement in water, tend to converge toward a similar conformation. This is confirmed by the analysis of dihedral angles from the ensemble of structures (Table 5.13), which show no significant differences for the majority of angles (Fig. 5.27). The only considerable difference concerns the angles ψ_7 and ϕ_8 of peptide 1. Figure 5.27 shows the dispersion of the dihedral angles for peptides 1 to 4.

Due to nonstandard structural elements, residues 7 and 8 are not recognized as peptides by MOLMOL and are therefore not presented in Fig. 5.27. As mentioned above, dihedral angles ψ_7 and ψ_8 can take two conformations in peptide 1 at (-90±10 , 76±15) and (-143±30 , -115±4) or (-73±7 , -16±7) and (-73±10 , -116±5). The two sets of structures differ in the orientation of the amide group in residue 8. The values of the dihedral angle in the first set of structures are very similar to those calculated for the other peptides. In these the carbonyl group of residue 8 points out of the peptide ring. This contrasts with the second set of structures where this group points inwards and

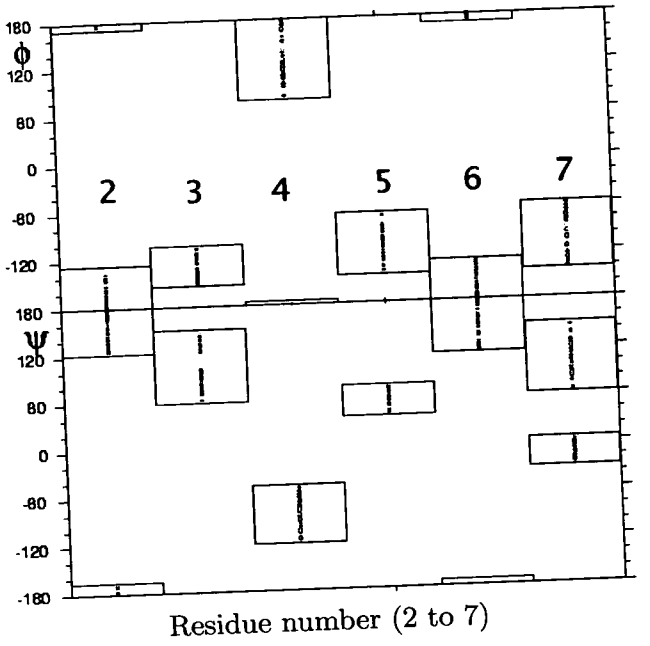


Figure 5.27: Dihedral angles of the ensemble of structures calculated using protocol D (or B for peptide 1) versus residue number in peptide 1 (black), 2 (red), 3 (grey) and 4 (blue). The ϕ and ψ angle are plotted in the upper and lower parts of the figure, respectively.

is close in space to the amide proton of Gly4.

Particular attention is paid to the angle ϕ_3, as this is always pushed to the edge of the allowed region when the dihedral restraints are used. This results in a very small standard deviation of this angle. Based on the values of the coupling constants (Table 4.6), the ϕ_3 angle was restrained to -90±30^o for the peptides 2, 3 and 4 with the resulting structures showing values close to -120^o (this angle was not restrained for peptide 1 because of the lack of value of the coupling constant $^3J_{H^NC^\beta}$). Calculations performed without dihedral restraints show that the ϕ_3 converges to values between -110^o and -160^o. The mean values for peptides 2, 3 and 4 were -116^o, -125^o, -131^o respectively which are close to those calculated with the dihedral angle restraints. Despite such small differences, the fact that the angles are pushed towards the edge of the allowed region indicates that the distance restraints are not compatible with the dihedral restraints which was interpreted as a sign of flexibility.

Figure 5.28 shows an ensemble of the water refined structures for peptides 1 and 2 to 4 (see 5.3). Two different views for each peptide illustrate the fact that the structures

share very similar conformations.

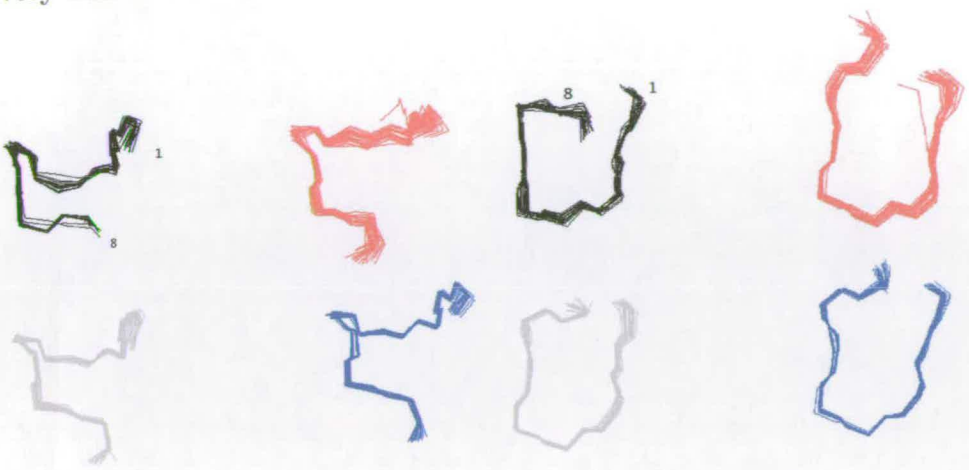


Figure 5.28: Peptide backbones overlaid for an ensemble of 20 lowest-energy refined structures P1w (black), P2w (red), P3w (grey) and P4w (blue) structures.

The closest-to-mean structures from the ensembles of structures shown in Fig. 5.28 are overlaid in Fig. 5.29. Figure 5.30 shows the structures including side chains.

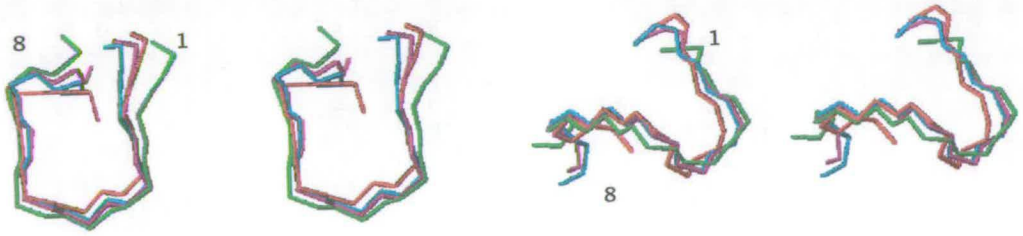


Figure 5.29: Stereo-views of the backbone of the representative structures of peptides 1 (red), 2 (green), 3 (cyan) and 4 (magenta).

Conformation of side chains was analyzed using χ^1 angles. A statistical approach was used to determine the population of side chain rotamers in peptides using the Pachler's equations as explained in section 4.4 [106, 108, 110, 111, 121]. In the peptides in water the H^β protons of residues 2, 7 and 8 are strongly coupled or degenerate - which by itself is a sign of conformational averaging. The coupling constant extracted from residues 1, 3, 5 and 6 and the intensity of the NOE cross peaks indicate that conformational averaging takes place also for these residues. Experimental data thus indicate that the

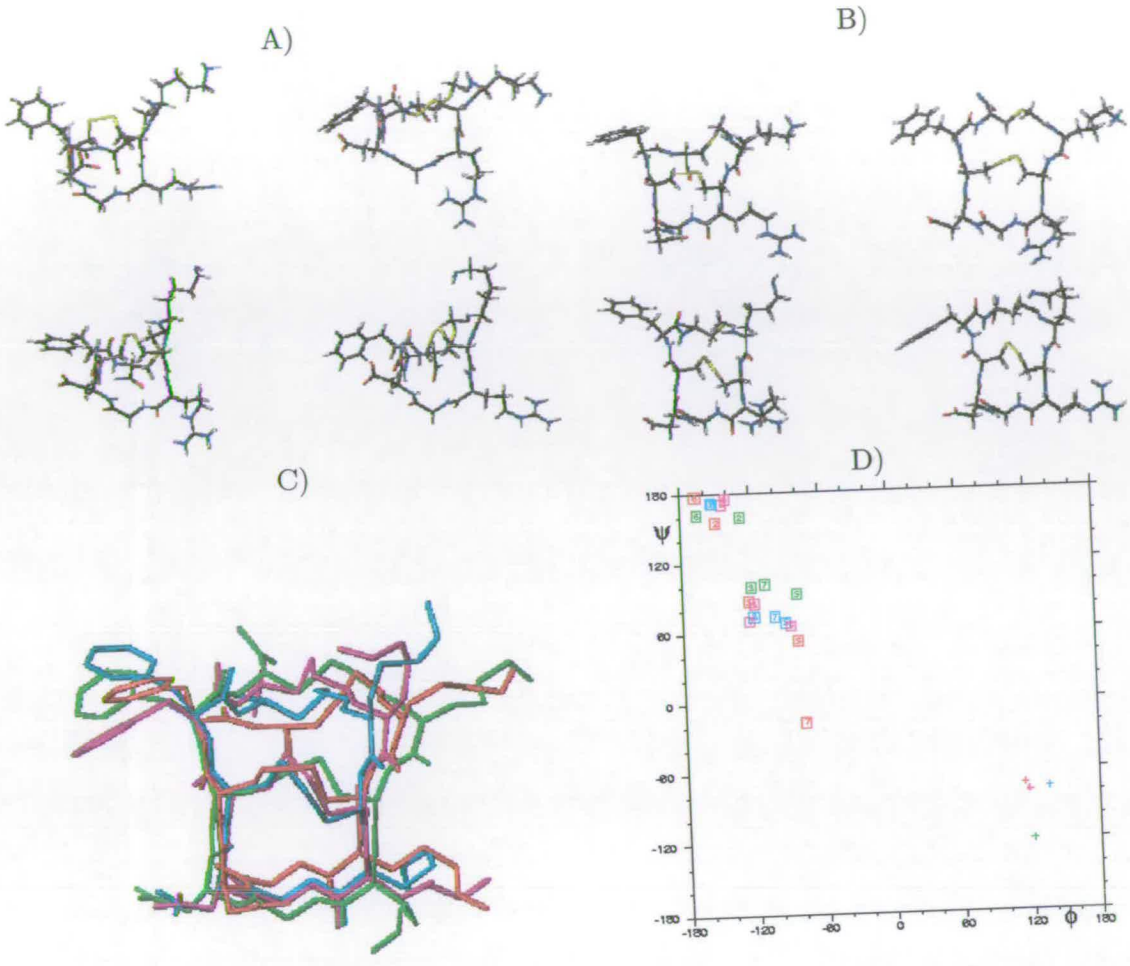


Figure 5.30: **Comparison of the structures also showing side chains for peptides:** 1 (red), 2 (green), 3 (cyan) and 4 (magenta). A and B show two different views, C an overlay of peptides 1-4; D Ramachandran plot for the residues 2 to 7 of the structures shown in C.

side chains of the peptides are not rigid, and are involved in conformational exchange.

This also applies to the cysteine side chains with a consequence that no preferred conformation could be established for the disulfide bridge. The H^{β} protons of Cys2 residue are degenerate or strongly coupled, therefore no coupling constant could be extracted for the Cys2 χ^1 angle. The coupling constants extracted for Cys6 χ^1 converged towards a dihedral angle of 120^o which is impossible considering the geometry of the side chain. Together, these data suggest that the disulfide bridge undergoes a conformational averaging in all the peptides in water.

The Ramachandran plot given in Fig. 5.30D illustrates the similarities between the

conformations of peptides 1-4. These similarities must be supported by the experimental data that were not used to restrain the structures: temperature coefficients and chemical shifts. These will be inspected next.

5.8 Consistency of the calculated structures with NMR data

Unlike proteins, where the overall shape of the backbone is determined by long-range interactions, the conformation of small peptides is driven by short range interactions. The structures of studied peptides were solved by converting NOEs and vicinal coupling constants to geometrical restraints and using molecular dynamics calculations. It was observed that omitting or adding restraints could dramatically change the shape of the molecules without creating violations. Therefore, it is essential that the structures obtained are cross-validated against a different type of data than that used to derive these structures and also additional experimental data that were not used during the process. Here we look first at the relationship between the backbone dihedral and the cross peaks.

It can be observed that in the peptide structures, ϕ dihedral angles are restricted to the region (0 , -180^o) and ψ falls in the region between 0 and +180^o (apart from the glycine and ψ_7 which adopts two conformations). These conformations are supported by the cross peak intensities observed in the 2D NOESY and ROESY spectra for intra- and sequential cross peaks. The comparison of the distances (related to the intensities) $d_{n\alpha}(i,i)$ and $d_{n\alpha}(i,i+1)$ shows that $d_{n\alpha}(i,i+1)$ is systematically shorter than $d_{n\alpha}(i,i)$. Meanwhile, due to the geometry of amino acids, the distance $d_{n\alpha}$(i,i) can oscillate between 2.8 Å and 2.2 Å [79] which means that $d_{n\alpha}$(i,i+1) must be shorter than 2.8 Å. This distance corresponds to a positive value for the ψ angle. Now, considering that the distance $d_{n\alpha}$(i,i+1) in the favorable positive regions for ψ varies between 2.2 and 2.6 Å, the distance $d_{n\alpha}$(i,i) must be larger than 2.6 Å. Alongside with the weak $d_{nn}(i,i+1)$, these observations match with a combination of $\phi-\psi$ angles of the β-sheet region of the Ramachandran plot and confirm the results of the calculated structures.

All of the bicyclic peptides share an identical peptide core and are structurally homo-

geneous. The suggested homogeneity of structures can be validated by comparing the experimental data such as chemical shifts, temperature coefficients and amplitude of the coupling constants between the peptides.

5.8.1 Chemical shifts

The chemical shifts are sensitive to the chemical environment and the conformation of the molecules and are a good indicator of structural homogeneity/differences. In this work the values of the chemical shifts were compared in order to assess structural homogeneity between the peptide cores [41]. For peptides 1-4 a comparison of backbone chemical shifts is presented in Fig. 5.31. The similarity in ^{15}N , C' and H^N chemical shifts for residues 2 to 6 indicates structural homogeneity of the RGD motif of these peptides. Larger differences were only observed for residues 1, 7 an 8 between peptide 1 and the other peptides. These differences are likely a result of two contributions. Differences in the primary structure (attachment of tails to residues 1 and 8 and) and conformational differences where residues 7, 8 and CLL link are arranged differently in peptide 1.

5.8.2 Hydrogen bonds

Hydrogen bonds were calculated in MOLMOL using ensembles of structures and minimum acceptor-donor distance of 2.4 Å and maximal (O–H-N) angle of 35^o.

The hydrogen bond $6H^N \cdots 4CO$ was observed in many water refined structures (Table 5.15). The presence of hydrogen bonds in peptides often correlates with low temperature dependence of the donors, the amide protons. A temperature coefficient higher than -4.5 ppb.K^{-1} has been suggested to indicate H-bonded protons [122, 123]. According to the temperature coefficients of the peptides (Table 4.8), it appears that $6H^N$ is the only possible donor. Its temperature dependance (-3.5 ppb.K^{-1}) is lower than that determined for the other residues in all peptides where values between -5.0 and -12.5 ppb.K^{-1} were obtained. Large temperature coefficients of other amide protons in all peptides indicate that these protons cannot be involved in hydrogen bonds. However, cyclic peptides are generally less structured than globular proteins, and the temperature dependent changes of H^N chemical shifts - from which the temperature

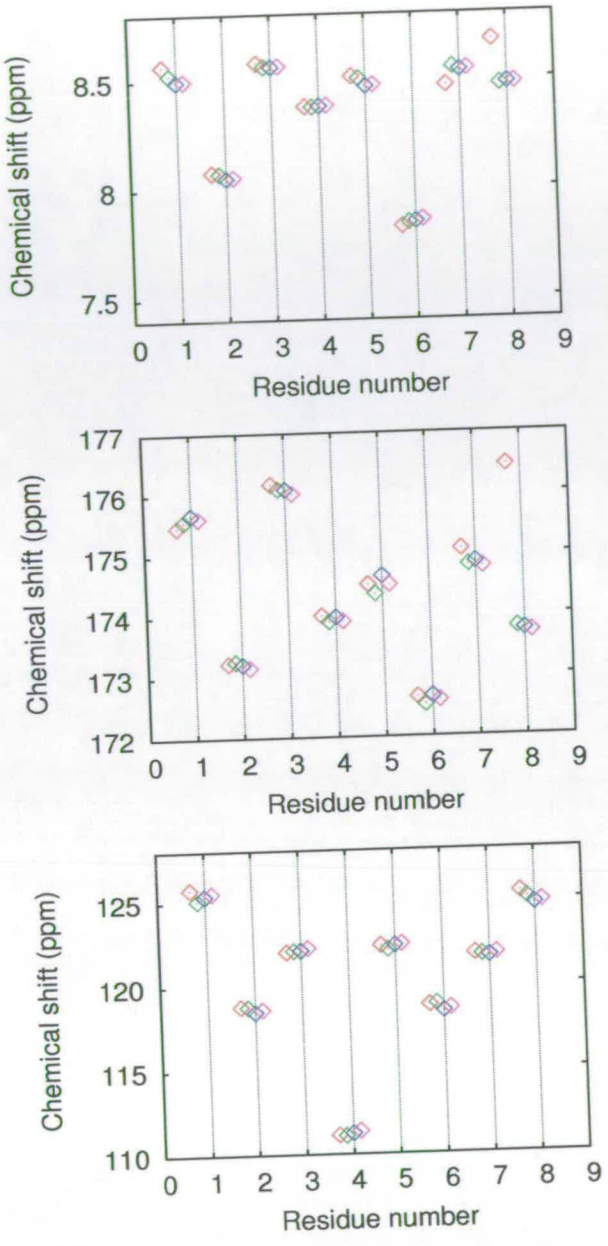


Figure 5.31: Comparison of the H^N , C' and N (top to the bottom) chemical shifts for peptides 1 (red), 2 (green), 3 (blue) and 4 (magenta).

coefficients are derived - must be interpreted carefully [27, 121, 124, 125]. Depending on the exchange regime, conformation averaging can also cause temperature dependent movement of chemical shifts. It has also been shown that the ring current of aromatic amino acids can lower the temperature dependence of protons in the neighboring residues [55]. In all peptides, residues Cys6 and Phe7 are sequential and the ring current of the latter could influence the values of the temperature coefficients of

$6H^N$ protons. However, the distance between the phenyl ring and the $6H^N$ proton exceeds 7 Å in all peptides and the ring current would therefore be observed preferably on $7H^N$ or $8H^N$ protons which are closer to the phenyl ring than on the $6H^N$ proton.

The presence of a hydrogen bond $6H^N \cdots 4CO$, repeatedly found in all peptides is consistent with the temperature dependence of $6H^N$ protons. However, considering a limit at -4.5 ppb.K^{-1}, the values of the temperature coefficient indicate that this is a weak or transiently formed hydrogen bond. Another supporting evidence for the existence of this bond comes from a relatively weak water-$6H^N$ cross peak in the ROESY spectra indicating that this amide proton is partially protected from the solvent. Finally, the line width of the $6H^N$ protons decreases with higher temperature for all compounds. This correlates with slower quadrupolar relaxation of ^{14}N at higher temperature causing decreased line broadening. Therefore this amide proton is in slow exchange with the solvent and small temperature changes of its chemical shift can safely be related to the existence of a hydrogen bond. As illustrated in figure 5.32, this contrasts with the amide proton $1H^N$ which shows an increase in linewidths at higher temperature, indicating faster exchange with water for all peptides.

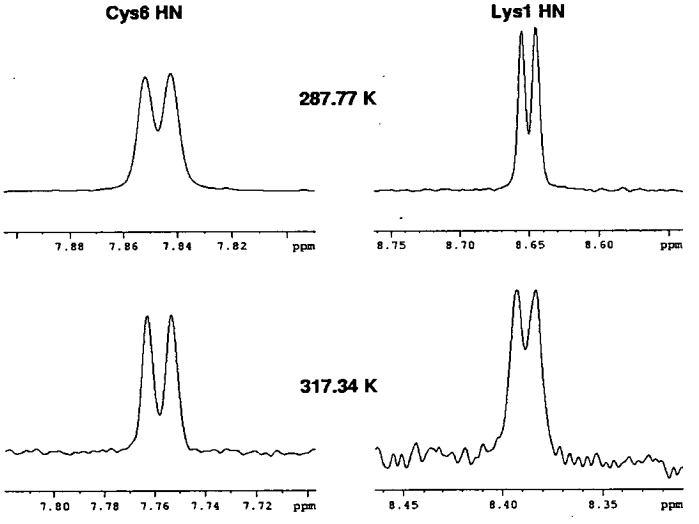


Figure 5.32: Amide protons $6H^N$ and $1H^N$ in peptide 1 acquired at 287.77K and 317.34K illustrate the opposite trends in line broadening of the amide protons at different temperatures.

The temperature coefficients of the other residues in all compounds (smaller than -5.0 ppb.K^{-1}) indicate that no stable intra molecular hydrogen bonds are formed in the peptides. The temperature coefficients can also be compared to the "random coil" temperature coefficients calculated in water at pH 5 [55] (Table 5.14).

Table 5.14: **Comparison of experimental and random coil temperature coefficients ppb.K^{-1}:** The random coil temperature coefficients were measure at pH 5 in water. The pH of the peptide solution is described in chapter 3.

Amino acid	Random coil value	P1w	P2w	P3w	P4w
Lys1	-7.9	-9.2	-7.8	-7.8	-8.1
Cys2	-7.4	-5.2	-5.9	-5.0	-5.2
Arg3	-7.7	-7.5	-7.6	-7.6	-7.7
Gly4	-7.0	-9.3	-8.6	-8.3	-8.2
Asp5	-6.4	-8.8	-7.8	-8.0	-8.2
Cys6	-7.4	-2.7	-3.5	-3.1	-3.5
Phe7	-8.1	-5.1	-7.1	-6.3	-6.4
Cya[a]	-8.0	-12.5	-8.6	-8.9	-9.0

[a] Without values for unusual residues, Cya was compared to the random coil value of Met

Experimental temperatures coefficients are similar to those of the random coil values for residues 1 and 3, lower for residues 4, 5 and 8 and higher for residues 2, 6 and 7 (in P1w only). The temperature coefficients smaller than the random coil values can be accounted for by fully exposed protons or conformational averaging. The relatively high values of Cy2 amide temperature coefficients can be accounted for by partially populated turns stabilized by a hydrogen bond. This temperature coefficient could be explained by transitory H-bond $2H^N \cdots 8C\zeta$ observed in some structures (Table 5.15).

Table 5.15: Population of hydrogen bonds (in %) observed in peptides 1-4. The population with 20 FRM water refined structures of water refined structures.

Donor	Acceptor	Peptide 1	Peptide 2	Peptide 3	Peptide 4
$6H^N$	4C'	95	30	65	35
$2H^N$	$8C^\zeta$	15	30	0	5

The temperature coefficient of residue 7 in P1w can be explained by the random orientation of the ring in P1w which decreases the temperature dependence of $7H^N$. In

other peptides, this orientation is not entirely random as NOE cross peaks are observed between Ncc9 and the protons of the phenyl ring, therefore orientating the phenyl ring towards Ncc9 and Cya8.

5.8.3 Structures based on T-ROESY spectra

T-ROESY

Test structure calculations were also carried out with the data extracted from ROESY and T-ROESY experiments. Calculations performed using ROESY data showed distance violations involving side chains, presumably because of the TOCSY artifacts. The advantage of T-ROESY experiments compared to ROESY experiments is that they effectively remove the off-resonance effects and the TOCSY transfers, yielding more reliable intensities, albeit with some reduction in cross peak intensities. It was noticed that the HN HN region did not give rise to as many cross peaks as seen in NOESY spectra. In T-ROESY spectra, it was possible to calibrate the distances using the isolated spin-pair approximation using the methylene proton cross peak of the glycine. The distances were calibrated and given an error of ±20%.

Nevertheless, the data obtained from T-ROESY spectra yielded structures similar to those calculated using one mixing time NOESY spectrum and the set of distances given in Table 5.4. SInce the full relaxation matrix was used to calculate the final structures for peptides 2-4 and since the NOESY spectrum of peptide 1 at 900 MHz yielded accurate structures, T-ROESY experiments were not used for structure determination of all peptides.

Structures of peptide 2 calculated based on cross peak intensities extracted from 200 ms mixing time T-ROESY spectra can be compared with those calculated using the full relaxation matrix approach and a series of NOESY spectra. These structures are referred to in this chapter as P2tr and P2frm. Structural statistics of P2tr are given in Table 5.16.

Both ensembles of structures show similar rmsds of 0.35 (C2tr) and 0.37 Å (C2frm) for 20 best structures with similar backbone conformation for almost all residues (Table 5.17).

Table 5.16: Structure statistics for 20 structures of peptide 2 calculated with T-ROESY restraints and refined in explicit solvent

Potential energies	DR
Etot	-4192.03 ± 237.03
Ebond	136.80 ± 7.88
Eangle	34.22 ± 5.65
Eimpro	5.25 ± 2.04
Evdw	528.54 ± 36.48
Enoe	19.36 ± 12.00
Ecdih	0.04 ± 0.08
Distance restraints used in calculations	98
Intra., seq., med., adr.	58, 21, 8, 0
Violation > 0.3Å	10
Dihedral restraints used in calculations	9
Violations > 5^o	0
Pairwise rmsd (Å)	2-8
backbone	0.35 ± 0.13
non-hydrogen	1.29 ± 0.33

Table 5.17: Comparison of dihedral angles for structures calculated with FRM and T-ROESY data.

Residues	T-ROESY	FRM
ϕ dihedral angles		
Lys1	-87 ± 5	-105 ± 15
Cys2	-159 ± 14	-147± 15
Arg3	-111 ± 8	-112 ± 10
Gly4	156 ± 17	151± 30
Asp5	-92 ± 8	-113± 18
Cys6	-135 ± 9	-144± 18
Phe7	-99 ± 9	-110± 14
Cya8	-118 ± 19	-117 ± 20
ψ dihedral angles		
Lys1	10 ± 5	63 ± 10
Cys2	140 ± 10	166 ± 10
Arg3	102 ± 16	89± 23
Gly4	-104 ± 12	-101± 21
Asp5	41 ± 12	75 ± 21
Cys6	167 ± 7	158 ± 8
Phe7	91 ± 14	102± 20
Cya8	60 ± 35	105 ± 29

Significant differences were only observed for ψ_1 and ψ_8 dihedral angles. The compared distances also show that P2tr is somehow more compact that P2frm. This is indicated by systematically shorter distances $d(C^{\alpha}C^{\alpha})$ in P2tr compared to P2frm (Table 5.18). Yet, these distance differences do not have dramatic effect apart from the largest effect

Table 5.18: Comparison of distances of peptide 2 calculated with the full relaxation matrix approach and T-ROESY data.

Residue1	Atom1	Residue2	Atom2	d_{TR} (Å)	d_{FRM} (Å)	d_{TR}-d_{FRM} (Å)
Lys1	C^{α}	Cys2	C^{α}	3.8	3.8	0
Lys1	C^{α}	Arg3	C^{α}	6.9	7.0	-0.1
Lys1	C^{α}	Gly4	C^{α}	9.6	9.8	-0.2
Lys1	C^{α}	Asp5	C^{α}	10.6	11.0	-0.4
Lys1	C^{α}	Cys6	C^{α}	8.0	8.8	-0.8
Lys1	C^{α}	Phe7	C^{α}	8.9	9.3	-0.4
Lys1	C^{α}	Cya8	C^{α}	6.9	7.2	-0.3
Cys2	C^{α}	Arg3	C^{α}	3.8	3.8	0
Cys2	C^{α}	Gly4	C^{α}	6.0	6.4	-0.4
Cys2	C^{α}	Asp5	C^{α}	7.4	7.6	-0.2
Cys2	C^{α}	Cys6	C^{α}	5.9	6.3	-0.4
Cys2	C^{α}	Phe7	C^{α}	7.9	8.3	-0.4
Cys2	C^{α}	Cya8	C^{α}	6.7	7.7	-1
Arg3	C^{α}	Gly4	C^{α}	3.8	3.8	0
Arg3	C^{α}	Asp5	C^{α}	6.8	6.8	0
Arg3	C^{α}	Cys6	C^{α}	6.4	7.3	-0.9
Arg3	C^{α}	Phe7	C^{α}	8.1	10.0	-1.9
Arg3	C^{α}	Cya8	C^{α}	6.6	9.8	-3.2
Gly4	C^{α}	Asp5	C^{α}	3.8	3.8	0
Gly4	C^{α}	Cys6	C^{α}	5.4	6.0	-0.6
Gly4	C^{α}	Phe7	C^{α}	8.0	9.1	-1.1
Gly4	C^{α}	Cya8	C^{α}	8.3	10.1	-1.8
Asp5	C^{α}	Cys6	C^{α}	3.8	3.8	0
Asp5	C^{α}	Phe7	C^{α}	7.0	7.2	-0.2
Asp5	C^{α}	Cya8	C^{α}	8.9	9.6	-0.7
Cys6	C^{α}	Phe7	C^{α}	3.8	3.8	0
Cys6	C^{α}	Cya8	C^{α}	5.8	6.3	-0.5
Phe7	C^{α}	Cya8	C^{α}	3.8	3.8	0

seen for Arg3-Cya8 which is the consequence of changes in ψ angles for residues 1 and 8. This results in a change of the conformation of the bridge connecting residues Lys1 and Cya8. In fact, P2tr adopts a conformation which is closer to conformations of peptides 3 and 4. This can been seen from Fig. 5.28 and Fig. 5.33.

Similarly to structures calculated from NOESY data, a hydrogen bond is observed

Figure 5.33: The comparison of the the ensemble of structures calculated for P2frm (red), P2troe (black) and peptide 3 (blue).

between $6H^N$ and 4CO. It is found in 85 % of P2tr structures.

In conclusion, structures calculated from off-resonance ROESY yielded conformations that agreed well with the NOESY-derived structures. It was decided that the full relaxation matrix approach using a series of NOESY spectra with different mixing times is a better way of determining the inter-proton distances. When signal to noise ratio is an issue, the T-ROESY experiment should be used.

5.9 Structure of peptides 1-4 in H_2O

The conformation of residue 8 and CLL link in peptide 1 are the most significant differences between this and the other peptides. Residues 7 and 8 exist in two conformations in peptide 1; these can be described as closed and open. The closed conformation cannot exist in peptides 2, 3 or 4 because of the steric effects of the Ncc residues which prevent the carbonyl group of residue 8 from interacting with $4H^N$. NOE cross peaks between the aromatic ring protons of Phe7 and Ncc9 also point to the existence of an extended conformation in peptides 2-4.

Other than that, the substituents (metal binding site and Ncc residues) have a negligible influence on the conformation of the CRGDC fragment as indicated by the similarities of all the NMR parameters discussed above. Presumably, the metal binding site is too far from the backbone and attached to a flexible side chain to have an influence on the conformation of the peptides. The influence of the Ncc residue can be found by comparing the experimental data of peptide 1 and 2. The chemical shifts (H^N and C') of Phe7, Cya8 and Lys1 are different between the two peptides. The value of the temperature coefficients also show that residues 7 and 8 are different. The high temperature dependence of Cya8 for peptide 1 can be related to the flexibility of

this residue in the absence of Ncc which can stabilize the conformation.

Based on the similarity of the NMR parameters and convergence of all peptides to similar structures, we can conclude that the conformation of the RGD motif is determined by the arrangement of the ring formed by CRDGC residues and the disulfide bridge. The analysis of experimental data show that the conformation of the disulfide bridge is time-averaged. The intensity of NOESY cross peaks related to $d_{n\beta}$(i;i) are equal for both non-degenerate H^{β} protons of Cys6 in all spectra. For diastereotopic protons, equal distance would indicate a χ^1 angle around 180° which is not consistent with the coupling constants related to this angle. The disulfide bond also displays signs of flexibility as none of the side chains of $CH^{\beta 2}$ protons of Cys2 and Cys6 is structurally restrained. The analysis of the $\chi 1$ angle of Cys6 using $^3J_{H^{\alpha}H^{\beta}}$, $^3J_{H^{\beta}C'}$ and $^3J_{NH^{\beta}}$ coupling constants showed that they are not compatible with a single angle; H^{β} protons of Cys2 are degenerated for peptides 2-4 and strongly coupled for peptide 1. The structural flexibility of cysteine side chains contrasts with the relative rigidity of their backbones. The size of coupling constants ($^3J_{H^NH^{\alpha}}$, $^3J_{H^NC'}$ and $^3J_{H^NC^{\beta}}$) and the distances extracted from NOESY spectra strongly indicate that the backbone of Cys2 and Cys6 are restrained at ϕ=150±30° and ψ=150±30°.

On the other hand, the RGD motif is more flexible. The dihedral backbone angles of residues Asp5 and Arg3 tend to be pushed to the edge of the allowed regions leading to discrepancies between the distance and dihedral restraints. This is a sign of flexibility and can be attributed to the lack of strong hydrogen bonding which would stabilize of the conformation of this turn.

Cyclic RGD peptides are designed to mimic, but at the same time, restrain the flexibility of turns occurring in proteins. The presence of a Gly-Asp dipeptide favors the formation of βII'-turns [126]. The RGD loops in the NMR structures of the studied peptides do not meet all the criteria to be classified as beta turns. ϕ, ψ dihedral angles of glycine (-120, 120) are in agreement with the βII' turn but the conformation of residues i+1 and i+2 deviates more than 30° from an ideal dihedral angle of βII' turn which is defined by i+1 (60,-120) and i+2 (-80,0). In the loop formed by residues 3 to 6 (RGDC), the distance between C_i^{α} to C_{i+3}^{α} oscillates between 5.9 and 8.2 Å and no hydrogen bonds connect residues H_{i+2}^{N} and CO_i because of the equatorial orientation

of Arg3 with respect to the peptide ring and the position of the $5H^N$. The hydrogen bond between $6H^N$ and 4CO seen in the NMR-determined structures influences the ϕ_5 angle pushing the $5H^N$ to point to the outside of the peptide ring. Therefore, the GDC turn can be classified as a γ_i-turn with dihedral angles of Asp5 occupying ideal values of (-79, 69) with an allowed variation of 40^o and a hydrogen bond between CO_i and H^N_{i+3}.

Another turn occurs in 15%, 30% and 5% of structures calculated for peptides 1, 2 and 4, respectively. The turn is centered on the Lys1 and a hydrogen bond connects the residues $8CO^\zeta$ and $2H^N$. These small populations are consistent with the temperature coefficients of $2H^N$ which are higher than the random coil values.

5.10 Structure of peptide 2 in different solvents

Structural studies of potent RDG-containing cyclic ligands for $\alpha_v\beta_3$ are often carried out in DMSO [27, 38, 39, 127]. Problems of limited solubility and inefficient NOE transfers in H_2O have been overcome by several groups using DMSO as a solvent, either pure or in a mixture with H_2O. In this work, the results in water showed that the polyethylene glycol moiety and the metal binding site have no effect on the structure of the peptide core. Therefore only peptide 2 was taken forward for NMR studies in DMSO. The structure obtained is referred to as P2d. Peptide 2 was also studied in pure D_2O which, when combined with the H_2O data, enabled extraction of a similar number of cross peaks as available in DMSO (this in particularly true for H^α protons, which are obscured by the water signal in H_2O). The P2d structure can therefore be compared well with the P2w structure and also with the literature structures of RGD ligands of $\alpha_v\beta_3$ in DMSO [27, 38].

Few examples of NMR structures of cyclic RGD peptides in either DMSO or water have been published thus far [27, 39, 109]. However some report experimental parameters and comparison of structures obtained in both solvents [36, 128]. These studies suggest that the conformations in DMSO are similar to those in water and imply that there is less conformational averaging in DMSO compared to water. Locardi *et al.*[39] systematically examined a series of six RGD containing peptides in both solvents and

for five of them noticed differences. In some cases a different arrangement of β-turns was observed, in others the absence of coupling constants or NOE cross peaks in one of the solvent prevented the comparison of structures in DMSO and water. In the last cases, the weight of populations of different conformers was changed.

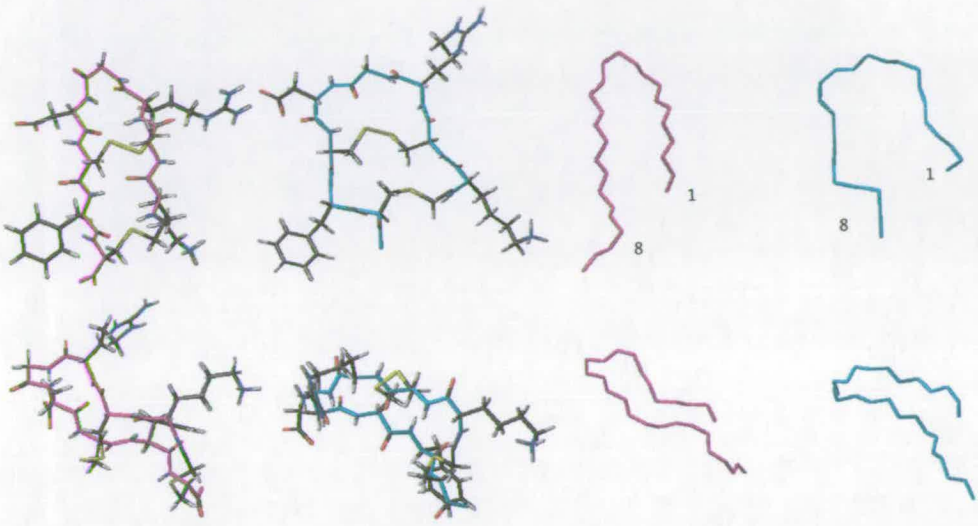


Figure 5.34: Conformation of peptide 2 in DMSO (magenta) and water (cyan).

The NMR-derived structures of peptide 2 in water and DMSO are very different (Fig. 5.34).This is despite the fact that a comparison of coupling constants and the temperature coefficients does not indicate such conformational changes (Tables 4.2, 4.3 and 4.8). However, the coupling constants used are related to the ϕ angle, for which few differences were observed (ϕ_5 and ϕ_7, Table 5.19).

A comparison of the backbone dihedral angles illustrates the differences between P2d and P2w structures (table 5.19). The non-glycine residues of the P2w and P2d are all located in the allowed β region of the Ramachandran plot. No significant difference can be observed between the ϕ angles of P2d and P2w. This is in line with the measurements, carried out in this work, which show similar values of $^3J_{H^NH^\alpha}$ coupling constants in both solvents. However, the values of ψ angles for residues 1, 3, 4 and 5 (all part of the loops) show large differences.

The conformation of the disulfide bridge changes dramatically in DMSO. The H^α protons of Cys2 and Cys6 are brought closer which rotates the disulfide bridge by 180^o relative to the backbone plane and orients the amide protons of the cysteines

Table 5.19: Dihedral angle of P2w and P2d structures: the angles and the standard deviation are calculated from an ensemble of 20 FRM structures with solvent refinement.

Residues	Peptide 2 (H_2O)	Peptide 2 (DMSO)
ϕ dihedral angles		
1LYS	-109 ±11	-101 ± 14
2CYS	-142± 13	-125± 12
3ARG	-119 ± 2	-147 ± 13
4GLY	126± 22	118± 16
5ASP	-105± 18	-144 ± 8
6CYS	-135± 12	-118± 11
7PHE	-79± 15	-128.5 ± 13
8CYA	-138± 14	-117 ± 7
ψ dihedral angles		
1LYS	58± 15	124 ± 14
2CYS	161 ± 14	121 ± 16
3ARG	102± 18	55 ± 10
4GLY	-127± 14	-35± 5
5ASP	76 ± 13	163 ± 4
6CYS	167 ± 7	133 ± 15
7PHE	115± 20	83 ± 8
8CYA	87± 15	61 ± 9

towards the outside of the ring (Fig. 5.34). The close conformation of P2d is also associated with four hydrogen bonds found in P2d structures (compared to one in P2w structures). An antiparallel β-sheet like structure is formed between residues 1-3 and 6-8 (Table 5.20 and Fig.5.35). The conformation of the RGD loop could not be categorized into usual types of turns found for RGD peptides despite the hydrogen bond formed between $5H^N_{i+2}$ and $3CO_i$. It could be classified as a β-hairpin connecting two antiparallel β-strands formed by Lys1-Cys2 and Cys6-Phe7.

Table 5.20: Number and occurrence of hydrogen bonds in peptide 2 in DMSO. The hydrogen bond $6H^N \cdots 5O_\delta$ occurs in all the structures.

Donor	Acceptor	Occurrence (%)
$1H^N$	7C'	55
$3H^N$	5C'	85
$6H^N$	$5O_{\delta 1}$	40
$6H^N$	$5O_{\delta 2}$	60
$7H^N$	1C'	95

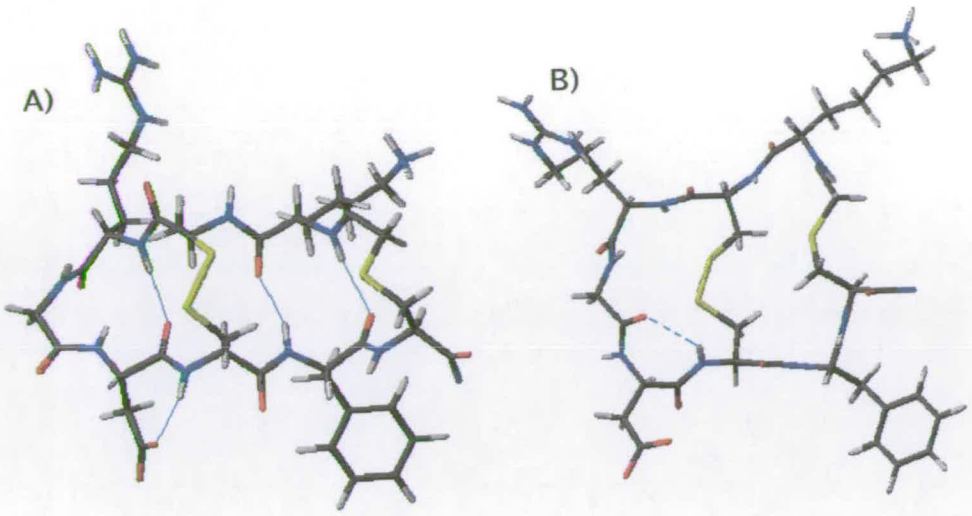


Figure 5.35: The hydrogen bonds are indicated by blue lines for the peptide core of peptide 2 in DMSO (A) and water (B) structures.

Overall, the conformations found in water and DMSO can be characterized as open and closed, respectively. The distances are changed dramatically (Table 5.21). The DMSO structure is arranged in a hairpin structure with Gly4 and the side chain of Cya8 at its extremities.

Such large changes of the conformation between the two solvents require a careful examination of experimental evidence, particularly as several studies suggested that the conformation of the RGD containing peptide does not change in DMSO [36, 128].

Distance restraints

A systematic comparison of NOESY cross peak assignments and their intensities was carried out for peptide 2 in order to validate the differences seen between the structure of this peptides obtained in different solvents. Chemical shifts, NOE enhancements and linewidth of amide protons were sometimes found to be different between H_2O and DMSO samples. For example, some NOESY cross peaks could be identified in DMSO only because of the dispersion of chemical shifts allowed them to be observed readily. This does not mean that these cross peaks are not present in water NOESY spectra where they could be overlapped with some other cross peaks. A comparison of the fingerprint region of the NOESY spectra immediately shows that a larger number of cross peaks is present in DMSO. The inspection focused on H^N , H^α , H^β and $8H^\delta$

Table 5.21: C^{α} C^{α} **distances for peptide 2 in water and DMSO:** Distances for residues 1 to 8 calculated with a cutoff of 15 Å.

Residue1	Atom1	Residue2	Atom2	d_{DMSO} (Å)	d_W (Å)	d_{DMSO}-d_W (Å)
Lys1	C^{α}	Cys2	C^{α}	3.8	3.8	0
Lys1	C^{α}	Arg3	C^{α}	6.6	7.0	-0.4
Lys1	C^{α}	Gly4	C^{α}	9.4	9.8	-0.4
Lys1	C^{α}	Asp5	C^{α}	8.0	11.2	-3.2
Lys1	C^{α}	Cys6	C^{α}	5.5	9.0	-3.5
Lys1	C^{α}	Phe7	C^{α}	5.3	8.9	-3.6
Lys1	C^{α}	Cya8	C^{α}	5.9	6.8	-0.9
Cys2	C^{α}	Arg3	C^{α}	3.8	3.8	0
Cys2	C^{α}	Gly4	C^{α}	6.1	6.4	-0.3
Cys2	C^{α}	Asp5	C^{α}	5.4	8.0	-2.6
Cys2	C^{α}	Cys6	C^{α}	4.2	6.4	-2.2
Cys2	C^{α}	Phe7	C^{α}	6.5	7.8	-1.3
Cys2	C^{α}	Cya8	C^{α}	8.2	7.1	1.1
Arg3	C^{α}	Gly4	C^{α}	3.8	3.8	0
Arg3	C^{α}	Asp5	C^{α}	5.3	7.0	-1.7
Arg3	C^{α}	Cys6	C^{α}	6.5	7.2	-0.7
Arg3	C^{α}	Phe7	C^{α}	8.9	9.9	-1
Arg3	C^{α}	Cya8	C^{α}	11.4	9.6	1.8
Gly4	C^{α}	Asp5	C^{α}	3.7	3.8	-0.1
Gly4	C^{α}	Cys6	C^{α}	6.9	5.9	1
Gly4	C^{α}	Phe7	C^{α}	9.9	9.2	0.7
Gly4	C^{α}	Cya8	C^{α}	13.0	9.9	3.1
Asp5	C^{α}	Cys6	C^{α}	3.8	3.8	0
Asp5	C^{α}	Phe7	C^{α}	6.8	7.1	-0.3
Asp5	C^{α}	Cya8	C^{α}	10.1	8.9	1.2
Cys6	C^{α}	Phe7	C^{α}	3.8	3.8	0
Cys6	C^{α}	Cya8	C^{α}	6.4	6.2	0.2
Phe7	C^{α}	Cya8	C^{α}	3.8	3.8	0

of residues 1 to 8 (peptide core) and the most structurally important cross peaks were analyzed carefully. Two lists of NOE cross peaks were created for each spectrum: the first contained only unique occurrences (either cross peaks observed only in the water or DMSO NOESY spectra) while the second one contained only common cross peaks. The inter proton distances were measured for both compounds using their FRM structures and checked against the distance restraints generated from the spectra collected in the other solvent. When a cross peak was compatible with both structures, the distance was not considered as structurally important. On the other hand, when a cross peak from one set was incompatible with the structure obtained with the other

set of restraints, it was classified as structurally important cross peak. The existence of such cross peaks supports the differences observed between P2w and P2d structures. The results of this analysis are summarized in Table 5.22 where structurally important cross peaks are reported.

Cross peaks	Water	DMSO
$8H_1^{\delta}$ $2H^N$	+	-
$8H_{1,2}^{\beta}$ $2H^N$	+	-
$1H^N$ $7H^N$	*	+
$7H^{\alpha}$ $8H_1^{\delta}$	-	+
$4H_1^{\alpha}$ $5H_1^{\beta}$	-	+
$1H^N$ $6H_1^{\beta}$	-	+
$2H^{\alpha}$ $6H^{\alpha}$	-	+ (very intense)
$7H^{\alpha}$ $8H_2^{\delta}$	-	+
$6H^{\alpha}$ $5H_2^{\beta}$	-	+

(-) not observed ; (+) observed ; (*) impossible to observe (overlaps or noise)

Table 5.22: **Structurally important NOE cross peaks in DMSO and water:** The differences reported concerns P2w and P2d.

The differences in table 5.22 indicate that the segments Lys1-Cys2 and Cys6-8Cya are closer in space in the P2d than in P2w structure. This is supported in particular by the existence of intense $6H^{\alpha}$-$2H^{\alpha}$ and $1H^N$-$7H^N$ cross peaks. The identification of structurally important cross peaks was difficult for the RGD loop because short sequential cross peaks are present in both DMSO and water.

The $6H^{\alpha}$-$2H^{\alpha}$ cross peak is very important as its existence strongly influences the final conformation of the peptide: When this cross peak was excluded from the list of restraints only a minority of structures converged to the closed conformation. Chemical shifts of $6H^{\alpha}$ and $2H^{\alpha}$ protons are very close in both solvents (Appendix F), which places their cross peak very close to the diagonal. To ensure that this cross peak is not artefactual, e.g. appears because of close diagonal peaks or that its intensity is not overestimated, a series of selective 1D NOESY experiments was acquired. The separation of the chemical shifts of $2H^{\alpha}$ and $6H^{\alpha}$ protons was large enough in DMSO in order to selectively excite one of them. Proton $6H^{\alpha}$ was chosen and the NOE buildup on close by protons was followed by repeating the experiments while increasing the mixing time. In water, the resonances of $2H^{\alpha}$ and $6H^{\alpha}$ were severely overlapped and could

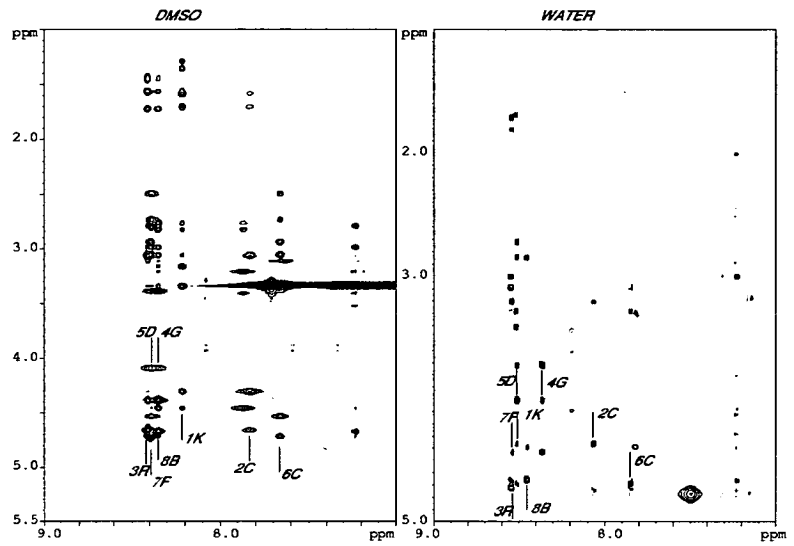


Figure 5.36: **Comparison of the fingerprint region of NOESY spectra of peptide 2 acquired in H_2O (200 ms mixing time) and DMSO (100 ms mixing time):** The H^N resonances are indicated by vertical lines.

not be to selectively excited. A series of 1D COSY-NOESY was therefore acquired by generating the magnetization of $6H^{\alpha}$ in a COSY step followed by a selective NOESY transfer from this proton. The COSY transfer started on $6H^{\beta}_2$ proton and the $6H^{\alpha}$,$6H^{\beta}$ coupling was refocused prior to the NOESY step. The increase in the intensity of signals of $2H^{\alpha}$ and both $6H^{\beta}$ protons was followed as a function of the mixing time (Fig. 5.37).

Experiments using a 2 ms mixing time were acquired in order to inspect a possibility of any artefactual signals. The results illustrated that clean selective excitation of $6H^{\alpha}$ in DMSO could be achieved without any partial excitation of proton $2H^{\alpha}$ (Fig. 5.37b) and that a sufficient amount of $6H^{\alpha}$ signal is generated by COSY transfer (Fig. 5.37f). The intensity of the $2H^{\alpha}$ proton in a series of 1D NOESY spectra in DMSO (Fig. 5.37b-d) increased with the increasing mixing time whilst the intensity of the selected $6H^{\alpha}$ proton decreased at the same time. This is to be expected for a pair of dipolar coupled protons between which the NOE is building up. Quantitatively, these experiments indicates that the enhancement of $2H^{\alpha}$ is slightly less, but still comparable with that of $6H^{\beta}$ protons. Although the intensity of $2H^{\alpha}$ is biased by the presence of the nearby signal of $6H^{\alpha}$ in DMSO, the categorization of the cross peak between these two protons into < 3.3 Å category is justified: H^{α} H^{β} intra-residue distances of < 2.9

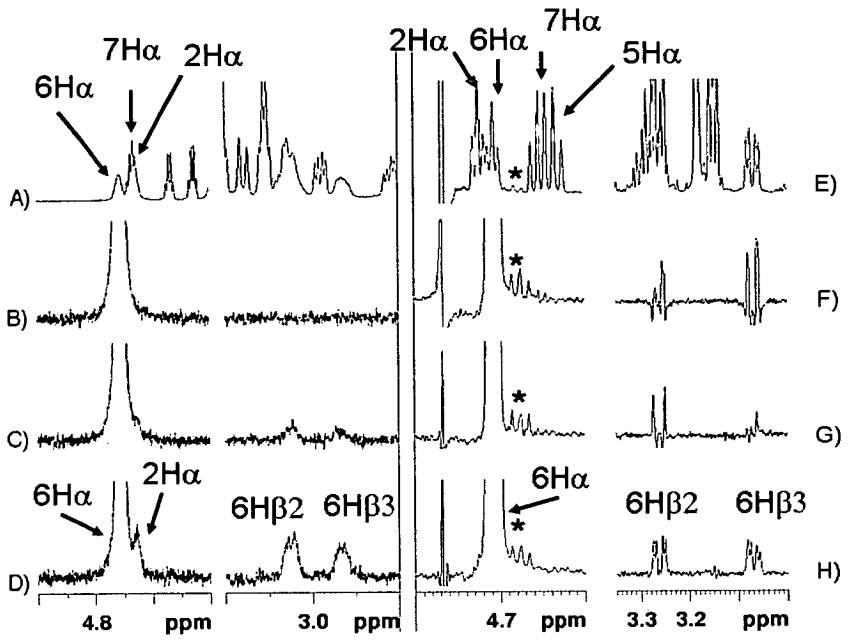


Figure 5.37: **Series of 800 MHz NOE enhancements from proton 6H$^\alpha$ of peptide 2:** The panels on the left (A-D) show spectra acquired in DMSO, while panels on the right (E-F) show spectra acquired in water. A and E: partial spectra showing the overlap of resonances. B-D: 1D NOESY spectra with mixing times 2, 75 and 200 ms respectively. F-H: 1D COSY-NOESY spectra with COSY transfer 6H$^\beta$ 6H$^\alpha$ and selective NOESY step from 6H$^\alpha$ using mixing times of 2,75 and 200 ms. An asterisk indicates impurity.

Å are typically observed in proteins [79]. The series of 1D NOESY experiments thus confirmed the existence of an intense 2H$^\alpha$ 6H$^\alpha$ cross peak observed in the 2D NOESY spectra of peptide 2 in DMSO indicating that the atoms 2H$^\alpha$ and 6H$^\alpha$ are close in space. This contrasts with peptide 2 in water where there is no evidence of spatial proximity between these two atoms. As can be seen in the spectrum of Fig. 5.22h, only a small hint of 2H$^\alpha$ signal is visible in the 200 ms mixing time spectrum, with a total intensity of only a fraction of the intensity of H$^\beta$ protons.

Next attention was paid to the disulfide bridge. There should be clear experimental evidences in the spectra to confirm the observed change of its conformation (Fig. 5.34, 5.35) between the two solvents. The position of the disulphide bridge with respect to the peptide backbone is intimately linked with the orientation of the H$^\alpha$ and H^N protons of the cysteines. In water, H^N protons of cysteines point toward the center of

the peptide ring whereas in DMSO the H^{α} protons are oriented toward the center of the molecule. Consequently, in DMSO the H^{α} protons are closer in space to the H^{β} protons of the other cysteine forming the disulfide bond. To illustrate this difference, F_2 traces through $6H_2^{\beta}$ protons were extracted from 2D NOESY spectra acquired in different solvents (Fig. 5.38) and the intensity of $2H^{\alpha}$ $6H^{\beta 2}$ cross peaks was compared. This cross peak is very intense in the DMSO spectrum while very weak in H_2O. A higher intensity of the cross peak observed in DMSO confirmed that in DMSO these distances are shorter, leading to a more compact structure compared to structures in water.

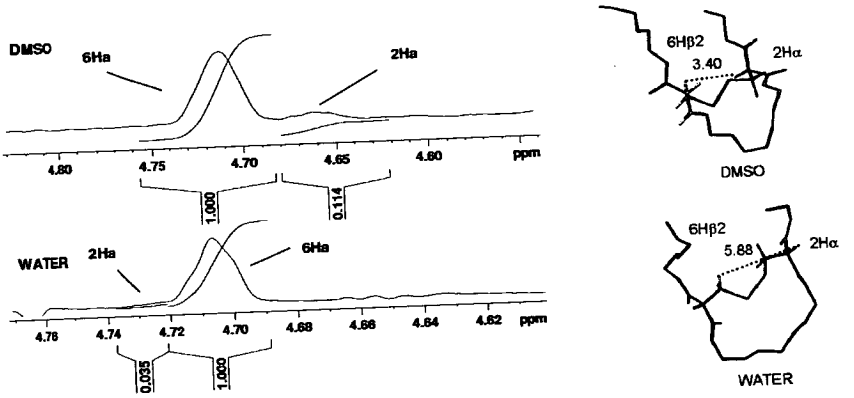


Figure 5.38: **Distances between $2H^N$ and 6_2^{β} in P2d and P2w:** F_2 traces of $6H_2^{\beta}$ extracted from 2D NOESY spectra acquired in (a) DMSO and (b) D_2O . The $2H^{\alpha}$ $6H_2^{\beta}$ cross peak is significantly more intense than the $6H^{\alpha}$ $6H_2^{\beta}$ cross peak. The distances between $2H^{\alpha}$ $6H_2^{\beta}$ protons are shown as a dashed line in P2w and P2d structures.

A comparison of the cross peaks intensities in 2D NOESY spectra acquired at short mixing times (55ms) and in both solvents suggests higher flexibility of the side chains in water. In general a comparison of H^{α} to H^{β} and H^N to H^{β} distances deduced from the relative intensities of the corresponding cross peaks reports on the conformation of the side chains [38, 105]. Nevertheless, it is instructive to compare the $d_{\alpha\beta}$ and $d_{n\beta}$ distances in order to rationalize the differences in the conformation of side chains between P2w and P2d structures, e.g. Cys6 in Fig. 5.39.

In water, the intensities of H^{α} H_2^{β} and H^{α} H_3^{β} cross peaks are similar as are those of H^N H_2^{β} and H^N H_3^{β} cross peaks. The implied equidistance involving H^{β} protons

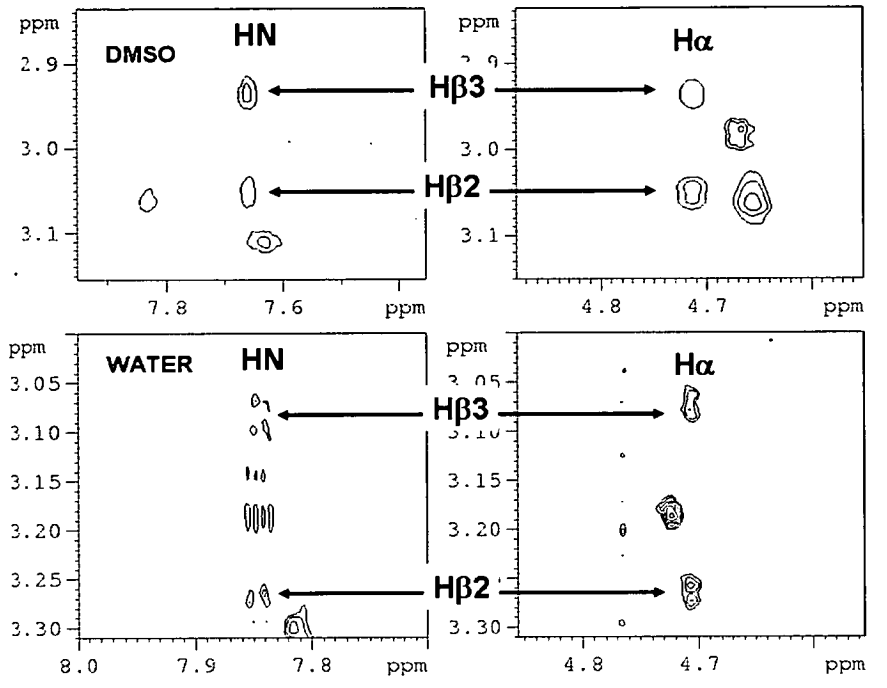


Figure 5.39: Comparison of (H^N H^β) and (H^α H^β) 55 ms NOESY cross peaks in water and DMSO for residue Cys6.

is impossible considering the geometry of the amino acids. The uniform cross peak intensity can only be achieved by extensive averaging that takes place on NMR time scale. This contrasts with the unequal intensities of the NOESY cross peaks observed in DMSO spectra indicating some rigidity. However, as discussed before (chapter 4), the analysis of scalar coupling constants did not show any convergence in DMSO.

Scalar coupling constants

The $^3J_{H^NH^\alpha}$ coupling constants are systematically higher in DMSO than in H_2O with an exception of Cys2 and Asp5 where their values could not be measured. Their size (>8 Hz) strongly indicate a conformational preference in residues 1, 3, 6, 7 and 8. These values were used to restrain ϕ dihedral angles for these residues. Similarly, large $^3J_{H^NH^\alpha}$ coupling constant for glycine is interpreted as a sign of rigidity. The difference between $J_{H^NH_1^\alpha}$ and $J_{H^NH_2^\alpha}$ coupling constants of Gly4 indicates rigidity in P2d where this difference was 4.3 Hz in DMSO and less than 1 Hz in water (Table 4.2) [51]. Concerning the χ dihedral angles, they could not be restrained in a particular

conformation. Similarly to water, the coupling constants did not converged towards any of the staggered conformers which indicates a certain degree of flexibility. However, the H^{β} protons were not degenerate in DMSO, with exception of $2H^{\beta}$, which contrasts with the spectra in water where the H^{β} protons were degenerate for residue 2, 7 and 8.

Chemical shifts dispersion

Rather than directly comparing chemical shifts of corresponding protons in different solvent, which is difficult to interpret, we focus on the H^{β} protons and inspect the separation of chemical shifts of these diastereotopic protons from the spectra acquired at 800 MHz. The H^{β} protons are degenerate (or strongly coupled) in water for residues Cys2, Phe7 and Cya8. In contrast, $7H^{\beta}$ protons in DMSO are separated by 154 Hz. Separation of the $5H^{\beta}$ protons has increased to 196 Hz in DMSO from 96 Hz in water. Similarly, H^{α} protons of the glycine are 452 Hz apart in DMSO while only 234 Hz in water. These data, though qualitative, indicate larger flexibility of the studied peptides in aqueous solution. This is in line with observations made by others [129].

Temperature coefficients

The temperature coefficients are significantly higher in DMSO than in water and suggest that more hydrogen bonds are present in DMSO. High temperature coefficients are observed for the H^{N} proton of residues 1, 2, 3, 6 and 7. Except for Cys2, for which no hydrogen bond was observed, all the other residues were involved in hydrogen bonds in the NMR structures of P2d (Table 5.20).

Chapter 6

Docking studies of peptides on $\alpha_V\beta_3$ integrin

Docking was performed using the AUTODOCK 3.05 [94]. The closest-to-mean water-refined NMR structures of peptides 1-4 were docked into their receptor using the X-ray structure of $\alpha_V\beta_3$ integrin solved by Xiong *et al.* [25].

6.1 Interactions of peptides docked into $\alpha_V\beta_3$ integrin.

6.1.1 Cilengitide

AUTODOCK was initially used to verify the ability of the program to reproduce the complex between Cilengitide and $\alpha_V\beta_3$ integrin as well as the structures found by Marinelli *et al.*[29] in docking studies using AUTODOCK. For the most populated cluster (22 structures out of 50) (Fig. 6.1A), structures of peptides 1-4 are almost identical to those described in the crystal structure and found in docking studies. The second cluster found in docking calculations (10 structures) is also mentioned by the authors and presented as another possible binding mode (Fig. 6.1B). This conformation of Cilengitide differs by the orientations of the N-methyl-valine and D-phenylalanine residues. In both modes of binding, the RGD triad interact in the same manner with the integrin. Rather than the T-shaped interaction with (β)-T122 of conformation A, in the conformation B the hydrophobic peptide residues interact with integrin residues (β)-K253, (β)-N313 and (α)-I216.

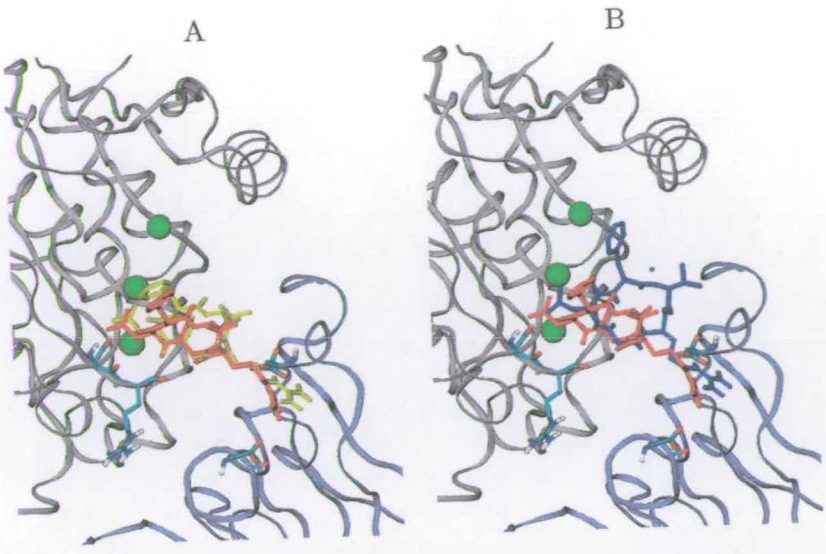


Figure 6.1: **Conformations of Cilengitide in crystal structure and after docking runs:** The figures show the two conformations of Cilengitide after docking (conformations A and B in yellow and blue, respectively) compared to the crystal structure (red). Calcium ion are represented with green spheres. Residues forming hydrogen bonds (sticks) from chain α (grey) and β (iceblue)

6.1.2 Peptides 1 to 4

NMR studies have shown that substituents attached to residues 1 and 8 do not modify the conformation of the peptide core. Based on this observation, Ncc and Lym residues were truncated in order to improve the convergence of the docking and to reduce the number of rotatable bonds.

The clusters of the lowest energy complexes were also the most populated (13/50, 4/50, 6/50, 10/50 for peptides 1, 2, 3 and 4, respectively) with an exception of peptide 2. Here the most populated cluster in a complex corresponds to an artifact where the truncated Ncc residues interacts with the receptor. These structures had energies 3.7 kcal.mol^{-1} higher than the lowest energy structure. The lowest energy structures of truncated peptides showed that Ncc and Lym do not interact with the receptor and these residues are oriented toward the solvent. Overall, the ligand-receptor interactions were not perturbed by the truncation of the peptides and it was assumed that the Ncc and metal binding site parts of the molecule were not interacting with the binding site of the integrin.

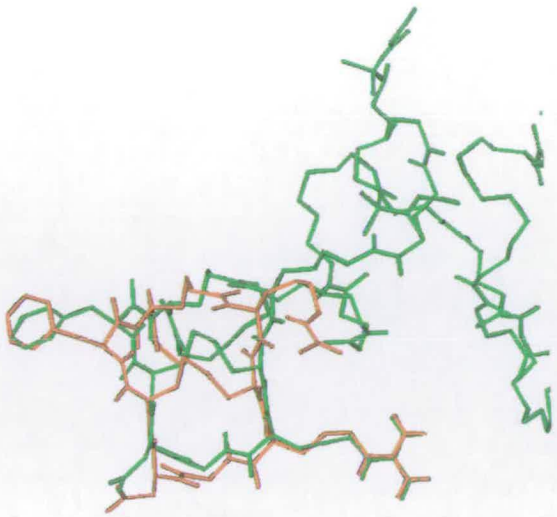


Figure 6.2: **Comparison of peptide 4 in the bound state using truncated and entire peptide:** The truncated and the entire peptide are shown in orange and green, respectively.

This assumption was confirmed by docking calculations using the entire peptides. The convergence of these calculations was not as good as with truncated peptides. Moreover, the analysis based on clusters was more difficult to perform because of the flexibility of Ncc and Lym residues which increased the rmsds of peptide conformations. Clustering of the structure with a threshold of 5 Å would generally resulted in populations containing no more that one individual. Hence, due to the degree of freedom and the flexibility of Ncc and Lym, no cluster analysis was relevant when entire peptides were used. Structures using entire peptides were analysed by comparing them with the lowest energy clusters of the truncated peptides. Typically only a few structures of the complex containing the entire peptide had similar conformations to the truncated ones. This was attributed to the existence of the large conformational heterogeneity of the structure of intact peptides. The occurrence of each conformation within 50 calculated structures was therefore low. Nonetheless, conformations of the peptide core for the lowest-energy structures were very similar for the truncated and intact peptides (Fig. 6.2). This indicated that the interactions with the protein are predominently mediated by the peptide core of the ligand.

The peptide-protein interactions using the low energy clusters of intact peptides were compared to those using the truncated peptides. Similar interactions were observed

for both cases. This confirmed that the peptides could safely be truncated without affecting the conformational search of the lowest energy conformation of the ligands.

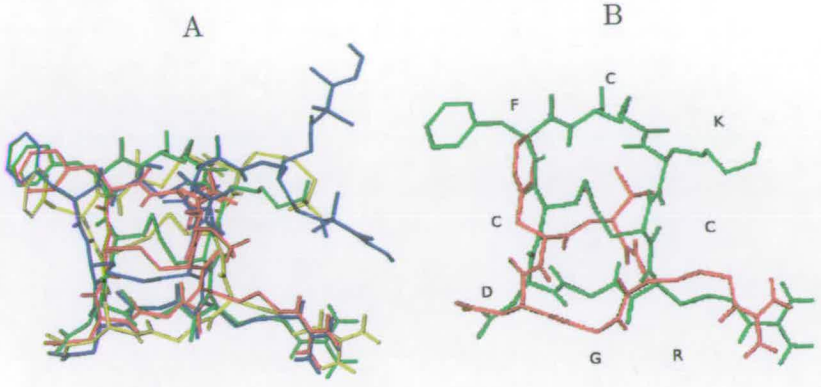


Figure 6.3: **Bound conformations of peptides 1 to 4:** (A) Overlay of the peptide cores of peptides 1, 2, 3 and 4 are shown in red, yellow, blue and green respectively. (B) Peptide 4 (green) is compared to the crystal structure of Cilengitide (red).

In addition, the ligand-receptor interactions were similar to those observed with Cilengitide. Similarly to Cilengitide, side chains of Arg and Asp are in an extended conformation and point in opposite directions with residue Gly lying at the interface between the α and β subunits of the integrin. The Arg3 residue of the ligand is inserted into a negatively charged groove forming hydrogen bonds between the guanidium group and α-D150, α-Q180 or α-D218. Side chains of Arg3 also interacts with the phenyl ring of α-Y178. The aspartic acid of the ligand forms ionic interaction with the calcium ion of MIDAS. However, this residue, unlike Cilengitide, does not interact with β-R216 nor β-N215. It forms a hydrogen bond between its carboxylate and the side chain of β-S123. The interactions of Asp5 with β-S123, in peptides, increase the distance between the receptor and backbone atoms of Gly4-Asp5 (Fig. 6.3B). For peptides 1-4 docked, the location of the aspartic acid is closer to MIDAS and makes the above interactions impossible. The phenyl group of Phe7 is positioned into an hydrophobic pocket formed by the side chain of β-K125 and β-Y122 but not at the same location as Cilengitide. This difference is a consequence of an extra residue, Cys6, inserted between aspartic and phenylalanine amino acids in the peptides which increases the distance between the aromatic ring of Phe7 and carboxylate group of Asp5 (Fig. 6.3 B). Residue α-Y178 is placed between the side chains of Lys and Arg of peptides 1-4 and mainly interacts with backbone atoms of Lys1, Cys2 and Arg3 of the ligand.

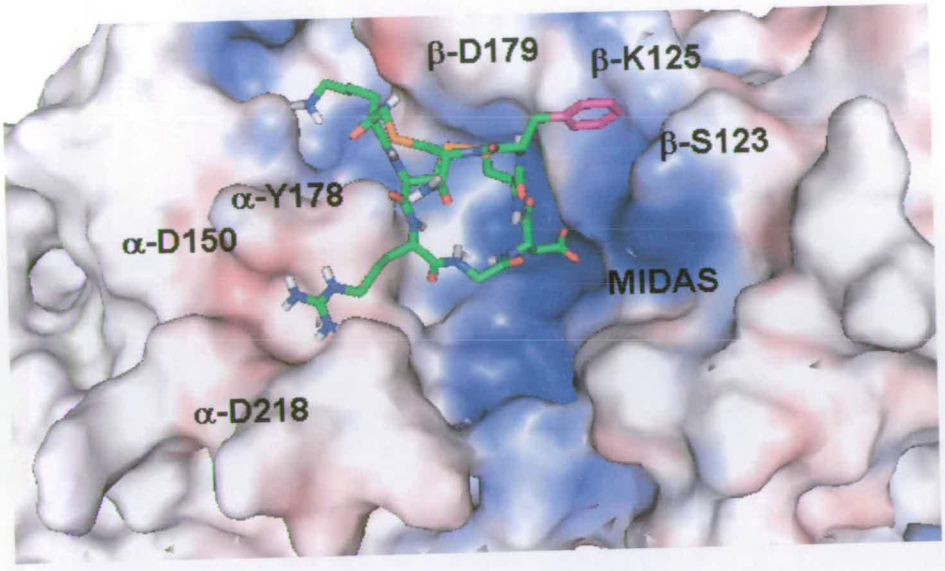


Figure 6.4: **Interactions between peptide 1 and $\alpha_v\beta_3$ integrin:** The electrostatic potential map of the binding site of the receptor was calculated at pH 7 using PBB2PQR [130] and CHARMM force field with Ca^{2+} parameters and Adaptive Poisson-Boltzmann Solver (APBS) 4.0 [131]. Negative and positive charges are indicated in red and blue, respectively. For clarity, only polar hydrogens are shown on peptide 1.

In addition to the above common features, some differences were observed between the docked conformations of individual peptides 1-4. In the top cluster of peptide 1, the orientation of the lysine residue of the ligand has no preferred conformation. This residue forms hydrogen bonds with either α-D148 or β-D179 in peptide 1. In the case of peptide 2, this lysine residue is close to the negatively charged region around α-D150 but does not form hydrogen bonds. Finally, in peptides 3 and 4, the modifed lysine, Lym, does not interact with the receptor. In these peptides, Ncc residues are oriented away from the receptor and do not interact with the RGD binding site.

For truncated peptides, the estimated free energy of binding of the lowest energy structures of truncated peptides is similar -34.9, -32.9, -27.6 and -29.4 kcal.mol^{-1} for peptide 1, 2, 3 and 4 respctively. Interactions and conformations of the peptides being similar, the difference observed between peptide 1/2 and 3/4 can be explained by the presence of a positively charged group (Lysine) which interacts with the receptor in peptides 1 and 2, which is not present in peptides 3 and 4.

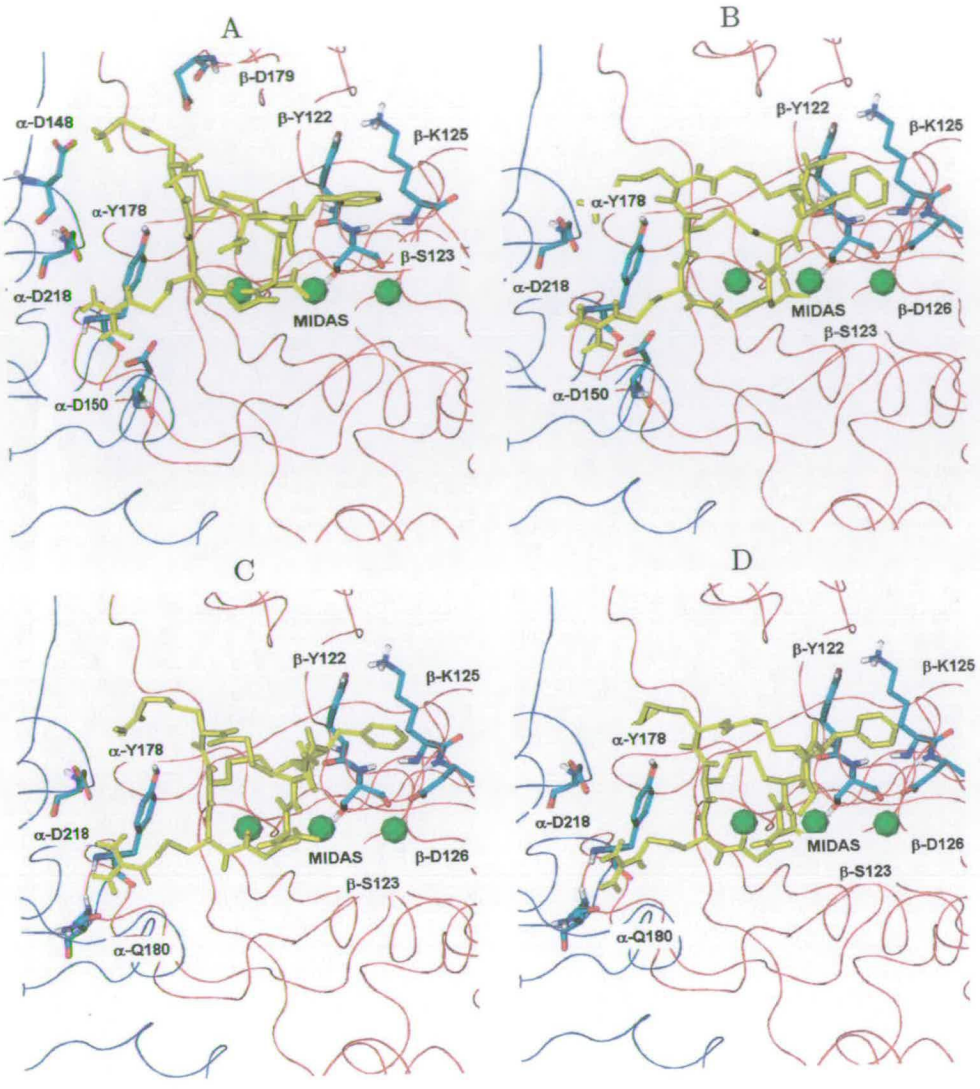


Figure 6.5: **The interactions of the docked peptides 1 - 4 (A - D) with the receptor:** Interacting residues of the receptor are indicated using sticks. Docked peptides are represented by yellow sticks.

6.2 Conclusions.

Docking studies showed that the binding mode of peptides 1-4 to $\alpha_v\beta_3$ integrin is similar to that of Cilengitide. Three sites are essential for the binding of these ligands to the $\alpha_v\beta_3$ integrin [29, 132]: the positively charged pocket around MIDAS where the aspartic acid of the RGD peptide binds, the negatively charged pocket around α-D150, α-Q180 and α-D218 and hydrophobic interactions with β-K125 and β-Y122 (Fig. 6.4). All these were also confirmed by the docking studies.

The conformation of the peptides is suitable to position Arg, Asp and phenyl group into the three pharmacophoric points which are believed to be essential for a specific interaction of RGD peptides with $\alpha_v\beta_3$ integrin. The presence of Ncc or the metal binding site do not conflict with the binding mode of the peptides.

Chapter 7

Discussion

7.1 Summary of the measured coupling constants

Several NMR methods were tested for their suitability for the measurement of coupling constants of the small cyclic peptides. Inevitably, some methods were unsuccessful, delaying the progress. The set of experiments that were found to work well are ω_1-filtered TOCSY experiments (^{13}C and ^{15}N HETLOC), refocused/decoupled HMBC for the determination of heteronuclear coupling constants and ^{1}H-^{15}N HSQC, 1D TOCSY and a combination of 1D TOCSY and 1D COSY. The other experiments considered were not suitable because of inadequate sensitivity or complications with extraction of coupling constants. Within the suite of experiments used, some work only for intra-residue coupling constants, while others can also provide the inter-residue coupling constants. In this regard they are complementary.

7.2 A structural comparison of RGD peptides binding $\alpha_v\beta_3$

The structures of several ligands selectively binding $\alpha_v\beta_3$ integrin have been solved using NMR spectroscopy [27, 38, 39, 127] or free molecular dynamics filtered by NMR data [36, 37]. One of the most potent ligands, Cilengitide [27] (under clinical trial) has also been co-crystallized with the $\alpha_v\beta_3$ integrin [25]. This structure provided a valuable insight into the binding of RDG ligands to integrins. Until this structure appeared in 2002 (pdb code 1L5G) [133] the specificity of RGD ligands was thought

to require a distance between the ArgC$^{\beta}$ - AspC$^{\beta}$ atoms shorter than 6.7 Å [38, 39, 127]. This distance was based on studies of RGD containing cyclic peptides and their selectivity toward the $\alpha_v\beta_3$ integrin carried out by Kessler et al. [40, 41]. However, the solution structure of the most potent ligand for the $\alpha_v\beta_3$ integrin, Cilengitide, showed that this distance is 8.0 Å and made the authors withdraw their claims regarding the distance between the C$^{\beta}$ atoms. According to the crystal structure and docking studies carried out [29], simultaneous binding of the α and β subunits of the integrin would be impossible for shorter distances between the C$^{\beta}$ atoms. We are therefore going to compare cyclic RGD peptides 1-4 with the crystal structure of Cilengitide in complex with $\alpha_v\beta_3$ [25]. Also available are the 3D coordinates of one peptide originally isolated from a phage-display peptide library. This peptide containes 4 cysteins that spontaneously forms two isomers with different combinations of disulfide bonds. The binding of these isomers was tested and their structures were solved by NMR in DMSO by Assa-Munt et al. [38]. One of these cyclic peptides was found to be a potent ligand for the $\alpha_v\beta_3$ integrin (pdb code 1FUV).

Cilengitide

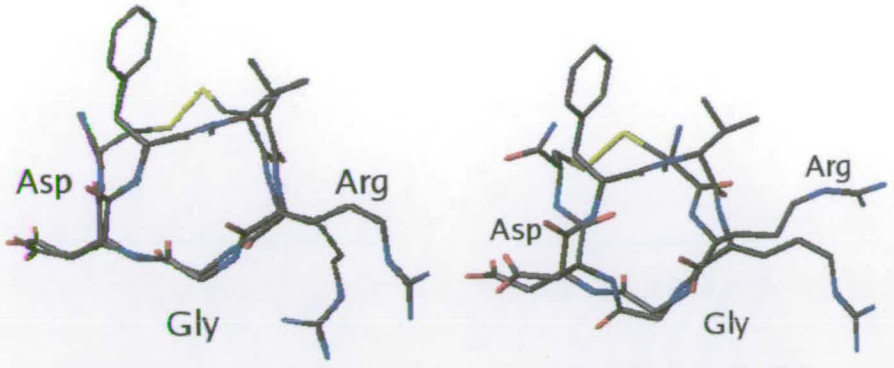
The rational design of RGD pentapeptides by Kessler's group led to a highly selective and active ligand for $\alpha_v\beta_3$ integrin: cyclo(RGDf-N(Me)V-) or Cilengitide [27]. The structure of this peptide was solved by NMR in water and DMSO, however the coordinates of these structures are not available. In water, the restrained molecular dynamics showed conformational averaging of Cilengitide where only the turn centered on the Arg is distinct, the others turns are distorted. The water structures of this cyclic peptide contain $\gamma_i - \gamma - \gamma_i$ turns with residues Arg, Gly and Asp occupying the central positions. However, when the free molecular dynamics in water was performed the above structural elements were only present for 36%, 31% and 15% of the time, respectively. The flexibility of this compound was also manifested by the lack of a preferred conformation of side chains. The free molecular dynamics also revealed backbone flexibility, particularly associate with the glycine, whose angle showed oscillations up to 70^o, while only 40^o deviations were observed for arginine and phenylalanine. Overall, a single conformation was not able to satisfy NMR data obtained in DMSO indicating

large flexibility also in this solvent. No major differences were observed between the structures in DMSO and water.

Based on the published NMR data [27], the compounds share some common features with the solution structure of Cilengitide. The values of the proton-proton ($^3J_{H^N H^\alpha}$, $^3J_{H^\alpha H_2^\beta}$ and $^3J_{H^\alpha H_3^\beta}$) coupling constant in water are similar for the RGD part. Furthermore, the free molecular dynamics also shows several hydrogen bonds (stabilizing the γ-turn) appearing during the prequilibrium period and distorting the NMR restrained structures at the position reported by Kessler on residues Arg, Gly and Asp. However, none of these appear to be stable during the free molecular dynamics runs. Regarding the NMR data, the main difference concerns the interpretation of the temperature coefficients. In all peptides, there is no line broadening observed at higher temperature for H^N protons thought to be involved in a hydrogen bond (Cys6). The temperature coefficients are also different for the RGD loop. In Cilengitide glycine has the highest value (the least negative). This indicates that the structure of all peptides is different and perhaps more rigid as also indicated by lower standard deviation (typically 20^o) of the ϕ angles observed during the free molecular dynamics.

The coordinates of the peptide in a bound state are available via Xiong et. al. (2002) who co-crystallized Cilengitide with the $\alpha_v\beta_3$ integrin. A comparison of NMR structures of peptide 2 in water and DMSO (structures P2w and P2d) with Cilengitide bound to $\alpha_v\beta_3$ integrin (Fig. 7.1) shows that the RGD motif of P2w is very similar to that of Cilengitide (backbone rmsd of heavy atoms of RGD, 0.41 Å). The side chains (Arg and Asp) are oriented in the same manner in both peptides. The distances between the C^β of the side chains are also similar with 9.0Å for the ensemble of peptides 1-4 against 8.9Å for the crystal structure of Cilengitide. This contrasts with the P2d structure solved in DMSO were the residues 2-6 of peptide 2 form a β hairpin. The distance between the C^β of the side chains is 6.9 Å. The backbone rmsd of the RDG heavy atoms of 0.69 Å is also higher than with the water structure P2w (Fig. 7.1).

The X-ray structure shows that the aromatic ring of the Phe residue of Cilengitide participates in the binding. The distance between the phenyl ring and the RGD loop is, in peptides 1 to 4, necessarily larger due to the inclusion of an extra amino-acid (Cys6) between the aspartic acid and the phenylalanine. Cross peaks observed between



Cilengitide and P2w Cilengitide and P2d

Figure 7.1: RGD heavy atoms of Cilengitide overlaid with P2w and P2d backbone atoms of RGD.

the aromatic protons of Phe and the Ncc residue protons indicates that, in peptides 2-4, this ring is in contact with the PEG moiety. This may affect the affinity of the binding.

Peptide from phage display library, 1FUV

The structure of a bicyclic undecapeptide whose sequence is ACDCRGDCFCG with disulfide bonds 2-10 and 4-8, has been solved in DMSO by Assa-Munt et al. [38] The authors used distance restraints, chemical shifts restraints and dihedral restraints to calculate the structure. The dihedral angle restraints were only used when $^{3}J_{H^{N}H^{\alpha}}$ was larger than 8.0 Hz. Sides chains χ^1 angle were restrained through a comparison of $d_{n\alpha}$ and $d_{\alpha\beta}$. This peptide forms two cycles via cysteine disulfide bridges. One of the isomers (1FUV) shows high specificity toward the $\alpha_v\beta_3$ integrin. The structure presents a distorted βI-turn and a distance lower than 7 Å between the C^{β} of Arg and Asp of the RGD loop. Contrasting with the structure of Cilengitide, the vector $C^{\alpha}C^{\beta}$ of the arginine and the aspartic acid of the RGD motif are almost parallel which brings the side chains of these residues relatively close in space. The Ramachandran plot of the ensemble of structures deposited in the PDB was inspected. This plot indicates that the residue Arg5 and Cys2 are pushed towards the edge of the region allowed by dihedral restraints (figure 7.2).

The Ramachandran plot reveals that the distance restraints and the dihedral restraints

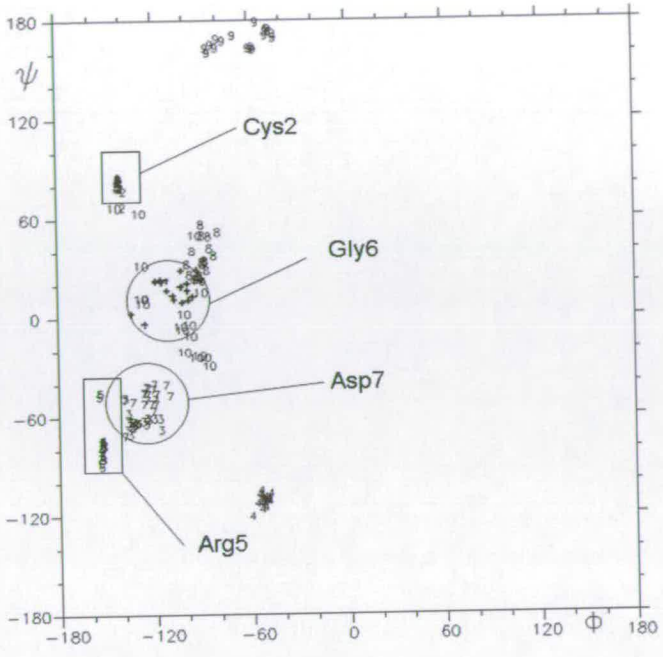
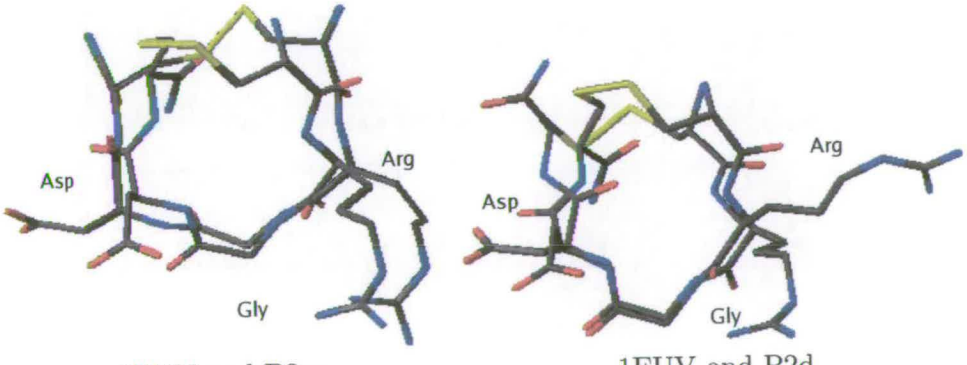


Figure 7.2: Ramachandran plot of the 10 best structures of 1FUV. The boxes indicate residues with ϕ dihedral angles that are pushed to the edge of the allowed region which residues Arg5 and Cys2 were set to 105±50 and 93±45, respectively. The regions of Gly6 and Asp7 are indicated by circles.

did not converge toward the same conformations. This reinforces the interpretation of the data that, in this structure, the flexibility of the 1FUV peptide is not fully accounted for. In this work, it was also the case for the water-refined structures and residue Arg3.



1FUV and P2w 1FUV and P2d

Figure 7.3: Heavy backbone atoms for the RGD motif 1FUV overlaid with P2w and P2d. Only the sequence common to FUV and peptides 1-4, CRGDCF, is represented.

The geometry of the backbone of 1FUV differs from that of P2w and P2d structures. First, the overall geometry of the peptide backbone is rather planar in all peptide while the structures of 1FUV is kinked. Focusing on the common sequence CRGDCF a better overlay is possible between 1FUV and P2d (1.62 Å) than between 1FUV and P2w (2.50 Å). The distance between $ArgC^{\beta}$ and $AspC^{\beta}$ is 6.9 Å in 1FUV is identical to the one measured in P2d (6.9 Å) and shorter than in P2w (8.9 Å). Similarly to P2d structures, 1FUV forms a closed structure. Nethertheless, there are significant differences between the 1FUV structure and the structures of peptides 1-4. The amide protons of the CRGDC motif point towards the inside of the ring (located in the inner part of the motif) in 1FUV whereas their orientation alternates in P2d and P2w structures which allows for hydrogen bonds to be formed. The Arg and Asp side chains ($C^{\alpha}C^{\beta}$ vectors) are both perpendicular to the plane of the turn in the 1FUV similar to P2d structure while they are more in the plane of the peptide ring in P2w. The position of the disulfide bridge relative to the peptide ring for 1FUV is similar to P2w (above the plane) and H^{α} protons are pointing outside of the peptide ring in 1FUV which is the opposite in P2d (beneath the plane) (FIg. 7.3). Also, the phenyl ring in 1FUV structures is close to the RGD motif whereas this group is distant from the loop in peptides 1-4. A comparison of the dihedral angles (Table 7.1) clearly shows that the shape of the molecules are completely different. This contrasts with relatively similar NMR parameters of 1FUV and peptide 2 in DMSO.

The temperature dependence of the amide protons for the RGD loop resembles those obtained for P2d and also the $^{3}J_{H^{N}H^{\alpha}}$ coupling constants are similar. However, several differences can be noticed. The chemical shifts of the amide protons of the first cysteine, Asp and Phe residues show differences of more than 0.1 ppm. Significant difference can also be seen for temperature coefficients of the first cysteine. Similarly to P2d, the cysteines surrounding the RGD motif show a large line broadening (no coupling constants could be measured). On the other hand the H^{N} proton of the aspartic acid of the RGD loop is not broadened, which contrasts with peptides 1-4.

Table 7.1: Compared dihedral angles of 1FUV, P2w and P2d for the common sequence CRGDCF: the angles and the standard deviation are calculated from an ensemble of 20 structures.

Residues	1FUV	P2w	P2d
ϕ dihedral angles			
Cys	-54 ± 3	-142± 13	-125± 12
Arg	-154 ± 1	-119 ± 2	-147 ± 13
Gly	-108 ±12	126± 22	118± 16
Asp	-125 ± 7	-105± 18	-144 ± 8
Cys	-91 ± 5	-135± 12	-118± 11
Phe	-62 ±15	-79± 15	-128 ± 13
ψ dihedral angles			
Cys	-111 ± 4	161 ± 14	121 ± 16
Arg	-77 ±12	102± 18	55 ± 10
Gly	17 ± 8	-127± 14	-35± 5
Asp	-48 ± 7	76 ± 13	163 ± 4
Cys	35 ±10	167 ± 7	133 ± 15
Phe	168 ± 5	115± 20	83 ± 8

Table 7.2: Comparison of the NMR parameters for the common sequence of 1FUV and peptide 2 in DMSO. For consistency, the chemical shifts δH^N and δH^α are compared (ppm) at the same temperature, 20°C. The coupling constant of the glycine residues is the sum of the coupling constants observed for each diastereotopic H^α protons. Δ_{coeff} are the temperature coefficients (ppb.K^{-1}).

Residues	FUV				P2d			
	H^N	H^α	$^3J_{H^NH^\alpha}$	Δ_{coeff}	H^N	H^α	$^3J_{H^NH^\alpha}$	Δ_{coeff}
Cys	7.69	4.70	8.5	-4.0	7.80	4.62	-	-0.1
Arg	8.49	4.48	8.5	-5.2	8.45	4.39	8.6	-5.3
Gly	8.37	4.13 3.45	10.3	-5.0	8.40	4.08 3.37	10.4	-6.0
Asp	8.56	4.56	8.6	-8.7	8.45	4.51	-	-7.6
Cys	7.69	4.71	-	-	7.62	4.68	9.7	1.3
Phe	8.70	4.66	8.6	-4.1	8.41	4.67	8.6	-3.4

Chapter 8

Conclusions

The structure of four bicyclic octapeptides was determined using an extensive set of NMR experiments. These peptides are used as contrast agents in cancer imaging. They rely on binding of the RGD motif to the $\alpha_v\beta_3$ integrin, which is up regulated during angiogenesis accompanying the growth of tumors. The conformation of the RGD loop was compared between all peptides and the influence of the substituents was assessed. The methods employed discarded, as much as possible, the size problem which is closely related to the quantity of available distance information for medium sized molecules. This enabled a comparison of structures of all peptides.

The convergence of structures was achieved by a combined use of the full relaxation matrix, coupling constants and refinement in explicit solvent. The full relaxation matrix appeared as the best method as the calculated distances rely on the variations of intensities of the NOEs. The use of several coupling constants also provides a valuable tool for determination of the local geometries of the peptides restraining them to a particular conformation. Finally, refinements in explicit solvents were important to account for electrostatic interactions and formation of hydrogen bonds.

As indicated by the chemical shifts, temperature coefficients and coupling constants, the conformation of the four peptides is similar. The method employed converged, for all peptides, towards comparable structures. In these cyclic peptides, the conjugated groups have a local influence. In the case of the polyethylene glycol tail, directly attached to the peptide core, the conformational changes are limited to the modified residue. The metal binding site, conjugated to the long side chain of lysine, has no

influence on the conformation of the peptide core.

The cyclic peptides form a reverse β-turn centered on the aspartic acid of the RGD motif which, according to free molecular dynamics, competes with other hydrogen bonds. Arginine residues present a certain degree of flexibility as revealed by the discrepancies observed between the distance and coupling constants restraints. From the analysis of the coupling constants related to the side chains dihedral angles and the intensities of cross peaks involving H^{β} protons, the side chains of the peptides were flexible and no preferred conformation could be identified, including the disulfide bridge.

Based on a similar conformation for all peptides, one peptide was studied in DMSO. In this solvent, the conformation was more rigid, stabilized by numerous hydrogen bonds. Overall the shape of the backbone changed to a closed conformation arranged in a β-hairpin. Structures in water and DMSO were compared to published structures of related compounds, Cilengitide and an RGD containing bicyclic undecapeptide, 1FUV. The water structures of the cyclic peptides were similar to those of the Cilengitide, a highly active and specific ligand for the $\alpha_v\beta_3$ integrin. The orientation of the side chain ($C^{\alpha}C^{\beta}$) segments of Asp and Arg relative to the backbone were similar as well as the dihedral angle of the RGD motif. The comparison with 1FUV showed that both structures adopt a compact conformation in DMSO, but are different despite sharing a similar sequence.

Docking studies were carried out using the structures of peptides 1-4 in water and all showed similar binding modes to that of Cilengitide co-crystallysed with $\alpha_v\beta_3$ integrin. Interactions between the peptides and the receptor were driven by three pharmacophoric points of the ligand: the guanidinium of the arginine, the carboxylate of the aspartic acid and a hydrophic interaction of the phenyl ring of the phenylalanine. The RGD motif was found at the interface between the two subunits of the integrin. The PEG and metal binding site adopted random conformations above the receptor without interacting with the protein and did not influence the binding mode of the peptide core with the $\alpha_v\beta_3$ integrin.

Structure determination of peptides in solution is a valuable tool for the design of RDG

ligands as the specificity of the peptide-receptor interaction is strongly related to the conformation of the ligand. These solution structures can be used as a starting point for rational design of more potent ligands or to select receptor subtypes by restraining the conformation of the RGD loop and interacting residues. This can be achieved by rational modifications using, for example, bulky groups conjugated to the ligand or substitutions of amino-acids of the ligand. Using a combination of dihedral restraints and full relaxation matrix refinement, as described in this work, conformation of a library of RGD peptides can be rapidly accessed and evaluated by docking studies.

The design of more potent ligands can be supplemented by interaction studies using computational methods when no crystal structures of the integrins are available. Integrins head groups can be constructed *in silico*, by comparing the amino-acid sequences of integrins, and used to dock ligands.

The design of contrast agents for medical imaging is based upon the optimisation of the structure of a lead compound. This involves maximising their *in vivo* effects while allowing metal bonding and other ligands to be attached to the core structure. However, these modifications might alter the pharmacophoric potential of the lead molecule, decreasing the affinity or the selectivity. Based on structural studies of four peptidic analogues, it was shown that the conformation of the pharmacophore and the binding to $\alpha_v\beta_3$ were not modifed in the presence of functional groups. These peptides are therefore attractive templates to which various labelling groups can be conjugated, allowing angigogenesis to be detected using different techniques that vary in sensitivity and spatial resolution.

Appendix A

Amino acids: usual and unusual letter code

Amino acid	Three letters code	One letter code
Alinine	ALA	A
Arginine	ARG	R
Asparagine	ASN	N
Aspartic acid	ASP	D
Cysteine	CYS	C
Glutamic acid	GLU	E
Glutamine	GLN	Q
Glycine	GLY	G
Histidine	HIS	H
Isoleucine	ILE	I
Leucine	LEU	L
Lysine	LYS	K
Methionine	MET	M
Phenylalanine	PHE	F
Proline	PRO	P
Serine	SER	S
Threonine	THR	T
Tryptophan	TRP	W
Tyrosine	TYR	Y
Valine	VAL	V
Modified cysteine	CYA	B
PEG tail	NCC	X
Modified lysine	LYM	J

Appendix B

CNS learn.inp

```
flags exclude *
include bonds angles dihedrals impropers
end
?

structure @@residue.mtf end

parameter
   reset
end

parameters
 learn initialize sele=(all)
    mode statistics
    !mode average
  end

end

   ?
coordinates @@compound3_template.pdb
    parameters
        learn accumulate
      end
   remarks verbose
end

parameters
  learn
    terminate
  end
  verbose
end

parameters
  reduce sele=(all) mode=average end
end
write
  parameters output=NewParameters.param
end


stop
```

Appendix C

Topology file

```
set message off echo off end
autogenerate
  angles=true
  dihedrals=false
end

residue LYM  ! modified Lysine
 group
  atom N   type=NH1  charge=-0.360 end    atom HN  type=H    charge= 0.260 end
  atom CA  type=CH1E charge= 0.000 end    atom HA  type=HA   charge= 0.100 end
  atom CB  type=CH2E charge=-0.200 end    atom HB1 type=HA   charge= 0.100 end
  atom HB2 type=HA   charge= 0.100 end    atom CG  type=CH2E charge=-0.200 end
  atom HG1 type=HA   charge= 0.100 end    atom HG2 type=HA   charge= 0.100 end
  atom CD  type=CH2E charge=-0.200 end    atom HD1 type=HA   charge= 0.100 end
  atom HD2 type=HA   charge= 0.100 end    atom CE  type=CH2E charge= -0.200 end
  atom HE1 type=HA   charge= 0.100 end    atom HE2 type=HA   charge= 0.100 end
  atom NZ  type=NH1  charge= -0.3 end     atom HZ  type=H    charge= 0.3 end
  atom CH  type=C    charge= 0.480 end   ! begining of the link
  atom OH  type=O    charge=-0.480 end    atom CQ  type=CH2E charge=-0.200 end
  atom HQ1 type=HA   charge=0.100 end     atom HQ2 type=HA   charge=0.100 end
  atom CI  type=CH2E charge=-0.200 end    atom HI1 type=HA   charge=0.100 end
  atom HI2 type=HA   charge=0.100 end     atom CK  type=CH2E charge=-0.200 end
  atom HK1 type=HA   charge=0.100 end     atom HK2 type=HA   charge=0.100 end
  atom CL  type=C    charge=0.480 end     atom OL  type=O    charge=-0.480 end
  atom C   type=C    charge= 0.480 end    atom O   type=O    charge=-0.480 end
  atom NO1 type=NH1   charge=-0.3 end       ! begining of the matal bindind site
  atom HO2 type=H     charge=0.3 end   atom CO3 type=CH2E charge=-0.2 end
  atom HM1 type=HA    charge=0.1 end   atom HM2 type=HA   charge=0.1 end
  atom CO6 type=CH2E  charge=-0.2 end  atom HN1 type=HA   charge=0.1 end
  atom HN2 type=HA    charge=0.1 end   atom CO9 type=CH1E charge=-0.1 end
  atom H11 type=HA    charge=0.1 end   atom C11 type=CH2E charge=-0.2 end
  atom HP1 type=HA    charge=0.1 end   atom HP2 type=HA   charge=0.1 end
  atom C14 type=CH2E  charge=-0.2 end  atom HR1 type=HA   charge=0.1 end
  atom HR2 type=HA    charge=0.1 end   atom N17 type=NH1  charge=-0.3 end
  atom H18 type=H     charge=0.3 end   atom C19 type=CTM  charge=0.1 end
  atom C20 type=CNI   charge=0.1 end   atom N21 type=NOH  charge=-0.2 end
  atom O22 type=ONC   charge=-0.2 end  atom H23 type=HON  charge=0.2 end
  atom C24 type=CH3E  charge=-0.3 end  ! methyl branched @ 7
  atom HS1 type=HA charge=0.1 excl=(HS1 HS2 HS3) end atom HS2 type=HA charge=0.1 excl=(HS1 HS2 HS3) end
  atom HS3 type=HA charge=0.1 excl=(HS1 HS2 HS3) end atom C28 type=CH3E charge=-0.3 end
  atom HS4 type=HA charge=0.1 excl=(H23 HS5 HS6) end atom HS5 type=HA charge=0.1  excl=(H23 HS5 HS6) end
  atom HS6 type=HA charge=0.1 excl=(H23 HS5 HS6) end atom C32  type=CH3E charge=-0.3 end
  atom HT1 type=HA charge=0.1 excl=(HT1 HT2 HT3) end atom HT2 type=HA charge=0.1 excl= (HT1 HT2 HT3) end
  atom HT3 type=HA charge=0.1 excl=(HT1 HT2 HT3) end   atom C36 type=CH2E  charge=-0.2 end
  atom HP3 type=HA   charge=0.1 end  atom HP4 type=HA   charge=0.1 end
  atom C39 type=CH2E charge=-0.2 end atom HR3 type=HA   charge=0.1 end
  atom HR4 type=HA   charge=0.1 end  atom N42 type=NH1   charge=-0.3 end
```

```
  atom H43 type=H    charge=0.3 end  atom C44 type=CTM   charge=0.1 end
  atom C45 type=CNI  charge=0.1 end  atom N46 type=NOH   charge=-0.2 end
  atom O47 type=ONC  charge=-0.2 end atom H48 type=HON   charge=0.2 end
  atom C49 type=CH3E charge=-0.3 end    ! methyl branched @ 7
  atom HS7 type=HA   charge=0.1 excl=(HS7 HS8 HS9) end  atom HS8 type=HA   charge=0.1 excl=(HS7 HS8 HS9) end
  atom HS9 type=HA   charge=0.1 excl=(HS7 HS8 HS9) end  atom C53 type=CH3E charge=-0.3 end
  atom HS10 type=HA  charge=0.1 excl=(HS10 HS11 HS12) end atom HS11 type=HA charge=0.1 excl=(HS10 HS11 HS12)
  atom HS12 type=HA  charge=0.1 excl=(HS10 HS11 HS12) end  atom C57  type=CH3E charge=-0.3 end
  atom HT4 type=HA   charge=0.1 excl=(HT4 HT5 HT6) end  atom HT5 type=HA    charge=0.1 excl=(HT4 HT5 HT6) end
  atom HT6 type=HA   charge=0.1 excl=(HT4 HT5 HT6) end
  bond N HN  bond N CA  bond CA HA bond CA CB bond CB HB1 bond CB HB2 bond CB CG  bond CG HG1  bond CG HG2
  bond CG CD  bond CD HD1  bond CD HD2 bond CD CE  bond CE HE1  bond CE HE2 bond CE NZ  bond NZ HZ
  bond NZ CH bond CH OH  ! LINK start bond CH CQ  bond CQ HQ1 bond CQ HQ2
  bond CQ CI bond CI HI1 bond CI HI2 bond CI CK  bond CK HK1 bond CK HK2
  bond CK CL bond CL OL  ! LINK  end bond CA C
  bond C  O bond NO1 CL bond HO2 NO1 bond NO1 C03 bond C03 HM1 bond C03 HM2
  bond C03 C06 bond C06 HN1 bond C06 HN2 bond C06 C09 bond C09 H10 bond C09 C11  bond C11 HP1  bond C11 HP2
  bond C11 C14 bond C14 HR1  bond C14 HR2 bond C14 N17
  bond N17 C19 bond  N17 H18 bond C19 C20 bond C20 N21 bond N21 O22 bond O22  H23
  bond C19 C24 bond C24 HS1 bond C24 HS2 bond C24 HS3 bond C19 C28 bond C28 HS4 bond C28 HS5 bond C28 HS6
  bond C20 C32 bond C32 HT1  bond C32 HT2  bond C32 HT3 bond C09 C36  bond C36 HP3  bond C36 HP4
  bond C36 C39 bond C39 HR3  bond C39 HR4 bond C39 N42 bond N42 C44 bond  N42 H43 bond C44 C45
  bond C45 N46 bond N46 O47 bond O47  H48 bond C44 C49 bond C49  HS7 bond C49 HS8  bond C49 HS9
  bond C44 C53 bond C53 HS10  bond C53 HS11  bond C53 HS12
  bond C45 C57 bond C57 HT4  bond C57 HT5  bond C57 HT6
  dihedral CG  CB  CA  N   dihedral CD  CG  CB  CA  dihedral CE  CD  CG  CB dihedral NZ  CE  CD  CG
  dihedral NZ  CH  CQ  CI  dihedral CH  CQ  CI  CK  dihedral CQ  CI  CK  CL
  dihedral NO1 C03 C06 C09 dihedral C03 C06 C09 C11 dihedral C06 C09 C11 C14
  dihedral C09 C11 C14 N17 dihedral C11 C14 N17 C19 dihedral C14 N17 C19 C24
  dihedral C14 N17 C19 C28 dihedral N17 C19 C20 C32
  dihedral N17 C19 C20 N21 dihedral C03 C06 C09 C36 dihedral C06 C09 C36 C39 dihedral C09 C36 C39 N42
  dihedral C36 C39 N42 C44 dihedral C39 N42 C44 C49 dihedral  C39 N42 C44 C53 dihedral N42 C44 C45 C57
  dihedral N42 C44 C45 N46
  improper HA N C CB  improper HB1 HB2 CA CG   improper HG1 HG2 CB CD
  improper HD1 HD2 CG CE improper HE1 HE2 CD NZ improper CE NZ CH CQ  improper CE NZ CH OH
  improper HZ NZ CH CQ  improper CH OH NZ HZ  improper HQ1 HQ2 CI CH improper HI1 HI2 CQ CK
  improper HK1 HK2 CI CL improper CK CL NO1 C03   improper H10 C06 C11 C36  improper HM1 HM2 NO1 C06
  improper HP1 HP2 C09 C14  C11 improper HR1 HR2 C11 N17
  improper HS1 HS2 HS3 C19  improper HS4 HS5 HS6 C19 improper HT1 HT2 HT3 C20  improper O22 N21 C20 C19
  improper O22 N21 C20 C32  improper HP3 HP4 C09 C39 improper HR3 HR4 C36 N42  improper HS7 HS8 HS9 C44
  improper HS10 HS11 HS12 C44 improper HT4 HT5 HT6 C45 improper O47 N46 C45 C44 improper O47 N46 C45 C57
end

residue NCC !ethylamine, atom  CH2N  based on CH2G from GLY
group
  atom N   type=NH1 charge=-0.36 end atom HN  type=H  charge=0.26 end atom CA  type=CH2N charge=-0.20  end
  atom HA1 type=HA  charge=0.10 end atom HA2  type=HA charge=0.10 end atom CB  type=CH2E charge=-0.20  end
  atom HB1 type=HA  charge=0.10 end atom HB2  type=HA charge=0.10 end  atom OG  type=OE charge=-0.50  end
  atom CD  type=CH2E  charge=0.05 end atom HD1  type=HA charge=0.10 end atom HD2  type=HA charge=0.10  end
  atom CE  type=CH2E  charge=0.05 end atom HE1  type=HA charge=0.10 end atom HE2  type=HA charge=0.10  end
  atom OZ  type=OE   charge=-0.50 end atom CH type=CH2E charge=0.05 end atom HH1  type=HA charge=0.10  end
  atom HH2 type=HA   charge=0.10 end atom CQ  type=CH2E charge=0.05 end atom HQ1 type=HA charge=0.10  end
  atom HQ2 type=HA   charge=0.10 end atom OI  type=OE charge=-0.5 end atom CK type=CH2E charge=0.05  end
  atom HK1 type=HA   charge=0.10 end atom HK2 type=HA charge=0.10 end atom CL type=CH2E charge=0.05  end
  atom HL1 type=HA   charge=0.10 end atom HL2 type=HA charge=0.10 end atom NM  type=NH1 charge=-0.36  end
  atom HM  type=H    charge=0.36 end atom CN  type=C charge=0.48 end atom ON  type=O   charge=-0.48  end
  atom CX  type=CH2E charge=0.20 end atom HX1  type=HA charge=0.15 end atom HX2 type=HA charge=0.15  end
  atom OO  type=OE  charge=-0.5 end atom CP type=CH2E charge=-0.20 end atom HP1 type=HA charge=0.10  end
  atom HP2 type=HA  charge=0.10 end atom C type=C charge=0.48 end atom O type=O charge=-0.48 end
  bond N HN bond N CA
  bond CA HA1 bond CA HA2 bond CA CB bond CB HB1 bond CB HB2 bond CB OG bond OG CD bond CD HD1
  bond CD CE  bond CE HE1 bond CE HE2 bond CE OZ  bond OZ CH bond CH HH1 bond CH HH2
  bond CH CQ  bond CQ HQ1 bond CQ HQ2 bond CQ OI bond OI CK  bond CK HK1  bond CK HK2
  bond CK CL  bond CL HL1  bond CL HL2 bond CL NM  bond NM HM bond NM CN  bond CN ON
  bond CN CX  bond CX HX1  bond CX HX2 bond CX OO  bond OO CP bond CP HP1  bond CP HP2
  bond CP C  bond C O dihedral N CA CB OG dihedral CA CB OG CD dihedral CB OG CD CE
  bond CD HD2
  dihedral OG CD CE OZ dihedral CD CE OZ CH dihedral CE OZ CH CQ
```

```
  dihedral OZ CH CQ OI dihedral CH CQ OI CK dihedral CQ OI CK CL
  dihedral OI CK CL NM dihedral CK CL NM CN dihedral CL NM CN CX
  dihedral NM CN CX OO dihedral CN CX OO CP dihedral CX OO CP C
  improper HA1 HA2 N CB improper HB1 HB2 CA OG improper HD1 HD2 OG CE improper HE1 HE2 CD OZ
  improper HH1 HH2 OZ CQ improper HQ1 HQ2 CH OI improper HK1 HK2 OI CL improper HL1 HL2 CK NM
  improper CL NM CN CX improper HM NM CN CX improper CL NM CN ON
 end

residue CYA   !modified cys : CYS-S-CH2-CO
  group
  atom N   type=NH1 charge=-0.36 end atom HN  type=H   charge= 0.26 end atom CA  type=CH1E  charge= 0.00 end
  atom HA  type=HA  charge= 0.10 end atom CB  type=CH2E  charge=-0.20 end atom HB1 type=HA  charge= 0.10 end
  atom HB2 type=HA  charge= 0.10 end atom SG  type=SM   charge=-0.05 end atom CD  type=CH2E  charge=-0.15 end
  atom HD1 type=HA   charge= 0.1 end atom HD2 type=HA   charge= 0.1 end atom CE  type=C   charge= 0.48 end
  atom OZ  type=O   charge=-0.48 end atom C   type=C   charge= 0.48 end atom O   type=O   charge=-0.48 end
  bond N  HN bond N  CA  bond CA HA bond CA CB     bond CB HB1     bond CB HB2 bond CB SG bond SG CD
  bond CD CE     bond CD HD1     bond CD HD2 bond CE OZ bond CA C
  bond C  O improper HA N C CB improper HB1 HB2 CA SG improper HD1 HD2 SG CE
  dihedral SG  CB  CA  N dihedral CD  SG  CB  CA dihedral CE  CD  SG  CB dihedral OZ CE  CD  SG
 end

presidue CLL      ! Cya Lys Link

 add bond 1CE 2N add angle 1CD 1CE 2N add angle 1CE 2N  2CA add angle 1OZ 1CE 2N add angle 1CE 2N  2HN
add improper 1OZ 1CE 2N 2CA add improper 2HN 2N 1CE 1CD add improper 1CD 1CE 2N 2CA
end

set echo=true end
```

Appendix D

Parameter file

```
set echo off message off end
 checkversion 1.1

 BOND  CH3E CNI   1000.000 {sd=   0.001}  1.511 BOND  CH3E CTM   1000.000 {sd=  0.001}    1.544
 BOND  CNI  CTM   1000.000 {sd=   0.001}  1.529 BOND  CNI  NOH   1000.000 {sd=  0.001}    1.292
 BOND  CTM  NH1   1000.000 {sd=   0.001}  1.457 BOND  HON  ONC   1000.000 {sd=  0.001}    0.941
 BOND  NOH  ONC   1000.000 {sd=   0.001}  1.293 ANGLe  CH2E CH1E CH2E  500 {sd= 0.031}  111.5218
 ANGLe  CH2E NH1  CTM   500 {sd=  0.031}  117.7235 ANGLe  CH3E CNI  CTM   500 {sd=  0.031} 120.1329
 ANGLe  CH3E CNI  NOH   500 {sd=  0.031}  118.3587 ANGLe  CH3E CTM  CH3E  500 {sd=  0.031} 108.7556
 ANGLe  CH3E CTM  CNI   500 {sd=  0.031}  109.2507 ANGLe  CH3E CTM  NH1   500 {sd=  0.031} 109.8513
 ANGLe  CNI  CH3E HA    500 {sd=  0.031}  109.902  ANGLe  CNI  CTM  NH1   500 {sd=  0.031} 109.7335
 ANGLe  CNI  NOH  ONC   500 {sd=  0.031}  125.7441 ANGLe  CTM  CH3E HA    500 {sd=  0.031} 110.512
 ANGLe  CTM  CNI  NOH   500 {sd=  0.031}  121.4885 ANGLe  CTM  NH1  H     500 {sd=  0.031} 109.1780
 ANGLe  HON  ONC  NOH   500 {sd=  0.031}  120.7046
 IMPRoper C    CH1E NH1  CH2N  500.00 {sd= 0.031} 0   -0.0382
 IMPRoper C    CH2N N    CH1E  500.00 {sd= 0.031} 0 -179.9669
 IMPRoper C    CH2N NH1  CH1E  500.00 {sd= 0.031} 0    0.0484
 IMPRoper C    CH2N NH1  CH2N  500.00 {sd= 0.031} 0   -0.0248
 IMPRoper C    CH2N OC   OC    500.00 {sd= 0.031} 0    0.0223
 IMPRoper C    O    NH1  CH2N  500.00 {sd= 0.031} 0 -180.0127
 IMPRoper CH1E C    NH1  CH2N  500.00 {sd= 0.031} 0 -180.0018
 IMPRoper CH1E N    C    CH2N  500.00 {sd= 0.031} 0  179.9856
 IMPRoper CH1E NH1  C    CH2N  500.00 {sd= 0.031} 0 -179.9916
 IMPRoper CH2N C    CH2P N     500.00 {sd= 0.031} 0   -0.0151
 IMPRoper CH2E C    NH1  CH1E  500.00 {sd= 0.031} 0 -179.9900
 IMPRoper CH2N C    H    NH1   500.00 {sd= 0.031} 0  180.0020
 IMPRoper CH2N C    N    CH2P  500.00 {sd= 0.031} 0   -0.0116
 IMPRoper CH2N C    NH1  CH2N  500.00 {sd= 0.031} 0  179.9899
 IMPRoper CH2N C    NH1  H     500.00 {sd= 0.031} 0    0.0161
 IMPRoper CH2N NH1  C    O     500.00 {sd= 0.031} 0   -0.0048
 IMPRoper HC   CH2N HC   HC    500.00 {sd= 0.031} 0  -66.4073
 IMPRoper C    CH2E NH1  CH1E  500.00 {sd= 0.031} 0  -179.9829
 IMPRoper HA   HA   CH2N OE    500.00 {sd= 0.031} 0  -69.8494
 IMPRoper HA   HA   CH2E OE    500.00 {sd= 0.031} 0  -69.8494
 IMPRoper HA   HA   OE   CH2E  500.00 {sd= 0.031} 0  -69.8494
{ residue LYM }
 IMPRoper CH1E CH2E HA   HA  500 {sd= 0.031} 0  -23.7146
 IMPRoper CH2E CH2E CH2E HA  500 {sd= 0.031} 0  -65.5636
 IMPRoper CH2E NH1  HA   HA  500 {sd= 0.031} 0   24.1307
 IMPRoper CH3E CNI  NH1  ONC 500 {sd= 0.031} 0 -172.1809
 IMPRoper CH3E CNI  NOH  ONC 500 {sd= 0.031} 0  0.46
 IMPRoper CNI  HA   HA   HA  500 {sd= 0.031} 0   73.5752
 IMPRoper CTM  CNI  NOH  ONC 500 {sd= 0.031} 0  178.69
 IMPRoper CTM  HA   HA   HA  500 {sd= 0.031} 0   75.5643
{ ============================ END residue LYM ====================}
 DIHEdral  NH1  CH2N OE   CH2E   0.00 {sd= 0.031} 3  0.0000
 DIHEdral  CH2N CH2E OE   CH2E   0.00 {sd= 0.031} 3  0.0000
```

```
 DIHEdral  OE   CH2E CH2E OE     0.00 {sd= 0.031} 3  0.0000
 DIHEdral  NH1  CH2E CH2E OE     0.00 {sd= 0.031} 3  0.0000
 DIHEdral  CH2E NH1  C    CH2E   0.00 {sd= 0.031} 3  0.0000
 DIHEdral  NH1  C    CH2E OE     0.00 {sd= 0.031} 3  0.0000
 DIHEdral  C    CH2E OE   CH2E   0.00 {sd= 0.031} 3  0.0000
 DIHEdral  CH2E OE   CH2E C      0.00 {sd= 0.031} 3  0.0000
 DIHEdral  CH2E OE   CH2E CH2E   0.00 {sd= 0.031} 3  0.0000
 DIHEdral  NH1  CH2N CH2E OE     0.00 {sd= 0.031} 3  0.0000
{ residue LYM }
 DIHEdral  NH1  C    CH2E CH2E  0 {sd= 0.031} 3    0.0000
 DIHEdral  C    CH2E CH2E CH2E  0 {sd= 0.031} 3    0.0000
 DIHEdral  CH1E CH2E CH2E NH1    0 {sd= 0.031} 3 -205.9499
 DIHEdral  CH2E CH1E CH2E CH2E   0 {sd= 0.031} 3  234.9263
 DIHEdral  CH2E CH2E NH1  CTM    0 {sd= 0.031} 3 -172.5657
 DIHEdral  CH3E CNI  CTM  NH1    0 {sd= 0.031} 3  113.1366
 DIHEdral  CH3E CTM  NH1  CH2E   0 {sd= 0.031} 3  -52.9527
 DIHEdral  NOH  CNI  CTM  NH1    0 {sd= 0.031} 3  -68.5117
{ ======================== END residue LYM ========================}
 NONBonded  OE   0.1000  2.5836  0.1000   2.5836  NONBonded  CH2N 0.0903  3.3409  0.0903   3.3409
 NONBonded  CTM 0.0903   3.3409  0.0903   3.3409  NONBonded  CNI 0.0903   3.3409  0.0903   3.3409
 NONBonded  NOH 0.1592   3.0068  0.1592   3.0068  NONBonded  ONC 0.2342   2.7755  0.2342   2.7755
 NONBonded  HON 0.0498   2.2272  0.0498   2.2272
 set echo on message on end
```

Appendix E

Link file

```
remarks file toppar/protein.link
remarks
remarks this is a macro to define standard protein peptide bonds
remarks and termini to generate a protein sequence.

set echo=false end

checkversion 1.1

Creates peptide bond
link pept    head - arg   tail + *     end
link pept    head - asp   tail + *     end
link pept    head - cya   tail + *     end
link pept    head - gln   tail + *     end
link pept    head - glu   tail + *     end
link pept    head - gly   tail + *     end
link pept    head - lys   tail + *     end
link pept    head - lym   tail + *     end
link pept    head - ncc   tail + *     end
link pept    head - phe   tail + *     end


Creates NH3+ group in N-terminus
Commented out for LYS or LYM
{first nter    tail + lys       end}
{first nter    tail + lym       end}

Creates amide for in C-terminus for CYA and NCC
last ctn     head - ncc       end
last ctn     head - cya       end

Other Patches : create the cyles
* Link CYA-LYS or CYA-LYM
patch CLL reference=1=(resid 8) reference=2=(resid 1) end
* Disulfide bridge
patch DISU reference=1=(resid 2) reference=2=(resid 6) end
```

Appendix F

Resonance assignments of ^{1}H, ^{13}C and ^{15}N

Res	Name	Atom	Peptide 1	Peptide 2	Peptide 3	Peptide 4	Peptide 2 DMSO
1	LYS	N	125.770	125.096	125.363	125.538	121.897
1	LYS	HN	8.567	8.527	8.496	8.499	8.212
1	LYS	CA	56.183	56.287	56.544	56.514	55.399
1	LYS	HA	4.413	4.397	4.357	4.357	4.304
1	LYS	CB	31.765	31.905	32.247	32.274	33.130
1	LYS	HB1	1.833	1.849	1.808	1.808	1.705
1	LYS	HB2	1.713	1.733	1.706	1.702	1.584
1	LYS	CG	24.745	24.747	25.122	25.085	25.432
1	LYS	HG1	1.443	1.445	1.361	1.369	1.355
1	LYS	HG2	1.376	1.378	1.300	1.303	1.289
1	LYS	CD	29.092	29.084	30.587	30.557	29.800
1	LYS	HD1	1.679	1.686	1.504	1.506	1.546
1	LYS	HD2	1.679	1.686	1.504	1.506	1.546
1	LYS	CE	42.150	42.170	41.849	41.829	41.822
1	LYS	HE1	2.981	2.990	3.151	3.153	2.772
1	LYS	HE2	2.981	2.990	3.151	3.153	2.772
1	LYS	NZ	57.273	55.292	126.260	126.426	*
1	LYS	HZ	7.432	7.512	7.936	7.943	7.711
1	LYS	C	175.473	175.565	175.669	175.613	174.08
2	CYS	N	118.798	118.767	118.475	122.241	115.779
2	CYS	HN	8.078	8.073	8.052	8.056	7.831
2	CYS	CA	54.872	54.856	55.019	54.962	54.686
2	CYS	HA	4.773	4.749	4.768	4.752	4.659
2	CYS	HB1	3.307	3.241	3.220	3.225	3.064
2	CYS	HB2	3.211	3.241	3.220	3.225	3.064
2	CYS	C	173.227	173.258	173.211	173.155	171.67
2	CYS	CB	44.887	44.815	44.896	44.816	46.627
3	ARG	N	122.009	122.078	122.098	118.615	120.180
3	ARG	HN	8.579	8.562	8.561	8.564	8.410
3	ARG	CA	55.445	55.789	55.837	55.819	55.212
3	ARG	HA	4.482	4.467	4.457	4.454	4.382
3	ARG	CB	30.778	30.754	30.720	30.683	31.669
3	ARG	HB1	1.849	1.849	1.847	1.850	1.723
3	ARG	HB2	1.748	1.758	1.744	1.750	1.564

3	ARG	CG	27.063	27.098	27.073	26.018	28.244
3	ARG	HG1	1.652	1.652	1.639	1.641	1.453
3	ARG	HG2	1.693	1.694	1.579	1.582	1.430
3	ARG	CD	43.301	43.229	43.293	43.256	43.462
3	ARG	HD1	3.202	3.207	3.198	3.197	3.109
3	ARG	HD2	3.202	3.207	3.198	3.197	3.109
3	ARG	NE	84.823	85.162	79.823	84.738	84.942
3	ARG	HE	7.096	7.170	7.178	7.184	7.634
3	ARG	CZ	159.640	159.646	159.633	159.574	*
3	ARG	C	176.170	176.090	176.080	176.010	174.115
4	GLY	N	111.139	111.103	111.235	111.371	108.528
4	GLY	HN	8.378	8.376	8.380	8.383	8.348
4	GLY	CA	45.583	45.493	45.592	45.574	45.499
4	GLY	HA1	4.025	4.047	4.061	4.061	4.092
4	GLY	HA2	3.739	3.755	3.728	3.730	3.382
4	GLY	C	174.000	173.898	173.972	173.907	172.69
5	ASP	N	122.350	122.102	122.345	122.432	119.926
5	ASP	HN	8.514	8.504	8.469	8.474	8.386
5	ASP	CA	54.189	53.102	53.494	53.407	52.171
5	ASP	HA	4.618	4.666	4.676	4.682	4.532
5	ASP	CB	39.958	37.868	38.824	38.657	38.091
5	ASP	HB1	2.685	2.865	2.780	2.797	2.733
5	ASP	HB2	2.591	2.745	2.669	2.679	2.495
5	ASP	CG	178.197	177.950	178.894	178.559	174.76
5	ASP	C	174.501	174.349	174.642	174.497	172.99
6	CYS	N	118.797	118.878	118.430	118.596	115.443
6	CYS	HN	7.825	7.843	7.847	7.856	7.661
6	CYS	CA	55.481	55.310	55.451	55.400	54.802
6	CYS	HA	4.736	4.736	4.748	4.752	4.717
6	CYS	CB	44.897	45.338	45.369	45.306	47.886
6	CYS	HB1	3.294	3.314	3.299	3.304	3.046
6	CYS	HB2	3.135	3.128	3.107	3.107	2.932
6	CYS	C	172.648	172.526	172.660	172.595	171.718
7	PHE	N	121.809	121.730	121.67	121.825	119.043
7	PHE	HN	8.469	8.552	8.538	8.544	8.392
7	PHE	CA	57.815	57.765	57.668	57.627	56.737
7	PHE	HA	4.657	4.647	4.669	4.668	4.672
7	PHE	CB	39.772	39.709	39.748	39.714	40.814
7	PHE	HB1	3.093	3.041	3.035	3.034	2.984
7	PHE	HB2	3.033	3.041	3.035	3.034	2.792
7	PHE	CG	138.887	138.895	138.901	138.899	140.57
7	PHE	HD1	7.298	7.258	7.247	7.253	7.228
7	PHE	HD2	7.298	7.258	7.247	7.253	7.228
7	PHE	CD1	131.953	131.818	131.934	131.877	132.398
7	PHE	CD2	131.953	131.818	131.934	131.897	132.398
7	PHE	HE1	7.365	7.349	7.337	7.342	7.248
7	PHE	HE2	7.365	7.349	7.337	7.342	7.248
7	PHE	CE1	131.516	131.317	131.475	131.432	131.303
7	PHE	CE2	131.516	131.317	131.475	131.432	131.303
7	PHE	CZ	129.954	129.819	129.900	129.845	129.558
7	PHE	HZ	7.310	7.310	7.299	7.300	7.183
7	PHE	C	175.072	174.811	174.862	174.780	173.909
8	CYA	N	125.425	125.094	124.729	124.878	120.863
8	CYA	HN	8.675	8.472	8.479	8.479	8.348

8	CYA	CA	55.878	55.622	55.530	55.557	55.444
8	CYA	HA	4.451	4.425	4.435	4.437	4.457
8	CYA	CB	36.583	36.379	36.608	36.558	38.110
8	CYA	HB1	2.921	2.887	2.878	2.880	2.827
8	CYA	HB2	2.869	2.887	2.878	2.880	2.759
8	CYA	CD	38.226	38.031	37.957	37.933	37.855
8	CYA	HD1	3.445	3.449	3.443	3.435	3.340
8	CYA	HD2	3.355	3.317	3.298	3.298	3.158
8	CYA	CE	174.865	174.714	174.669	174.517	172.529
8	CYA	C	176.448	173.780	173.747	173.707	172.590

* indicates that no resonance assignment could be done.

Appendix G

Temperature dependence of amide protons

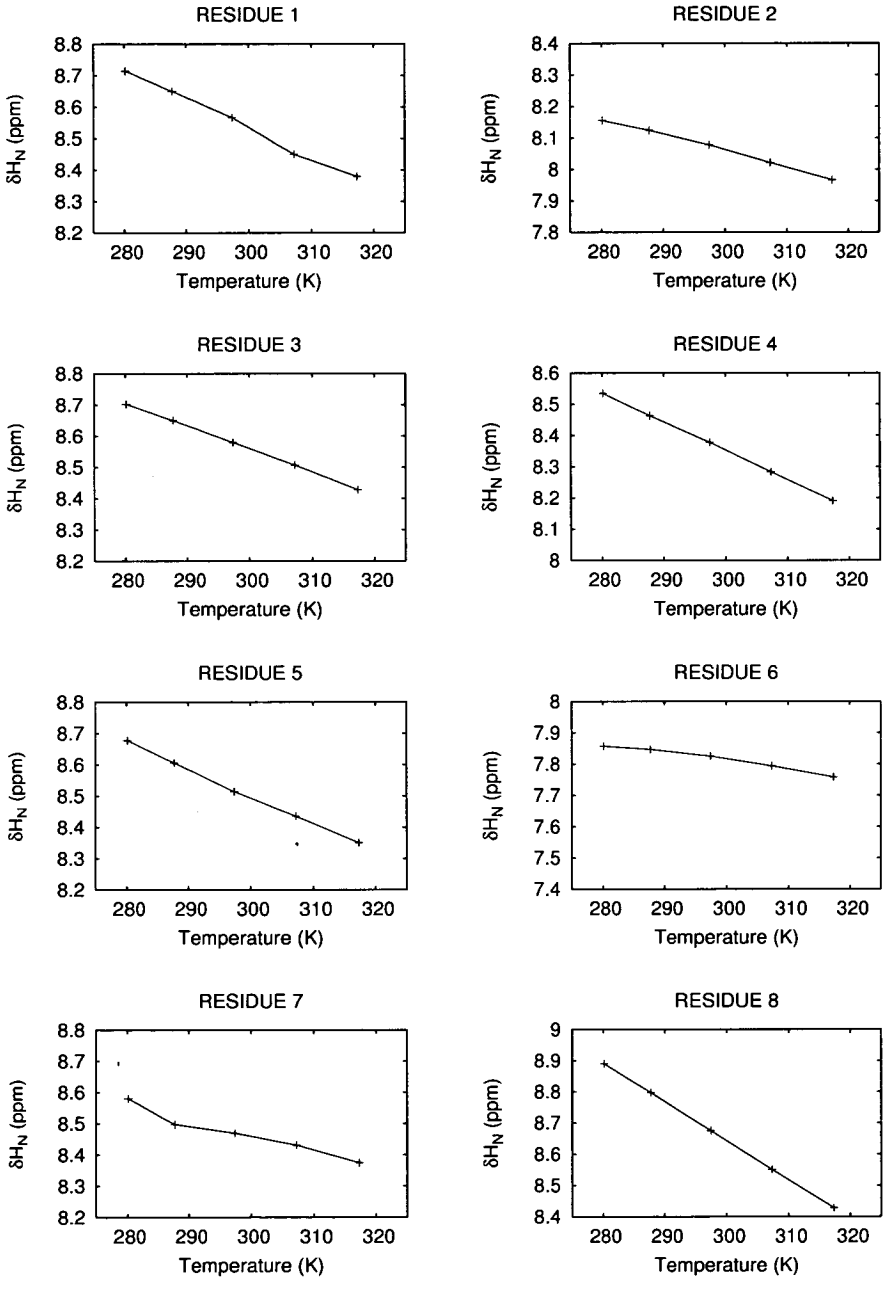


Figure G.1: Chemical shifts of H^N proton of peptide 1 as a function of temperature. The solid line joins individual points.

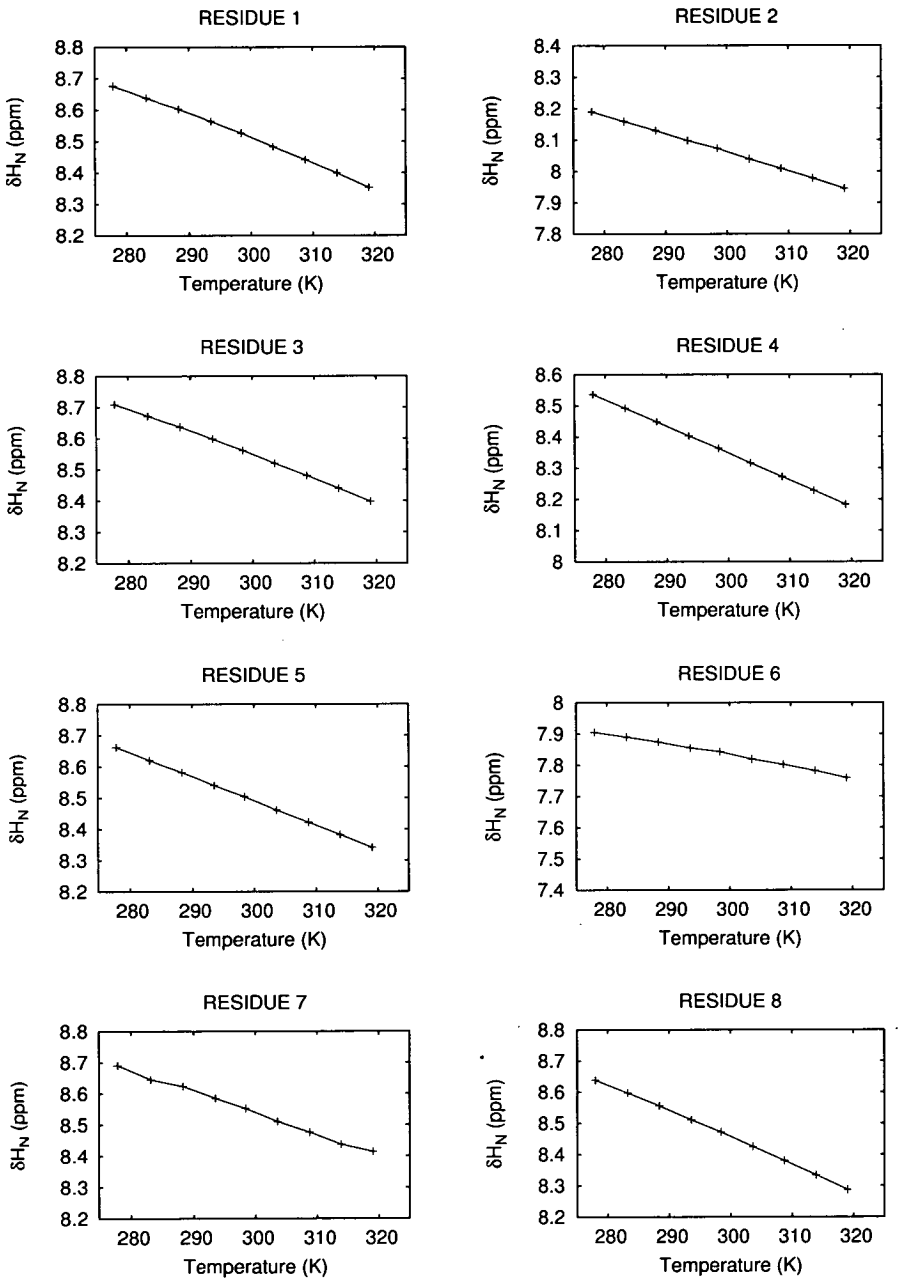


Figure G.2: Chemical shifts of H^N proton of peptide 2 as a function of temperature

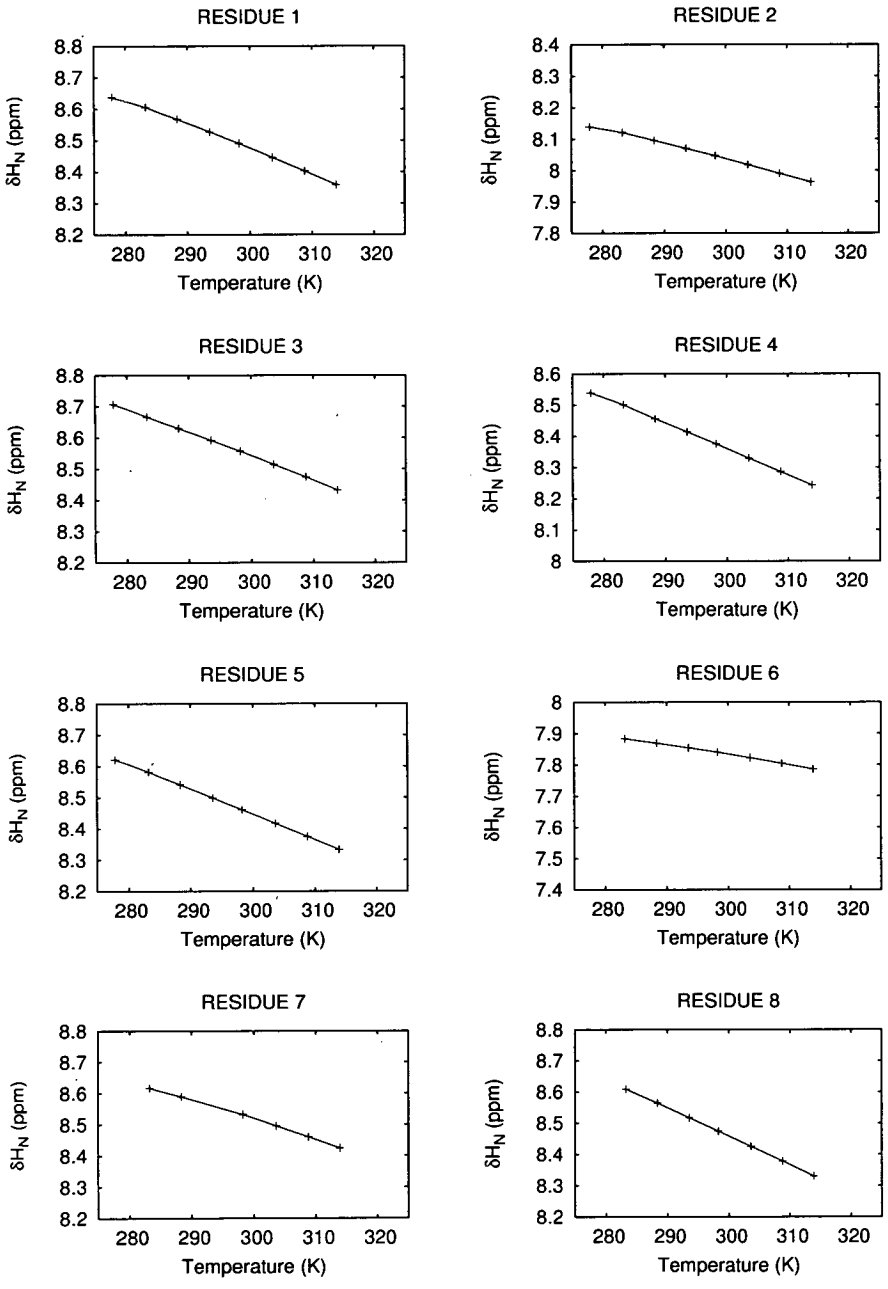


Figure G.3: Chemical shifts of H^N proton of peptide 3 as a function of temperature.

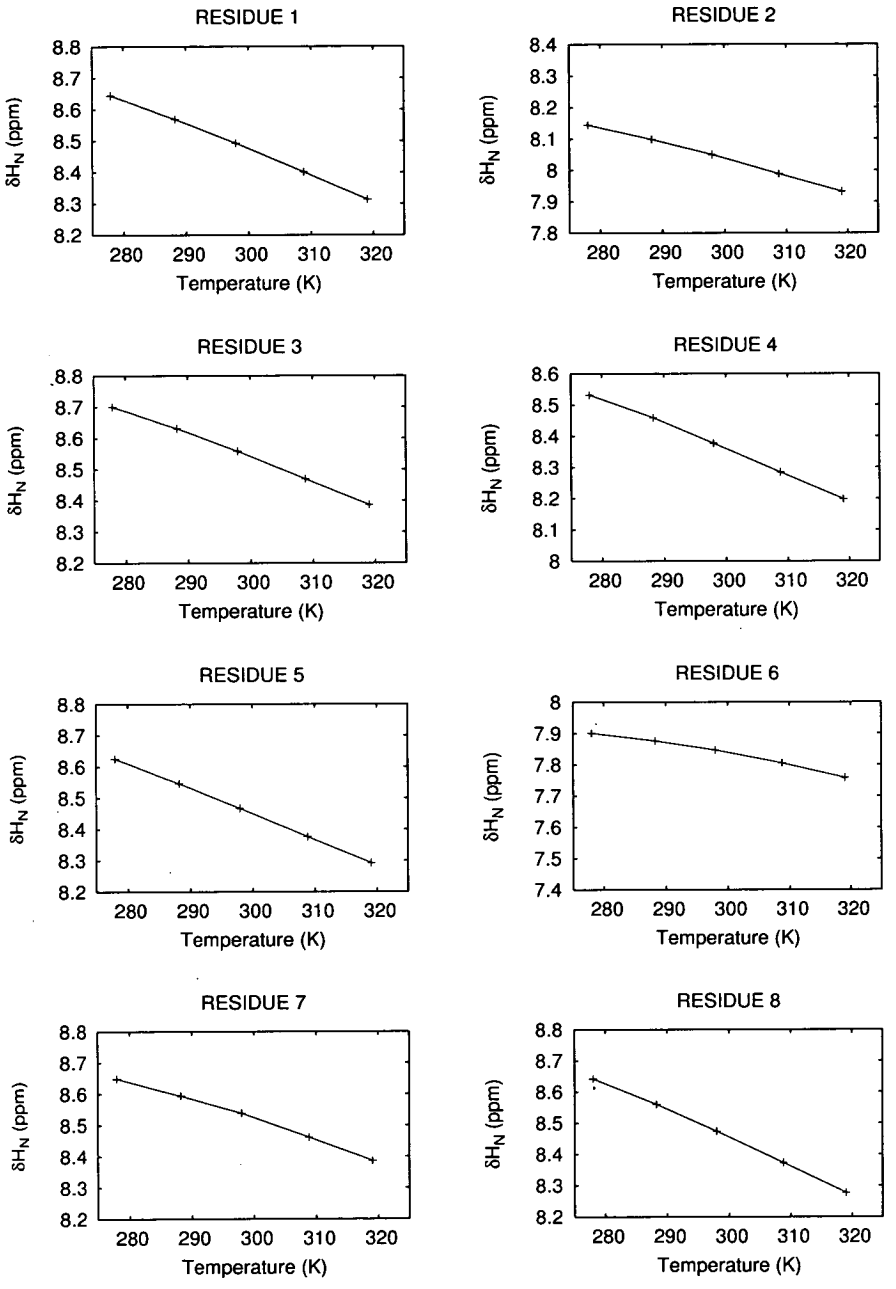


Figure G.4: Chemical shifts of H^N proton of peptide 4 as a function of temperature.

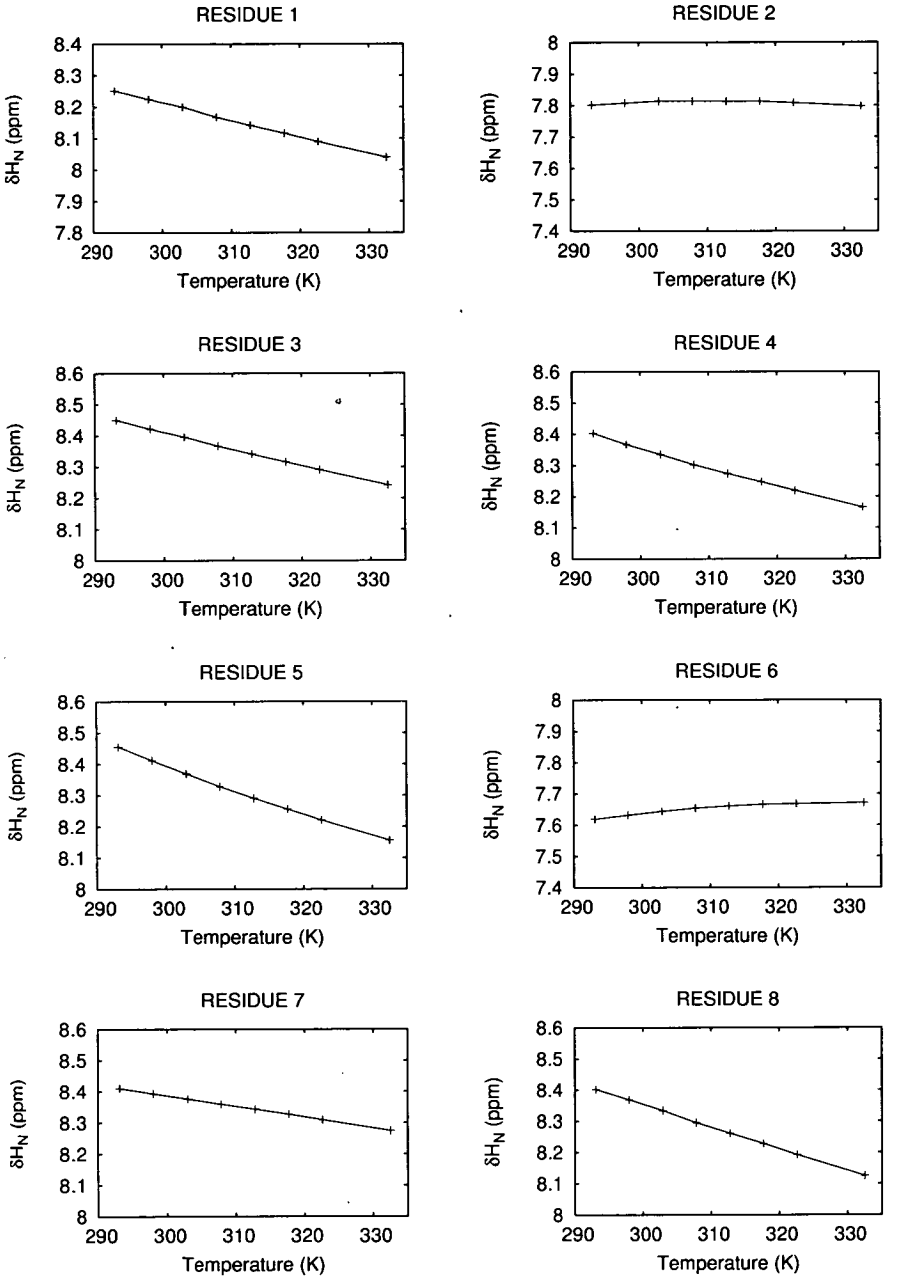


Figure G.5: Chemical shifts of H^N proton of peptide 2 in DMSO as a function of temperature

Bibliography

[1] Source : World Health Organization. http://www.who.int, 2006.

[2] D. Hanahan and J. Folkman. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. *Cell*, 86:353–364, 1996.

[3] R. van M and E. Voest. Angiogenesis in prostate cancer : its role in disease progression and possible therapeutic approaches. *Mol. Cell. Endocrinology*, 197:239–250, 2002.

[4] S. Madhusudan and A. Harris. Drug inhibition of angiogenesis. *Curr. opin. Pharm.*, 2:403–414, 2002.

[5] A. Jones and C. Fujiyama. Angiogenesis in urological malignacy: prognostic indicator and theraputic target. *BJU Int.*, 83:535–556, 1999.

[6] L. D. D'Andrea, A. Del Gatto, C. Pedone, and E. Benedetti. Peptide-based molecules in angiogenesis. *Chem. Biol Drug Des.*, 67(4):115–126, 2006.

[7] J. Folkman. Tumor angiogenesis: therapeutic implications. *N. Engl. J. Med.*, 285:1182–1186, 1971.

[8] P.C. Brooks. Role of integrins in angiogenesis. *Eur. J. Cancer*, 32 A(14):2423–2429, 1996.

[9] S. Strömblad and D. Cheresh. Cell adhesion and angiogenesis. *Trends in Cell Biol.*, 6:462–468, 1996.

[10] A. van der Flier and A. Sonnenberg. Function and interaction of integrins. *Cell and Tissue Research*, 305:285–298, 2001.

[11] S. Mousa. Anti-integrin as novel drug discovery targets: potential therapeutic and diagnosis applications. *Curr. opin. Chem. Biol.*, 6:534–541, 2002.

[12] K.M. Yamada and S. Even-Ram. Integrin regulation of growth factor receptors. *Nature Cell Biol.*, 4:75–76, 2002.

[13] M.J. Humphries. Integrin activation: the link between ligand binding and signal transduction. *Curr. opin. Cell Biol.*, 8:632–640, 1996.

[14] E.F. Plow, T.A. Haas, L. Zang, J. Loftus, and J.W. Smith. Ligand binding to integrins. *J. Biol. Chem.*, 275(29):21785–21788, 2000.

[15] P.N. Newham and M.J. Humphries. Integrin adhesion receptors: structure, function and implication for biomedicine. *Molecular Medicine Today*, 2:304–313, 1996.

[16] R.O. Hynes. Integrins: bidirectional, allosteric signalling machines. *Cell*, 110:673–687, 2002.

[17] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and J. Watson. *Molecular biology of the cell, 3rd edition.* Garland Publishing, 1994.

[18] J. Varner and D. Cheresh. Integrins and cancer. *Curr. opin. Cell Biol.*, 8:724–730, 1996.

[19] B. Eliceiri and D. Cheresh. The role of αv integrins during angiogenesis: insights into potential mechanism of action and clinicalc development. *J. Clinical Investigations*, 8:724–730, 1996.

[20] M.A. Horton. The $\alpha_v\beta_3$ integrin "vitronectin receptor". *Int. J. Biochem. Cell Biol.*, 29(5):721–725, 1997.

[21] R.L. Wilder. Integrin $\alpha_v\beta_3$ as a target for treatment of rheumatoid arthritis and related rheumatic diseases. *Ann. Rheum. Dis.*, 61:96–99, 2002.

[22] L.E. Reynolds, L. Wyder, J. Lively, D. Taverna, S.D. Robinson, X. Huang, D. Sheppard, R.O. Hynes, and K.M. Hodivala-Dilke. Enhanced pathological angiogenesis in mice lacking β_3 integrin or β_3 and β_5 integrins. *Nature Med.*, 8(1):27–34, 2002.

[23] R.O. Hynes. A reeavaluation of integrins as regulators of angiogenesis. *Nature Med.*, 9(9):919–921, 2002.

[24] P. Carmeliet. Integrin indecision. *Nature Med.*, 8:14–16, 2002.

[25] J-P. Xiong, T. Stehle, R. Zhang, A. Joachimiak, M. Frech, S. Goodman, and M. Arnaout. Crystal structure of the extracellular segment of integrin $\alpha_v\beta_3$ in complex with an Arg-Gly-Asp ligand. *Science*, 296:151–155, 2002.

[26] J-P. Xiong, T. Stehle, B. Diefenbach, R. Zhang, R. Dunker, D. Scott, A. Joachimiak, S. Goodman, and M. Arnaout. Crystal structure of the extracellular segment of integrin $\alpha_v\beta_3$. *Science*, 294:339–345, 2001.

[27] M. Dechantsreiter, E. Planker, B. Matha, E. Lohof, G. Hölzemann, A. Jonczyk, S. Goodman, and H. Kessler. N-methylated cyclic RGD peptide as highly active and selective $\alpha_v\beta_3$ integrin antagonist. *J. Med. Chem.*, 42:3033–3040, 1999.

[28] M.J. Humphries. Integrin Structure. *Biochem. Soc.*, 28(4):311–339, 2000.

[29] L. Marinelli, A. Lavecchia, K. Gottschalk, E. Novellino, and H. Kessler. Docking studies on $\alpha_v\beta_3$ integrin ligands : pharmacophore refinement and implication for drug design. *J. Med. Chem.*, 46:4393–4404, 2003.

[30] G. Bitan, L. Scheibler, D.F. Mierke, M. Rosenblatt, and M. Chorev. Ligand-integrin $\alpha_v\beta_3$ interaction determined by photoaffinity cross-linking: A challenge to the prevailing model. *Biochemistry*, 39:11014–11023, 2000.

[31] G. Tucker. Inhibitors of integrins. *Curr. opin. Pharm.*, 2:394–402, 2002.

[32] R. Kok, A. Schraa, E. Bos, H. Moorlag, S. Ásgeirsdóttir, M. Everts, D. Meijer, and G. Molema. Preparation and function evaluation of RGD-modified proteins as $\alpha_v\beta_3$ integrin directed therapeutics. *Bioconjugate chem.*, 13:128–135, 2002.

[33] K. Temming, R. Schiffelers, G. Molema, and R. Kok. RGD-based strategies for selective delivery of therapeutics and imaging agents to the tumour vasculature. *Drug Resistance Updates*, 8:381–402, 2005.

[34] W. Miller, R. Keenan, N. Willette, and M. Lark. Identification and *in vivo* efficacity of small-molecules antagonist of $\alpha_v\beta_3$ (the vitronectin receptor). *Drug Discovery Today*, 5(9):397–408, 2000.

[35] J.M. Goodman. *Chemical Application of Molecular Modelling.* Royal Society of Chemistry, 1998.

[36] A. Dal Pozzo, M. Ni, L. Muzi, R. de Castiglione, R. Mondelli, S. Mazzini, S. Penco, C. Pisano, M. Castorina, and G. Giannini. Incorporation of the unusual cα-fluoroakylamino acids into cyclopeptides : synthesis of arginine-glycine-aspartate (rgd) analogues and study of theur conformational and biological behaviour. *J. Med. Chem.*, 49:1808–1817, 2006.

[37] L. Belvisi, A. Bernardi, M. Colombo, L. Manzoni, D. Ptenza, C. Scolastico, G. Giannini, M. Marcellini, T. Riccioni, M. Castorina, P. LoGiudice, and C. Pisano. Targeting integrins : insights into structure and activity of cyclic RGD pentapeptide mimics containing azabicycloalkane amino acids. *Bioorg. Med. Chem.*, 14:169–180, 2006.

[38] N. Assat-Munt, X. Jia, P. Laakkonen, and E. Ruoslathi. Solution structures and integrin binding activities of an rgd petide with two isomers. *Biochemistry*, 40:2373–2378, 2001.

[39] E. Locardi, D. Mullen, R-E. Mattern, and M. Goodman. Conformations and pharmacophores of cyclic RGD containing peptides which selectively bind integrin $\alpha_v\beta_3$. *J. Peptide Science.*, 5:491–506, 1999.

[40] M. Pfaff, K. Tangemann, B. Muller, M. Gurrath, G. Muller, H. Kessler, R. Timpl, and J. Engel. Selective recognition of cyclic RGD peptides of NMR defined conformation by αIIβ3, $\alpha_v\beta_3$ and $\alpha v\beta 5$ integrins. *J. Biol. Chem.*, 269(32):20233–20238, 1994.

[41] R. Haubner, R. Gratias, B. Diefenbach, S. Goodman, A. Jonczyk, and H. Kessler. Structural and functional aspects of the RGD-containing cyclic pentapeptides as highly potent and selective integrin $\alpha_v\beta_3$ antagonists. *J. Am. Chem. Soc.*, 118:7461–7472, 1996.

[42] D. McDonald and P. Choyke. Imaging of angiogenesis: from microscope to clinic. *Nature Med.*, 9:713–725, 2003.

[43] Z. Su, G. Liu, S. Gupta, Z. Zhu, M. Rusckowski, and D. Hnatowich. In vitro and in vivo evaluation of a technetium-99m-labeled cyclic RGD peptide as a specific

marker of $\alpha_v\beta_3$ integrin for tumor imaging. *Bioconjugate Chem.*, 13:561–570, 2002.

[44] G. Sivolapenko, D. Skarlos, D. Pectasides, E. Stathopoulou, A. Milonakis, G. Sirmalis, A. Stuttle, N. Courtenay-Luck, K. Konstantinids, and A. Epenetos. Insight into integrin-ligand binding and activation from the first crystal structure. *Eur. J. Nucl. Med.*, 25(10):1383–1389, 1998.

[45] M. Janssen, W. Oyen, I. Dijkgraaf, L. Massuger, C. Frielink, D. Edwards, M. Rajopadhye, H. Boonstra anf F. Cortens, and O. Boerman. Insight into integrin-ligand binding and activation from the first crystal structure. *Cancer Research*, 62:16146–6151, 2002.

[46] E. Garanger, D. Boturyn, Z. Jin, P. Dumy, M. Favrot, and J. Coll. New multifunctional molecular conjugate vector for targeting, imaging and therapy of tumours. *Molecular Therapy.*, 12:1168–1175, 2005.

[47] X. Chen, S. Liu, Y. Hou, M. Tohme, R. Park, J. Bading, and P. Conti. MicroPET imaging of breast cancer α_v integrin expression with ^{6}4Cu labeled dimeric RGD peptide. *Molecular Imaging and Biology*, 6:350–359, 2004.

[48] R. Haubner. $\alpha_v\beta_3$ integrin imaging : a new approach to characterise angiogenesis? *Eur. J. Nucl. Med. Mol. Imag.*, 33:54–63, 2006.

[49] M. Werle and A. Bernkop-Schnürch. Strategies to improve plasma half life time of peptide and protein drugs. *Amino Acids*, 30:351–367, 2006.

[50] Cuthbertson, A.; Indrevoll, B.; Solbakken, M. Patent WO 03/006491, 2003.

[51] G. Rose, L. Gierash, and J. Smith. Turns in peptides and proteins. *Adv. Prot. Chem.*, 37:1–109, 1985.

[52] G. Wagner, A. Pardi, and K. Wüthrich. Hydrogen bond length and ^{1}H NMR chemical shifts in proteins. *J. Am. Chem. Soc.*, 105:5948–5949, 1983.

[53] T. Cierpicki and J. Otlewski. Amide proton temperature coefficients as hydrogen bond indicators in proteins. *J. Magn. Reson.*, 21:249–261, 2001.

[54] G. Casiraghi, G. Rassu, L. Auzzas, P. Burreddu, E. Gaetani, L. Battistini, F. Zanardi, C. Curti, G. Nicastro, L. Belvisi, I. Motto, M. Castorina, G. Giannini, and Claudio Pisano. Graffting aminocyclopentane carboxylic acids onto the RGD tripeptide sequence generates low nanomolar $\alpha_v\beta_3$ / $\alpha_v\beta_5$ integrin dual binders. *J. Med. Chem.*, 48:7675–7687, 2005.

[55] G. Merutka, H. Dyson, and P. Wright. Random coil ^{1}H chemical shifts obtained as a function of the temperature and trifluoroethanol concentration for peptide series GGXGG. *J. Biomol. NMR*, 5:14–24, 1995.

[56] V. Polshakov, T. Frenkiel, B. Birdsall, A. Soteriou, and J. Feeney. Determination of stereospecific assignments, torsion-angle constraints, and rotamer populations in proteins using the program AngleSearch. *J. Magn. Reson., series B*, 108:31–43, 1995.

[57] J.M. Schmidt. Conformation equilibria in polypeptides. II. Dihedral-angle distribution in antamanide based on three-bond coupling information. *J. Magn. Reson.*, 124:310–322, 1997.

[58] M. Piotto, V. Saudek, and V. Sklenar. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueuous solutions. *J. Biomol. NMR*, 2:661–665, 1992.

[59] T. Wang and A. Shaka. Water suppression that works. Exitation sculpting using arbitrary waveforms and pulse field gradients. *J. Magn. Reson.*, 112:275–279, 1995.

[60] J. Keeler. *Understanding NMR spectroscopy.* Wiley, 2005.

[61] D. Neuhaus and M. P. Williamson. *The Nuclear Overhauser Effect in Structural and Conformational Analysis.* Wiley-VCH, 2000.

[62] H. Günther. *NMR spectroscopy, basic principles, concepts, and applications in chemistry.* Wiley, 1995.

[63] J. Keepers and T. James. A theoretical study of distance determinations from NMR. Two-dimensional nuclear overhauser effect spectra. *J. Magn. Reson.*, 57:404–426, 1991.

[64] M. Nilges, J. Habazettl, A. Brünger, and T. Holak. Relaxation matrix refinement of the solution structure of squash trypsin inhibitor. *J. Mol. Biol.*, 219:499–510, 1991.

[65] M. Thrippleton and J. Keeler. Elimination of zero-quantum interferences in two-dimensional NMR spectra. *Angew. Chem. Int. Engl.*, 42:3938–3941, 2003.

[66] H. Desvaux and M. Goldman. A simple solution to decrease angular dispersion in off-resonance experiments. *J. Magn. Reson., series B*, 110:198–201, 1996.

[67] T. Maliavin, H. Desvaux, and M.-A Delsuc. Conditions for the measurement of quantitative off-resonance ROESY experiments. *Magn. Reson. Chem.*, 36:801–806, 1998.

[68] A. J. Shaka, C. J. Lee, and A. Pines. Iterative schemes for bilinear operators: application to spin decoupling . *J. Magn. Reson.*, 77:274–293, 1988.

[69] R.A.E Edden and J. Keeler. Development of a method for the measurement of long-range ^{13}C-^{1}H coupling constants from HMBC spectra. *J. Magn. Reson.*, 166:53–68, 2003.

[70] J.T. Titman, D. Neuhaus, and J. Keeler. Measurement of long-range heteronuclear coupling constants. *J. Magn. Reson.*, 85:111–131, 1989.

[71] S. Sheng and H. van Halbeek. Accurate and precise measurement of heteronuclear long-range couplings by a gradient-enhanced two-dimensional multiple-bond correlation experiment. *J. Magn. Reson.*, 130:296–299, 1998.

[72] D. Uhrín, V. Varma, and J. Brisson. A method for measurement of long-range heteronuclear coupling constants from 2D HMQC spectra. *J. Magn. Reson.*, 119:120–124, 1996.

[73] M. Kurz, P. Schmieder, and H. Kessler. HETLOC, an efficient method for determining heteronuclear long-range couplings with heteronuclei in natural abundance. *Angew. Chem. Int. Engl.*, 30:1329–1331, 1995.

[74] D. Uhrín, G. Batta, V. Hruby, P. Barlow, and K. Kövér. Sensitivity and gradient enhanced hetero (ω_1) half filtered TOCSY experiment for measuring long-range heteronuclear coupling constants. *J. Magn. Reson.*, 130:155–161, 1998.

[75] S. Goodman, G. Hölzemann, G. Sulyok, and H. Kessler. Nanomolar small molecules inhibitors for $\alpha_v\beta_6$, $\alpha_v\beta_5$ and $\alpha_v\beta_3$ integrins. *J. Med. Chem.*, 45:1045–1051, 2002.

[76] A. Hinchliffe. *Modelling Molecular Structures, second edition.* Wiley series in theoritical chemistry, 2000.

[77] J. Vinter and M. Gardner. *Modelling Molecular and drug design.* London, Macmillan, 1994.

[78] C. Schwieters, N. Tjandra J. Kuszewski, and G. Clore. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.*, 160:6–74, 2003.

[79] K. Wütrich. *NMR of proteins and nucleic acids.* Wiley-Interscience Publication, 1986.

[80] W. Boucher, Department of Biochemistry, and University of Cambridge. Azara, v2.7, 2003.

[81] P. J. Kraulis. ANSIG: A Program for the Assignment of Protein 1H 2D NMR spectra by Interactive Graphics. *J. Magn. Reson.*, 24:627–633, 1989.

[82] V. Sklenar, M. Piotto, R. Leppik, and V. Saudek. Gradient-Tailored Water Suppression for H-1-N-15 HSQC Experiments Optimized to Retain Full Sensitivity. *J. Magn. Reson.*, 102:241–245, 1993.

[83] D. Wishart, C. Bigam, J. Yao, F. Abilgaard, H. Dyson, E. Oldfield, J. Markley, and B.D. Sykes. H-1, C-13 and N-15 chemical-shift referencing in biomolecular NMR. *J. Biomol. NMR*, 6:135–140, 1995.

[84] A. T. Brünger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, , and G. L. Warren. Crystallography and NMR system : A new software suite for macromolecular structure determination. *Acta crystallographica*, D(54):905–921, 1998.

[85] J. P. Linge and M. Nilges. Influence of non-bonded parameters on the quality of the NMR structures: A new force field for NMR structure calculation. *J. Biomol. NMR*, 13:51–59, 1999.

[86] Allen F. H., S. Bellard, Brice M. D., Cartwright B. A., Doubleday A. an d Higgs H., Hummelink T., B. G. Hummelink-Peters, Kennard O., Motherwell W. D. S., Rodgers J. R., and Watson D. G. The Cambridge Crystallographic Data Centre: computer-based search, retrieval, analysis a nd display of information. *Acta crystallographica*, B(10):2331–2339, 1979.

[87] G. Kleywegt, K. Henrick, E. Dodson, and D. van Aalten. Pound-wise but penny-foolish - How well do micromolecules fare in macromolecular refinement ? *Structure*, 11:1051–1059, 2003.

[88] M. Nilges and S. O'Donoghue. Ambiguous NOEs and automated NOE assignment. *Progess in NMR spectroscopy*, 32:107–139, 1998.

[89] M. Nilges, G.M. Clore, and A.M. Gronenborn. Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms. *FEBS letters*, 229:129–136, 1988.

[90] M. Nilges, G.M. Clore, and A.M. Gronenborn. Determination of thre-dimensional structures of proteins from interproton distance data by hybrid distance geometry-dynamical simulated annealing calculations. *FEBS letters*, 239:317–324, 1988.

[91] J.P. Linge, M.A. Williams, C.A.E.M. Spronk, A.M.J.J. Bonvin, and M. Nilges. Refinement of protein structures in explicit solvent. *Structure, function and genetics*, 50:496–506, 2003.

[92] X. Liem, B. Dang, and M. Pettitt. Simple intramolecular model potentials for water. *J. Phys. Chem.*, 91:3349–3354, 1987.

[93] H. Liu, F. Müller-Plathe, and W. van Gunsteren. A molecular dynamics simulation study with a combined quantum mechanical and molecular mechanical potential energy function: solvent effects on the conformational equilibrium of dimethoxyethane. *J. Chem. Phys.*, 102:1722–1730, 1995.

[94] G. Morris, D. Goodsell, R. Halliday, R. Huey, W. Hart, R. Belew, and A. Olson. Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function. *J. Comput. Chemistry*, 19, 1998.

[95] J. Gasteiger and M. Marsili. Iterative partial equalization of orbital electronegativity- A rapid access to atomic charges. *Tetrahedron*, 36:3219–3228, 1980.

[96] M. Pellegrini, S. Liehr, A. Fisher, P. Laub, B. Cooperman, and D. Mierke. Structure-based optimization of peptide inhibitors of mammalian ribonucleic reductase. *Biochemistry*, 39:12210–12215, 2000.

[97] S. Chi, B. Olivera D. Kim, J. McIntosh, and K. Han. NMR structure determination of α-conotoxin BuIA, a novel neuronal nicotinic acetylcholine receptor antagonist with an unusual 4/4 disulfide scaffold. *Biochem. Biophys. Res. Comm.*, 349:1228–1234, 2006.

[98] B.L. Marquez, W.H. Gerwick, and R.T. Williamson. Survey of NMR experiments for the determination of ^{n}J(C,H) heteronuclear coupling constants in small molecules. *Magn. Reson. Chem.*, 39:499–530, 2001.

[99] H. Kessler, A. Müller, and H. Oschkinat. Differences and sums of traces within, COSY spectra (DISCO) for the extraction of coupling constants: Decoupling after the measurement. *Magn. Reson. Chem.*, 23:844–852, 1985.

[100] P. Schmieder and H. Kessler. Determination of the ϕ angle in a peptide backbone by NMR spectroscopy with a combination of homonuclear and heteronuclear coupling constant. *Biopolymers*, 32:435–440, 1992.

[101] A.C. Wang and A. Bax. Determination of the backbone dihedral angle ϕ in human ubiquitin from reparametrized empirical Karplus equations. *J. Am. Chem. Soc*, 118:2483–2494, 1996.

[102] R. Waylishe and T. Schaefer. INDO MO calculations of the conformational dependence of the vicinal ^{13}C H spinspin coupling constant in propane. *Can. J. Chem*, 50:2710–2712, 1972.

[103] R. Weisemann, H. Ruterjans, H. Schwalbe, J. Schleucher, W. Bermel, and C. Griesinger. Determination of H^N, H^α and $H^N C'$ coupling constants in ^{13}C , ^{15}N labeled proteins. *J. Biomol. NMR*, 4:231–240, 1994.

[104] P. Padrta and V. Sklenar. Program MULDER - A tool for extracting torsion angles from NMR data. *J. Biomol. NMR*, 24:339–349, 2002.

[105] G. C. K. Roberts. *NMR of macromolecules, a pratical approach.* Oxford University Press, 1993.

[106] H. Kessler, C. Griesinger, and K. Wagner. Peptide conformations. 42. Conformation of the side chains in peptides using heteronuclear coupling constants obtained by two-dimension NMR spectroscopy. *J. Am. Chem. Soc*, 109:6927–6933, 1987.

[107] W. Schmitt, G. Zanotti, T. Wieland, and H. Kessler. Conformation of different S-deoxo-Xaa-amaninamide analogues in DMSO solution determined by NMR spectroscopy. Strong CD effects induced by βI, βII conformation change. *J. Am. Chem. Soc.*, 118:4380–4387, 1996.

[108] K. Kövér, D. Jiao, S. Fang, and V. Hruby. Conformational properties of the unnatural amino acid β-methylphenylalanine in a linear octapeptide system; correlations of ^{13}C -NMR chemical shifts with the side-chain stereochemistry of the amino acid residues. *J. Org. Chem.*, 59:991–998, 1994.

[109] S. Grdadolnik, D. Mierke, G. Byk, I. Zeltser, C. Gilon, and H. Kessler. Comparison of the conformation of active and non-acitve backbone cyclic analogs of substance P as a tool to elucidate features of bioactive conformation: NMR and molecular dynamics in DMSO and water. *J. Med. Chem.*, 37:2145–2152, 1994.

[110] A.J. Fischman, D.H. Live, H.R. Wyssbrod, W.C. Agosta, and D. Cowburn. Torsion angle in the cysteine bridge of oxytocin in aqueous solution. Measurements of circumjacent vicinal coupling between ^{1}H ^{13}C and ^{15}N. *J. Am. Chem. Soc.*, 102(8):2533–2539, 1980.

[111] A. Fischman, H. Wyssbrod, W. Agosta, and D. Cowburn. Heteronuclear vicinal coupling constants and site-specific isotopic substitution in the investigation of rotational isomerism in leucine. *J. Am. Chem. Soc.*, 100:54–58, 1978.

[112] M. Kraszni, Z. Szakacs, and B. Noszal. Determination of rotamer populations and related parameters from NMR coupling constants : a critical review. *Anal. Bioanal. Chem.*, 378:1449–1463, 2004.

[113] Hyperchem, release 7.5. ©2003,Hypercube Inc.

[114] G.M. Clore and A.M. Gronenborn. New methods of structure refinement for macromolecular structure determination by NMR. *Proc. Nat. Acad. Sci.*, 95:5891–5898, 1998.

[115] R. Koradi, , M. Billeter, and K. Wüthrich. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics*, 14:51–55, 1996.

[116] J-P. Ryckaert, G. Ciccotti, and H. Berendsen. . *J. Comput. Phys.*, 23:327–341, 1977.

[117] C. D. Schwieters and G.M. Clore. The VMD-XPLOR Visualization Package for NMR Structure Refinement,. *J. Magn. Reson.*, 149:239–244, 2001.

[118] S. Oard and B. Karki. Mechanism of β-purothionin antimicobial peptide inhibition by metal ions: molecular dynamics study. *Biochem. Phys.*, 121:30–43, 2006.

[119] K. Bisetty, F. Corcho, J. Canto, H. Kruger, and J. Perez. A molecular dynamics study of the pentacyclo-undecane cage amino acid tripeptide. *J. Mol. Struct. : THEOCHEM*, 770:221–228, 2006.

[120] S. Pfeiffer, D. Fushman, and D. Cowburn. Simulated and NMR-derived backbone dynamics of a protein with significant flexibility : a comparison of spectral densities for the βARK1 PH domain. *J. Am. Chem. Soc.*, 123:3021–3036, 2001.

[121] N. Krishna, D. Huang, J. Vaughn Jr, G. Heavner, and G. Goldstein. Proton nuclear magnetic resonance study if an active pentapeptide fragment of ubiquitin. *Biochemistry*, 20:3933–3940, 1981.

[122] T. Cierpicki, I. Zhukov, A. Byrd, and J. Otlewski. Hydrogen bonds in human ubiquitin reflected in temperature coefficients of amide protons. *J. Magn. Reson.*, 157:178–180, 2002.

[123] N.J. Baxter and M.P. Williamson. Temperature dependence of ^{1}H chemical shifts in proteins. *J. Biomol NMR*, 9:359–369, 1997.

[124] N.H. Andersen, J.W. Neidigh, S.M. Harris, G.M. Lee, Z. Liu, and H. Tong. Extracting information from the temperature gradients of polypeptide NH chemical shifts. 1. The importance of conformation averaging. *J. Am. Chem. Soc.*, 119:8547–8561, 1997.

[125] N.H. Andersen, C. Chen, T. Marschner, S. Krystek, and D. Bassolino. Conformational isomerism of endothelin in acidic aqueous media : a quantitative NOESY analysis. *Biochemistry.*, 31:1280–1295, 1992.

[126] E. Hutchinson and J. Thornton. A revised set of potentials for β-turn formation in proteins. *Protein science*, 3:2207–2216, 1994.

[127] T-A. Tran, R-H. Mattern, Q. Zhu, and M. Goodman. A novel RGD containing dodecapeptidomimetic which exhibits selective binding to the $\alpha_v\beta_3$ receptor. *Bioorg. Med. Chem. letters*, 8(7):997–1002, 1997.

[128] A. Bach, C. Eyermann, J. Gross, M. Bower, R. Harlow, P. Weber, and W. DeGrado. Structural studies of a family of high affinity ligands for GPIIb / IIIa. *J. Am. Chem. Soc.*, 116:3207–3219, 1994.

[129] H. Kessler. Conformation and biological activity of cyclicpeptides. *Angew. Chem. Int. Engl.*, 21:7461–7472, 1982.

[130] H. Li, A. Robertson, and J. Jensen. Very Fast Empirical Prediction and Rationalization of Protein pKa Values. *Proteins*, 61:704–721, 2005.

[131] N. Baker, D. Sept, S. Joseph, M. Holst, and J. McCammon. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Nat. Acad. Sci.*, 98:10037–10041, 2001.

[132] N. Moitessier, C. Henry, B. Maigret, and Y. Chapleur. Combining pharmacore search, automated docking, and molecular dynamics simulations as a novel strategy for flexible dockin. Proof of concept: docking of arginine-glycine-aspartic acid-like compounds into the $\alpha_v\beta_3$ binding site. *J. Med. chem.*, 47:4178–4187, 2004.

[133] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne. The Protein Data Bank. *Nucleic Acids Res.*, 28:235–242, 2000.