Chapter 1 General Introduction

1.1 Nuclear Magnetic Resonance (NMR)

1.1.1 The nuclear spin Hamiltonian

NMR is a technique for investigating the chemical and spatial structure of compounds, by exploiting a property, that some nuclei have, called *spin*. The nuclear spin is a quantum effect, and the associated quantum numbers I are multiple of 1/2. The number of quantum states is 2I+1. In a simplified model, we can imagine the nucleus with spin 1/2 spinning around an axis with only two possible orientations, *up* and *down*. Since it is a moving charge, the spinning nucleus generates a magnetic field. When there is no external magnetic field present, there is no energy difference between up and down spin states. If an external magnetic field B_0 is applied, then the spin state that gives rise to a nuclear magnetic field aligned to the external magnetic field has a different, lower, energy from the other spin state. The energy difference ΔE is small and falls in the range of radio frequencies (Figure 1.1).

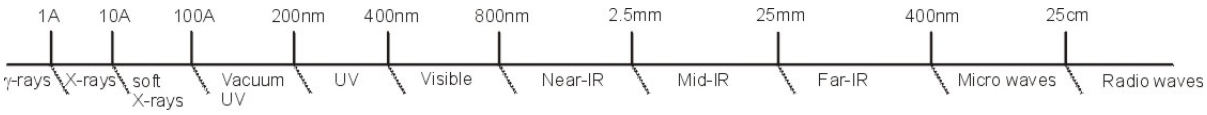

Figure 1.1 Range of frequencies exploited by different spectroscopies to investigate the matter.

The interaction between a nuclear spin **I** and a static magnetic field $\mathbf{B_0}$ is called *Zeeman interaction* and can be described by a second-rank tensor (3x3 matrix) $\hat{Z}$:

$$H_{zeeman} = \mathbf{I}\ \hat{Z}\ \mathbf{B_0} \qquad (1.1)$$

The energy separation between two different Zeeman states (*Larmor frequency*) is given by:

$$\Delta E = \hbar \gamma B_0 \tag{1.2}$$

and is directly proportional to the static magnetic field applied. The population differs for the two states and the ratio is given by Boltzmann distribution. The lowest-energy state, corresponding to a parallel orientation of nuclear spins along the magnetic field, is slightly more populated. As a result, the vector sum of all nuclear spins originates a macroscopic magnetic moment, named *magnetisation*.

By irradiation with a suitable energy, it is possible to convert a nucleus from one spin state to the other. Such irradiations with radiofrequencies of short duration are usually named *pulses*. The energy interaction between the spin I and the rf pulse is, analogously:

$$H_{rf} = \mathbf{I}\ \hat{Z}\ \mathbf{B_1} \tag{1.3}$$

The power of NMR spectroscopy relies on the fact that nuclei in different parts of the molecule experience different local magnetic fields according to the molecule's structure, and consequently *resonate* (adsorb energy) at different frequencies. It means that similar atoms in different environments, such as carbon-bonded or oxygen-bonded hydrogen atoms, show different frequencies. The absorbed frequency is called *chemical shift*. By defining a reference frequency, it is possible to express chemical shifts as dimensionless numbers, such that the value is not a function of the external magnetic field B_0, which allows direct comparison of spectra from machines with different magnetic field strengths and operating frequencies. Chemical shifts are a potent means of identifying different nuclei within the protein and discriminating between them. These effects of the environment on the nuclear spin depend on

the orientation of the sample with respect to the magnetic field and are therefore described by a second-rank *chemical shift* (or *shielding*) *tensor* $\hat{\Sigma}$:

$$H_{cs} = \gamma \mathbf{I} (\hat{\Sigma} \mathbf{B_0}), \tag{1.4}$$

where γ is the gyromagnetic ratio of the spin **I** and ($\hat{\Sigma}$ $\mathbf{B_0}$) represents the effective magnetic field experienced locally by the spin **I**.

This dependence of the chemical shift on the orientation is referred to as *chemical shift anisotropy* (CSA). Generally, it is possible to express the chemical shift tensor in a coordinate system, so that all off-diagonal elements vanish. In this principal axes system, the chemical shift tensor is fully described by the three diagonal elements – the principal components (δ_{11}, δ_{22} and δ_{33}) – and the three eigenvectors or Euler angles describing the orientation of the principal axes with respect to an arbitrary frame.

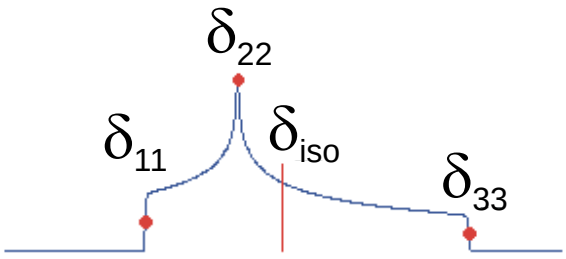

Figure 1.2 NMR Absorption line under anisotropic conditions.

Thus, δ_{11} corresponds to the direction of least shielding, with the highest frequency, while δ_{33} corresponds to the direction of highest shielding, with the lowest frequency:

$$\delta_{11} \geq \delta_{22} \geq \delta_{33} . \tag{1.5}$$

The isotropic values, δ_{iso}, are the average values of the principal components, and correspond to the centre of gravity of the line shape:

$$\delta_{iso} = \frac{(\delta_{11} + \delta_{22} + \delta_{33})}{3} . \qquad (1.6)$$

Under isotropic conditions (in solution, or, in the solid state, upon rapid sample spinning (§ 1.3)), an absorption line in correspondence to δ_{iso} will be observed and the Hamiltonian term simplifies into:

$$H_{cs} = \gamma \mathbf{I} (\sigma_{iso} \mathbf{B_0}), \qquad (1.7)$$

where σ_{iso} is now a scalar.

Nuclear spins within a molecule may interact with each other. Several interactions of different nature are usually simultaneously active. The *scalar coupling* (or *J* coupling) represents the interaction between nuclei relayed through the electrons in the chemical bonds. The strength of the scalar interaction between two spins **I** and **S** is represented by a coupling constant *J*:

$$H_J = \mathbf{I} J \mathbf{S} . \qquad (1.8)$$

Homonuclear (proton) couplings between two atoms distant by two bonds are largest (~15 Hz), smaller for three-bond couplings (5-10 Hz), and smallest for long-range couplings (~1 Hz). Heteronuclear coupling are substantially larger. For example, proton-nitrogen couplings are around 90 Hz and proton-carbon couplings are on the order of 140 Hz.

A second interaction between nuclear spins that takes place is the *dipolar coupling*. Nuclear spins behave indeed as magnetic moments, and consequently interact with each other through

space. The dipolar interaction is orientation-dependent and is therefore represented by a second-rank dipolar tensor $\hat{D}$:

$$H_D = \mathbf{I}\ \hat{D}\ \mathbf{S}. \qquad (1.9)$$

The dipolar tensor is a traceless tensor, that is, the sum of its diagonal elements is zero. Therefore, unlike the chemical shift, dipolar interactions are averaged to zero in solution due to the isotropic tumbling of molecules or in the solid-state under efficient sample spinning at the magic angle. However, the magnetic field that is generated from the dipolar coupling can have large effects on the spin-lattice relaxation rates of the spins (§ 1.2.1.1, *Nuclear Overhauser effect*).

Compared to other spectroscopies, a major advantage with NMR is the possibility to modify at will the Hamiltonian, with few restrictions, adapting it to the special requirements of the problem to be solved. The ease with which the nuclear spin Hamiltonian can be modified depends on the fact the nuclear interactions are very weak compared to the interactions exploited in other spectroscopic techniques, such as infrared spectroscopy or VIS-UV spectroscopy. In order to override an interaction, an alternative, competitive perturbation of the system has to be applied. To be effective, the energy of the perturbation has to be significantly larger than the interaction to manipulate. Examples of this are *spin decoupling* and *sample spinning*, vital tools in solution and solid-state NMR.

In general, by playing with the energy terms of the Hamiltonian, it is possible to design a large number of NMR experiments which provide different information: e.g., it is possible to obtain spectra which establish through-bond correlations between nuclei showing scalar coupling, or rather through-space correlations between nuclei showing dipolar coupling.

1.1.2 Magnetisation and pulses: 1D and multidimensional NMR spectra

Upon application of appropriate pulses, rotations of the magnetisation can be induced. A major property of the magnetisation is that its component in the plane perpendicular to the axis of the external magnetic field rotates around this axis with an angular velocity proportional to the Larmor frequency, that is the frequency ΔE of electromagnetic radiation required to excite transitions between Zeeman levels (Equation 1.2). If we now introduce a frame that rotates with this frequency, the magnetisation appears static. However, spins are also evolving under their chemical shift (Equation 1.7), which depends on the gyromagnetic ratio and the shielding by the chemical environment. The precession of the several components of the magnetisation along the static magnetic field constitutes a time-varying magnetic field, which in turn has the property to induce an electromotive force in a coil appropriately located close to the sample.

The measured intensity of the generated alternated current as a function of time is called FID (Free Induction Decay). In the case of a protein, the bulk magnetic moment is originated by nuclear spins with in general different chemical shifts. Hence, the FID contains all different resonance frequencies of the nuclei, corresponding to their chemical shifts. By Fourier transformation[a], the FID can be converted into a frequency function, that is, the NMR spectrum. To study biopolymers like proteins, DNA and RNA, 1D NMR spectroscopy is unable to resolve the frequency of the individual nuclear spins. Hence, 2D, 3D and even 4D spectroscopical techniques have been developed to increase resolution. Every 2D experiment, for example, can be described with a simple basic scheme, consisting of a *preparation* period, an *evolution* period t1 (during which the spins are labeled according to their chemical shift), a *mixing* period where spins are correlated to each other, and finally a *detection* period. For

[a] Fourier transformation is an important mathematical operation, which allows transforming time domain data into the frequency domain and vice versa.

measuring a 2D spectrum, many FIDs are acquired for incremented values of the t1 delay (*evolution period*) to generate a second frequency dimension. The recorded FIDs are then Fourier transformed with respect to both t1 and t2 (Figure 1.3). Signal with two different frequency coordinates (*cross-peaks*) indicate a correlation between two nuclei. The length of the mixing period is commonly referred to as *mixing time*. The intensity of each cross-peak as a function of the length of the mixing time represents the *bild-up curve* for the cross-peak.

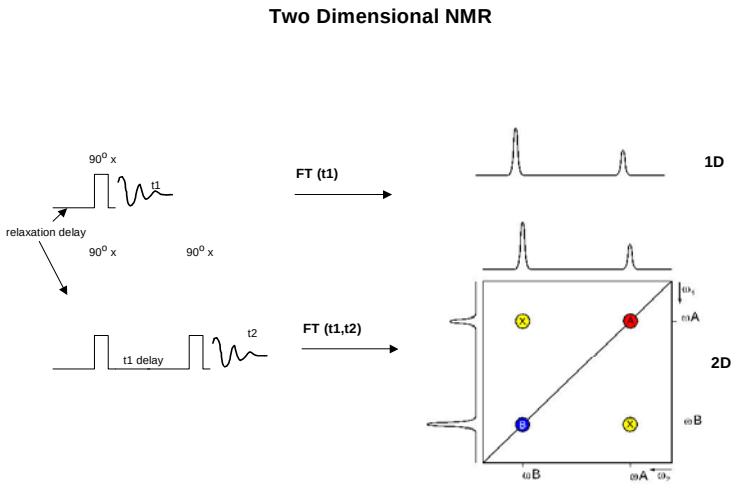

Figure 1.3 Principles of 2D NMR spectroscopy.

In 3D or 4D spectra, the 2D peaks are dispersed along one or two more orthogonal axes using the chemical shifts of one or two bonded heteronuclei.

1.2 Solution NMR

Over the last two decades, solution NMR has become a major technique in structural biology. It has proven to be a powerful technique to investigate protein structures in solution at atomic definition. Compared to X-ray crystallography, solution NMR allows not only to investigate the structures of biopolymers in a nearly physiological environment, but also to determine

their dynamic properties. Hence, solution NMR is a vital tool also in protein biophysics, allowing for the study of protein-ligand and protein-protein interactions, catalysis, hydratation.

1.2.1 Protein structure determination *via* solution NMR

The commonly employed procedure for structure determination of proteins based on solution NMR consists of several consecutive steps, covering sample preparation, NMR data acquisition, peak-picking, resonance assignment, collection of distance- and additional structural restraints and, finally, structure calculation and structure refinement[1]. Till date, this is the most widely followed strategy for determination of protein structures in solution, despite several alternative promising approaches that mostly aim at omission of the cumbersome and biologically irrelevant resonance-assignment[2].

1.2.1.1 Collection of structural restraints

In solution NMR, molecular tumbling produces an isotropic Hamiltonian, were only isotropic chemical shifts and J couplings are left (§ 1.1.1). A plethora of 2D and 3D experiments is present in the literature designed for assigning each observed chemical shift to a nuclear spin within the protein. This procedure is referred to as *resonance assignment*. The assignment of the resonances is usually done in two different separate steps. First, resonances within each single amino acid are assigned. As a result, chemical shifts are grouped in different spin systems, one for each amino acid. Second, the different spin systems are connected with each other by exploiting scalar couplings between heteronuclei in the backbone, or, alternatively, trough-space correlations between protons[1]. COSY and TOCSY[3] are the most commonly used homonuclear correlation experiments exploiting J couplings. HNCACB, HNCA and HNCO[1,4]

are routine experiments to establish correlations along the protein backbone *via* heteronuclear correlations.

The dipolar coupling vanishes in solution; nevertheless it affects nuclear spin relaxation. In particular, the relaxation of a spin is influenced *via* dipolar coupling by the presence of another close spin, whose spin population has been perturbed. These effects on spin relaxation are called Nuclear Overhauser effect[5-7], and have the important property of depending on internuclear distances, hence they provide a way to measure them. NOESY is the 2D solution NMR experiment, which directly exploits this effect to correlate nuclei which are close in space (distance smaller than 5Å).

Using a first-order approximation, the volume V of each NOE may be expressed as[5]:

$$V = \langle r_{ij}^{-6} \rangle f(\tau_c). \tag{1.10}$$

It depends on the distance between the two nuclei i and j, r_{ij}, as well as on the rotational correlation time τ_c (describing the Brownian tumbling motion of the protein in solution). By measuring cross-peak intensities V, *distance restraints* can be derived. *Dihedral angle restraints* represent a second important group of restraints that is possible to derive from solution NMR spectra. Dihedral angles in the protein backbone (Figure 1.4) influence directly the three-bond scalar couplings ^{3}J constants *via* Karplus equation[8]:

$$^3J(\tau) = A\cos^2(\tau) + B\cos(\tau) + C\,. \tag{1.11}$$

The constants A, B and C change depending on which dihedral angle (ϕ, ψ or χ_1) the letter τ stands for. Traditionally, ϕ angles are obtained by measurement of $^3J_{H^\alpha H^N}$ constants, whereby χ_1 angles from $^3J_{H^\alpha H^\beta}$ constants. Hence, after measuring J couplings, it is possible to restrain

the backbone torsion angles. Alternatively, dihedral angle restraints can be obtained exploiting the information contained in the secondary chemical shifts of the heteronuclei in the backbone (N, C^{α} and CO), which in turn depend on ϕ and ψ.

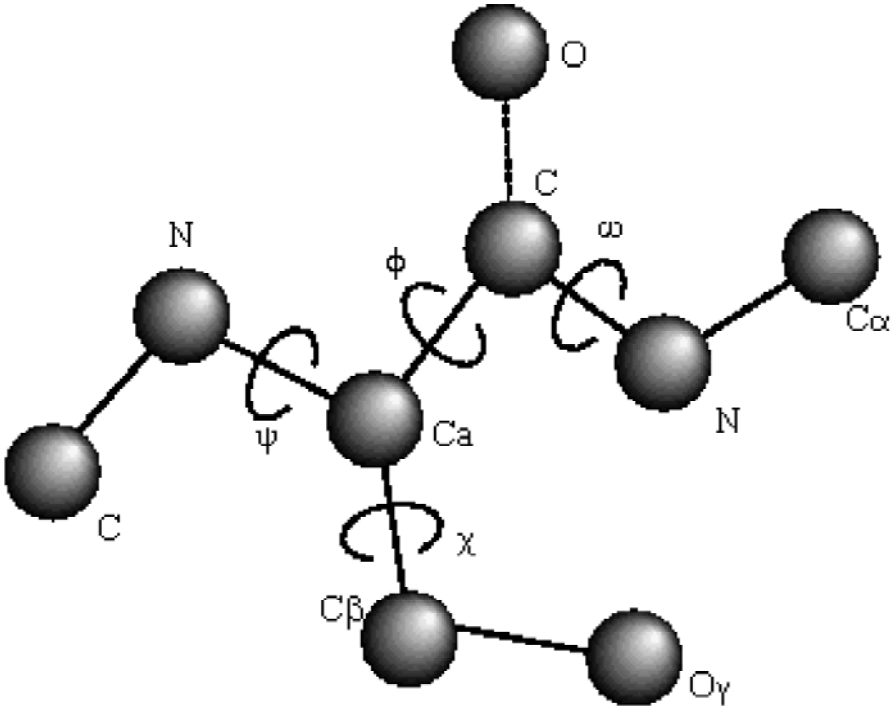

Figure 1.4 Dihedral angles of the protein backbone.

The *secondary chemical shift* is defined as the difference between the measured chemical shift and the chemical shifts in a random-coil (e.g., unstructured) protein and represents that component of the chemical shift that is induced by the tree-dimensional structure. Dihedral angle restraints can be derived automatically with the TALOS[9] programme, which relates the backbone secondary shifts with an internal database of high-resolution protein structures. Both strategies were used in this work to derive dihedral angle restraints.

Finally, with NMR it is possible to derive orientation angles θ of $N\text{-}H^{N}$ vectors with respect to the direction of the external magnetic field B_0 *via* measurement of Residual Dipolar Couplings (*RDC*). In these experiments (not carried out in this work), dipolar coupling is

partially reintroduced in solution by adding agents to the sample, which partially orient the proteins along a preferential direction.

The three different sources of structural information that can be obtained from NMR are indicated in Figure 1.5.

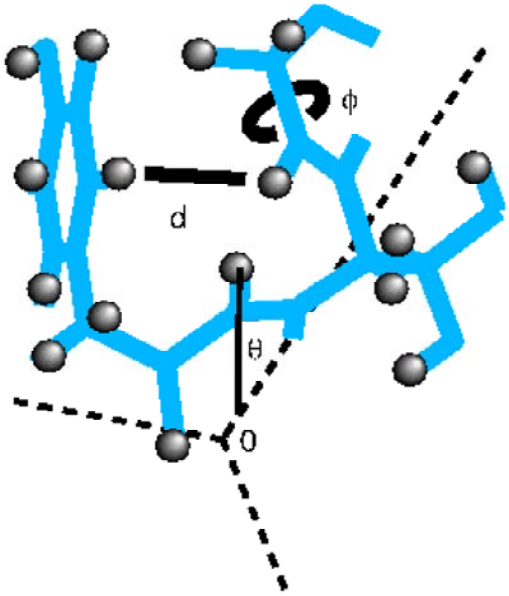

Figure 1.5 Experimental data from NMR: chemical shifts, short distances (d), torsion angles (ϕ) and orientation angles (θ).

1.2.1.2 Structure calculation *via* Simulated Annealing

Following the assignment of NOESY spectra, NOE-derived distance restraints (plus additional restraints like dihedral angle restraints, hydrogen bonds, disulphide bridges, if present) are used to calculate an ensemble of structures.

The ultimate goal of structure calculation algorithms is to operate an optimisation of the simultaneous agreement of an atomic model with the observed data and with *a priori* chemical information. Most algorithms used for structure calculation aim to find the global

minimum of a hybrid energy function E (*target function*), which includes *a priori* chemical knowledge of the system (the force field, defining bond lengths, bond angles, improper angles and non-bonded interactions) and experimental data (the structural restraints):

$$E = E_{chem} + E_{\exp} = \sum_i w_i E_i = \\ w_{\text{covalent}} E_{\text{covalent}} + w_{\text{angle}} E_{\text{angle}} + w_{\text{float}} E_{\text{float}} + w_{\text{vdW}} E_{\text{vdW}} + \\ w_{\text{NOE}} E_{\text{NOE}} + w_{\text{dihed}} E_{\text{dihed}} + \ldots \quad (1.12)$$

E_{chem} in Equation 1.12 contains energy terms for covalent bonds, bond angles, chirality, planarity and nonbonded repulsion. Electrostatic interactions and Van der Waals attraction forces are usually not included in structure calculation or refinement. Nonbonded repulsions are described by a quartic potential[10]. E_{exp} contains energy terms describing the experimental structural restraints.

The target function E is a function of many molecular parameters, most importantly of atomic coordinates. The large number of variables makes this function very complex and originates the so-called 'multiple minima problem': the target function contains many local minima In addition to the global minimum. The standard minimisation methods[11] tend to steer the system into local minima and frequently fail to reach the global minimum if the starting model is far away from the correct one. The sampling of a larger conformational space can be achieved by using Simulated Annealing (SA) optimisation techniques[11-13], which have the important property of enabling the system to overcome local minima (Figure 1.6). Monte Carlo-based and molecular dynamics-based SA have been used in structure prediction, molecular modelling, X-ray refinement and NMR structure determination for many years and have had a large impact in structural biology.

Structure calculation based on SA molecular dynamics consists in the integration of Newton's equation of motion:

$$\frac{d^2\vec{r}_i}{dt^2} = -\frac{c}{m_i}\frac{\nabla E}{\vec{r}_i}, \tag{1.13}$$

where $\vec{r}_i$ is the vector of Cartesian coordinates, m_i the mass of the atom i, c is a constant and ∇E is the gradient of the target function. The global minimum of the target function E is searched by reducing the temperature of the system after a high-temperature phase during a molecular dynamics simulation. In this context, the parameter 'temperature' has no physical meaning, but is simply a measure of the probability of the macromolecule to cross an energy barrier (i.e. its kinetic energy). Temperature control is performed by coupling the system with a thermal bath[14]. As a result, temperature coupling will cause "heat" (kinetic energy) to be added or removed from the system, as it is needed to maintain the temperature.

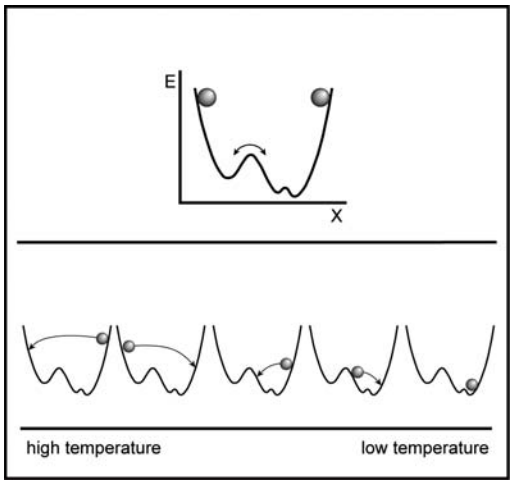

Figure 1.6 Compared to standard minimisation methods, SA allows the system to overcome local energy barriers, reducing the risk of the molecule to get trapped in local minima of the target function.

Energy terms for distance- and angle restraints are usually provided by specifying a lower and upper limit, L and U. During simulated annealing, a violation is generated whenever the

distance between the two atoms is not contained between the boundaries, leading to an increase in energy of the system. Typically, when restraints from manually assigned peaks are provided as input, energy terms for NOE-based distance restraints and dihedral angle restraints are present in the form of flat-bottomed parabolic functions (Figure 1.7). The flat bottom is delimited by L and U:

$$E_{dis\_res} = \begin{cases} (L-d)^2 & \text{if d < L} \\ 0 & \text{if L < d < U} \\ (d-U)^2 & \text{if U < d} \end{cases} \qquad (1.14)$$

$$E_{dih\_res} = \begin{cases} (L-\phi)^2 & \text{if } \phi < \text{L} \\ 0 & \text{if L} < \phi < \text{U} \\ (\phi-U)^2 & \text{if U} < \phi \end{cases}$$

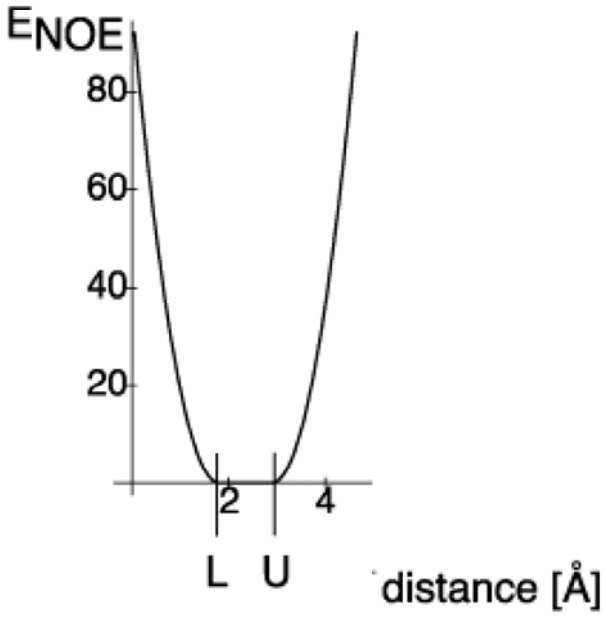

Figure 1.7 Standard "flat-bottom" potential used for NOE-derived distance restraints.

Structure calculations using iterative methods for automated cross-peak assignment (§ 1.4.1) are characterised by large violations of experimental distance restraints in the atomic model

during the early iterations. Hence, a linear function is substituted in Equation 1.14 for large violations (Equation 1.15), in order to avoid numeric instabilities arising from the high penalty for large violations (Figure 1.8).

$$E_{dis\_res} = \begin{cases} (L-d)^2 & \text{if d < L} \\ 0 & \text{if L < d < U} \\ (d-U)^2 & \text{if U < d < S} \\ \text{A(d - U)}^{-1} + B(d-U) + C & \text{if d > S} \end{cases} \qquad (1.15)$$

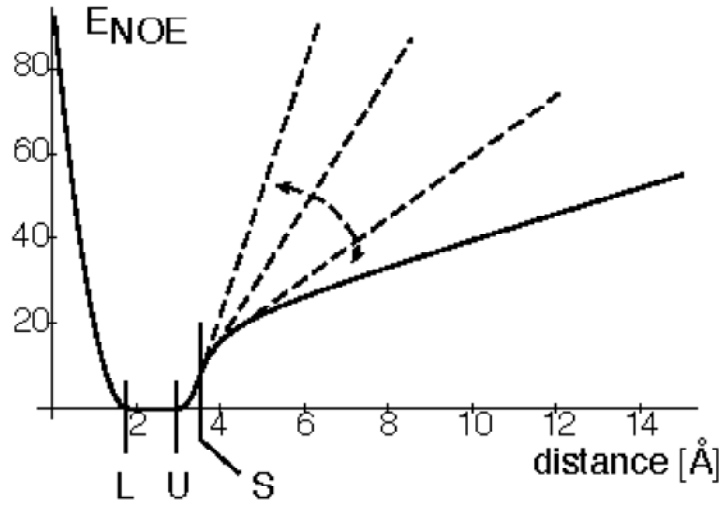

Figure 1.8 Soft potential for NOE-derived restraints used in iterative methods for automated cross-peak assignment.

1.3 Solid-State NMR

Recent developments in solid-state NMR methodology have led to an increasing number of pulse sequences and magic-angle spinning (MAS) techniques suitable for solid samples of proteins. As implied by the name, solid-state NMR is used for immobilised proteins, that is, samples where the reorientation of a molecule is very slow or non-existent, as is the case for

microcrystalline powders of soluble proteins or membrane proteins reconstituted in lipid bilayers. Therefore, solid-state NMR is a technique to investigate proteins whenever the dissolution of the sample is not possible or desirable and the X-ray approach is also not feasible. This may involve: 1) insoluble proteins, 2) proteins which aggregate in solution, 3) membrane-bound proteins, often insoluble or structurally altered in aqueous solution, in their synthetic or natural membrane environment. Compared to solution NMR, solid-state NMR has the disadvantage of a much lower resolution in terms of achievable line-widths relative to the chemical shift range. Conversely, molecular tumbling is not a band-narrowing mechanism in solid-state NMR; therefore, the size of the protein does not influence the line-width and, thus, is not an intrinsic limitation on resolution.

The chemical shift tensor and the dipolar interaction tensor all contain an orientation dependent factor:

$$(3\cos^2\theta - 1)\,. \tag{1.16}$$

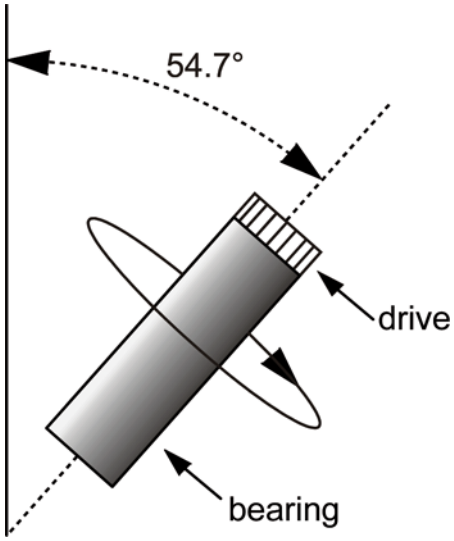

Figure 1.9 Principle of the magic-angle spinning (MAS) technique. The rotor containing the sample is tilted at 54.7° with respect to the external magnetic field.

Motional averaging, an intrinsic property of liquids, can be simulated by rapid sample spinning at the magic angle (θ_m = 54.733°) (Figure 1.9) relative to the magnetic field B_0, which leads to the vanishing of the anisotropy factor contained in Equation 1.16. Anisotropic interactions, like dipolar couplings and the anisotropic part of the chemical shift can be averaged out if the spinning is sufficiently rapid. Thus, also for solid–state samples it is possible to obtain, with some limitations, an isotropic Hamiltonian, giving rise to relatively narrow signals with absorption line shapes. In practice, to completely average out the anisotropies, the spinning rate has to be significantly larger than the interaction strength (in Hz). This can be easily achieved with interactions such as homonuclear and heteronuclear dipolar couplings not involving protons (< 10 kHz). At intermediate stages of averaging, a pattern of *sidebands* spaced at multiples of the spinning rate can be observed (Figure 1.10, II).

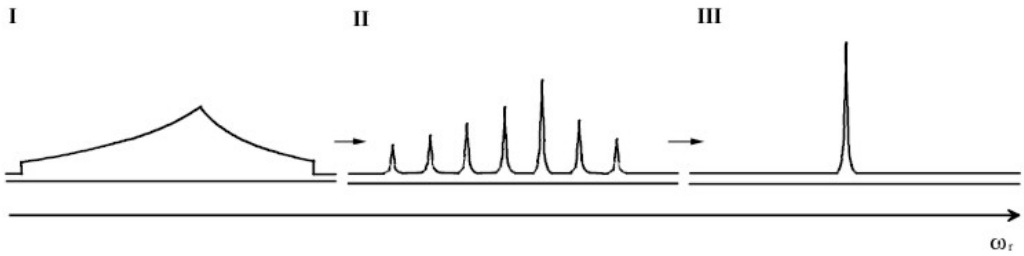

Figure 1.10 Solid-state NMR absorption line under I) anisotropic, II) intermediate III) isotropic conditions.

This is the case for stronger interactions such as ^{1}H-X dipolar couplings (up to 25 kHz) and homonuclear ^{1}H-^{1}H dipolar couplings (up to 80 kHz). Since spinning rates much larger than the ^{1}H-X dipolar coupling cannot be easily achieved without damaging the sample, high-power heteronuclear ^{1}H-decoupling is commonly used during the acquisition or evolution periods of ^{13}C and ^{15}N to resolve the isotropic chemical shifts of these nuclei. In general, the spinning rate should be adjusted such that the sidebands do not overlap with the centre-bands, in order to avoid the unwanted rotational resonance-condition and hence the re-introduction of dipolar couplings.

In biological solid-state NMR, homonuclear and heteronuclear correlations between ^{13}C and ^{15}N are most commonly measured. Such correlations are mostly mediated by dipolar interactions. Transfer pathways can involve either direct dipolar couplings between two spins (RFDR spectroscopy[15]), or indirect couplings *via* the protons, as in PDSD spectra[16], where the magnetisation transfer is 'driven' by the protons in a spin-diffusion process. For the first case, the transfer rate for cross peaks generally relates to the strength of the dipolar interaction, which scales with the inverse cube of the distance:

$$\Gamma_{ij} = \frac{1}{r_{ij}^3} A_{ij} . \tag{1.17}$$

In contrast to transfers mediated by direct dipolar interactions, the transfer rate Γ for cross peaks in PDSD spectra is proportional to the sixth power of the distance between the two nuclei[17]:

$$\Gamma_{ij} = \frac{1}{r_{ij}^6} B_{ij} . \tag{1.18}$$

The functions A_{ij} and B_{ij} in Equations 1.17 and 1.18 describe all the numerous effects which contribute to cross-peak intensities in MAS solid state NMR. First, mobile parts of the protein, like side-chains or flexible loops and termini, show different magnetisation transfer properties than rigid segments, since mobility interferes with and partly averages dipolar couplings. This leads to a much stronger variation in coherences for the restraint-delivering mixing process than in solution NMR. Mobility also has an effect on initial excitation of the low-γ nuclei by cross-polarization. As a result, it is more difficult to achieve a uniform excitation of all relevant spins in solid-state NMR than in liquid-state NMR. Further

complications may arise through offset-dependent transfer processes, interference with heteronuclear decoupling schemes and sample heterogeneity. Finally, even with a reduced ^{13}C-labelling of the samples (§ 1.3.1.1), the transfer efficiency between two spins can still be largely affected by the coupling to other nearby spins, a problem often referred to as *dipolar attenuation*[18,19]. Compared to solution NMR (Equation 1.10), these effects make it very difficult to find a satisfying uniform relation between experimental volumes and distances.

1.3.1 Protein structure determination *via* solid-state NMR

Recently, a simple methodology was developed in our group to investigate protein structures from solid-state MAS NMR data[20,21]. This approach bears many parallels with the strategy used for structure determination of proteins *via* solution NMR (§ 1.2.1).

1.3.1.1 Collection of structural restraints

For structure determination of proteins by MAS NMR it is first necessary to perform a resonance assignment[22]. Subsequently, ^{13}C-^{13}C correlation PDSD spectra are measured. By assigning cross-peaks, a large number of inaccurate ^{13}C-^{13}C distance restraints in the range of 2-7.5 Å are derived. The measurement of long-range ^{13}C-^{13}C correlations in the solid state is, however, not trivial due to dipolar attenuation problems. Due to strong dipolar couplings between chemically bonded nuclei, the measurement of weaker dipolar interactions between carbons that are farther apart is hampered. To solve this problem, two different strategies were employed: first, special pulse sequences were used which operate broad-band recoupling of the carbon spins[23-25]. Secondly, instead of using uniformly ^{13}C-labelled samples, "spin-diluted" samples were employed. These were obtained by growing the bacteria expressing the protein on a medium containing [1,3-^{13}C]-glycerol or [2-^{13}C]-glycerol as sole carbon source[20].

According to the labelling pattern observed, the twenty amino acids can be divided into two groups of ten, labelled A and B, as shown in Figure 1.11.

The amino acids of group A are characterised by either ~100% or ~0% ^{13}C-labelled carbon positions, whereas for those of group B there is fractional labelling, which arises from the fact that the residues exist as small ensembles of isotopomers. The percentages of labelling of the different carbon positions are, however, well known and do not depend on the particular protein studied, as they exclusively depend on the amino acid synthesis pathways of the bacterium used for the expression of the sample. In all isotopomers, the number of labelled carbons is strongly reduced, and, with few exceptions, the simultaneous labelling of two directly bonded carbons does not occur. This biosynthetically site-directed "spin-dilution" has several advantages: i) the simplification of the network of dipolarly coupled nuclei can be exploited to circumvent the problems due to dipolar attenuation and to measure weak long-range interactions important to define the structure; ii) the signal line-width is reduced because of the suppression of the one-bond homonuclear J coupling between chemically bonded carbons (this coupling is not averaged to 0 by magic angle spinning and is responsible for significant line broadening in uniformly labelled samples); iii) the reduced number of peaks in the spectra and the fewer assignment options simplify the cross-peak assignment procedure.

In solid-state spectra, distances are difficult to extract from peak volumes; however, a semi-quantitative estimation of the distance can be achieved by exploiting the fact that long-range correlations show a different bild-up curve (§ 1.1.2) compared to short-range correlations and appear in the spectrum only when longer mixing times are applied. Hence, prior to structure calculation, the cross-peaks from a series of 2D experiments measured with increasing mixing time were categorised in different distance classes based on a set of reference distances. Finally, a set of dihedral angle restraints can be obtained with the TALOS programme (§ 1.2.1.1).

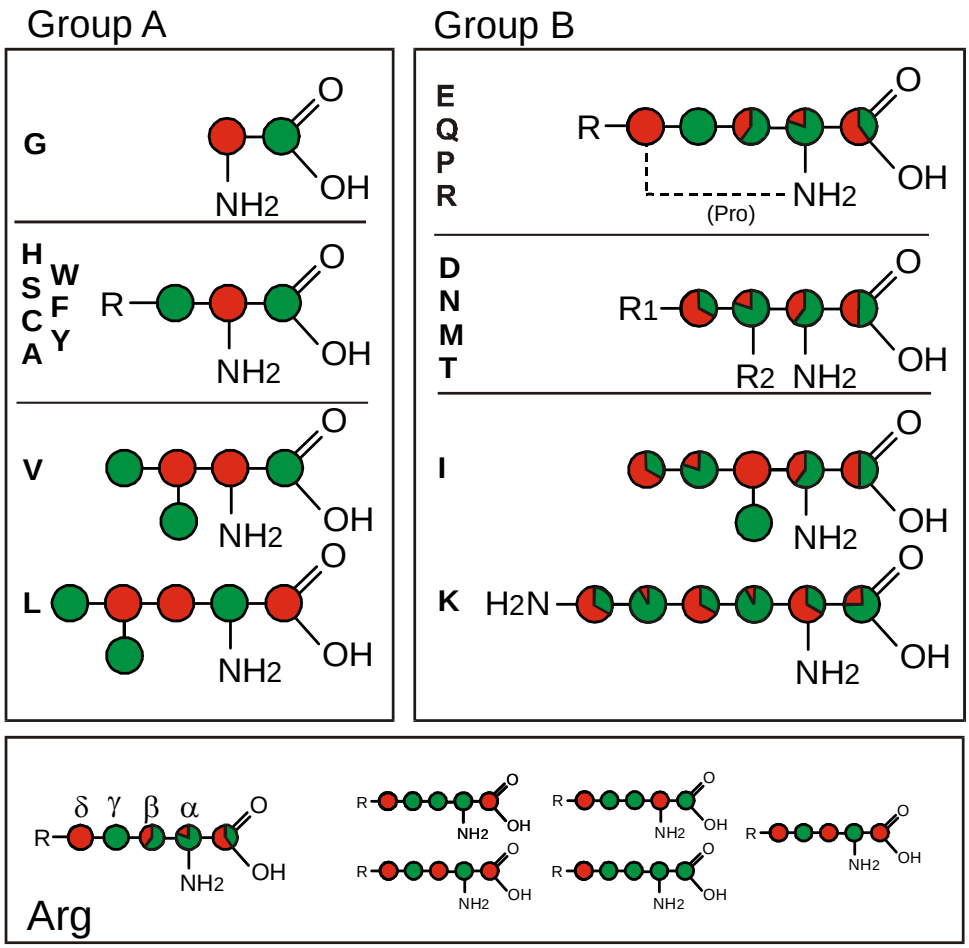

Figure 1.11 Labelling patterns for the different α-spectrin SH3 domain preparations used in these studies. Schematic representation of the effective ^{13}C-enrichment for the different residues, as obtained by protein expression in *E. coli* BL21 (DE3). The green colour corresponds to the degree of ^{13}C-labelling pattern obtained by growth on [1,3-^{13}C]-glycerol; the opposite labelling pattern, obtained by growth on [2-^{13}C]-glycerol, is represented in red. There are two groups of labelling (A and B). In group A, the various carbon sites are either ~100% or ~0% labelled, in group B the residues have fractional labelling. This fractional labelling is the result of the production of mixtures of residues with different labelling patterns, as illustrated for the arginine residue at the bottom of the figure. The labelling percentages for the residues in group B are represented using red/green pie diagrams. These percentages were estimated from solution NMR studies.

1.3.1.2 Structure Calculations of the α-spectrin SH3 domain

Once a sufficient number of long-range distance restraints are measured, a simulated annealing protocol is applied to calculate the structures, as in solution NMR (§ 1.2.1.2). In the case of the SH3 domain, structure calculations were carried out with the CNS[26] programme, one of the most widely used molecular dynamics-based simulated annealing protocol. An additional difficulty in structure calculations with solid-state data is represented by the presence of inter-molecular correlations in the spectra. In a standard NMR structure calculation of a monomeric protein in solution, only intra-molecular restraints are provided as input. Conversely, in solid-state NMR spectra it can be expected that due to the dense protein packing inter-molecular correlations occur as well. It is not *a priori* clear which correlations are intra-molecular and which are inter-molecular. The inter-molecular peaks may generate distance restraints which disturb the convergence of the structure calculation in a way similar to incorrectly assigned cross-peaks, 'peak-picked noise' and artefacts.

All CNS structure calculations using the SH3 domain failed if inter-molecular correlations were included in the peak list. Hence, a discrimination of inter- and intra-molecular contacts on an experimental basis was required: inter-molecular cross-peaks were identified by recording 2D PDSD spectra on isotopically diluted samples, made by mixing labelled and unlabelled proteins in a ratio of 1:4. When the inter-molecular correlations were removed, convergence could be achieved. Good results were obtained with distance restraints derived from 2D spectra only[20] (Figure 1.12a). The inclusion of restraints from three-dimensional spectra in the calculation led to significant improvements in accuracy and precision of the structures[21] (Figure 1.12b).

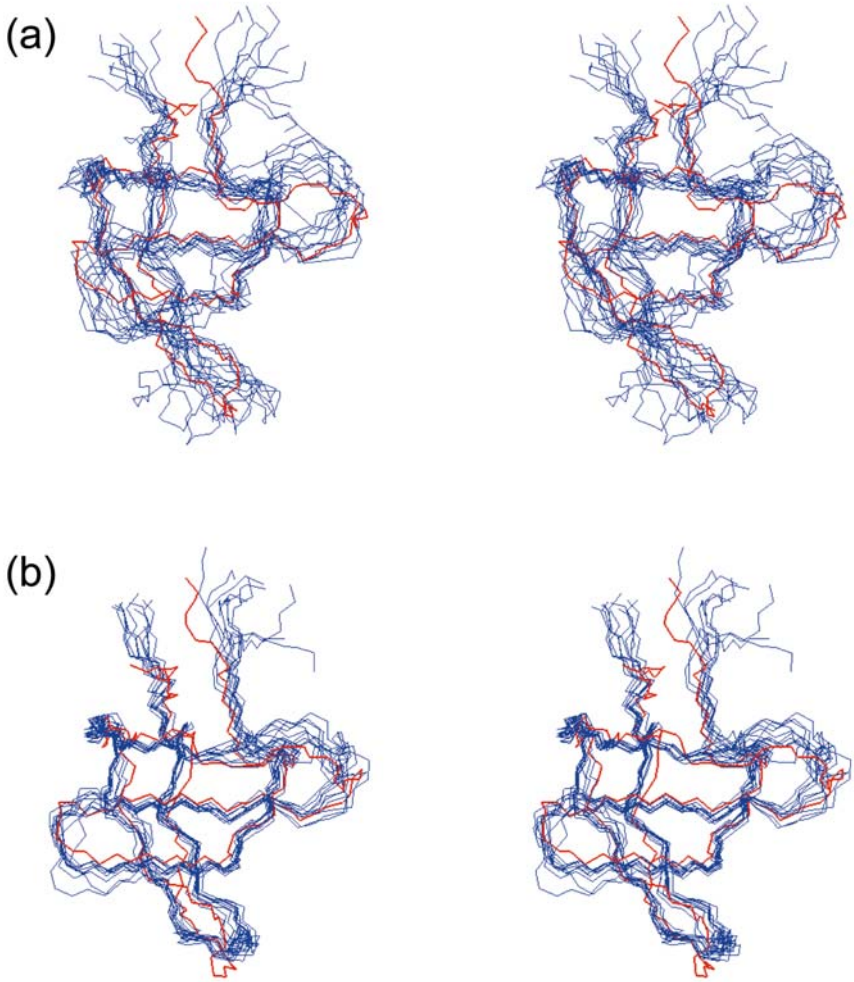

Figure 1.12 Stereo-view of the SH3 domain, as determined by solid-state NMR after manual assignment of the spectra, with only 2D (a) and both 2D and 3D (b) spectroscopy. The ensemble of 12 lowest-energy structures (blue) are overlaid with the reference X-ray structures (red).

1.4 ARIA

1.4.1 Unambiguous results from ambiguous data

NOESY spectra provide a large number of distances between protons which are close in space (not farther then 5 Å, approximately). Unfortunately, the interpretation of NOESY spectra is never straightforward. Signal degeneracy leads to several assignment options for many peaks.

With larger proteins, these effects are usually so dramatic, that a complete manual interpretation of the spectra is hampered. Consequently, the automation of this step has become an issue of great interest in the last years.

Until recently, software to calculate structures used to accept exclusively distance restraints from unambiguously assigned peaks, hence a large fraction of the distance information contained in the spectrum remained unexploited. In the case of larger proteins, however, most of the NOEs are ambiguous, and this approach does not usually allow deriving sufficient structural information from the spectrum to even calculate a preliminary structure.

In the past years, several semi-automated iterative approaches have been proposed to assign NOESY spectra[27-31]. Today, several software packages exist for a more or less fully-automated NOE assignment: ARIA[32-36], CANDID[37], DYANA[38], KNOWNOE[39], NOAH[40,41], AUTOSTRUCTURE[42,43] and have been recently reviewed[44]. All these programmes require the sequence-specific resonance assignment (§ 1.2.1.1) and lists with cross-peaks from NOESY-type spectra (peak lists) as input. Differently of the others, the CLOUDS[2] programme borrows its approach from X-ray crystallography and does not require any previous resonance assignment. A key-step towards complete automation was the introduction of *Ambiguous Distance Restraints*[45,46] (ADR) into structure calculation strategies based on simulated annealing. With this formalism (see § 1.4.2.2 for a detailed description), the ambiguity of an NOE cross-peak is correctly described in terms of the distances between all pairs of protons that may be involved. This leads to the evaluation of an averaged distance, which is then used to generate an ambiguous restrain term that is a function of all possible assignment options. Rather than rejecting the information contained in ambiguous peaks, this approach allows to include them all into the target function (Equation 1.12). The restraints used are still distances, but the requirement that each distance restraint is assigned to a single pair of protons has been removed. In particular, protein structures can be obtained from ambiguous data alone[46] (§ 3.3.2.2).

1.4.2 Description of the ARIA programme

ARIA[12,32-35] ("Ambiguous Restraints for Iterative Assignment") is a programme for fully automated, iterative assignment of NOESY spectra and structure calculation. Starting with a template file with extended backbone conformation, the programme first creates a list of assignment options for each cross-peak (*peak annotation*, § 1.4.2.1), calibrates cross-peak volumes and derives for each cross-peak a distance restraint by evaluation of a lower and an upper limit (*distance restraint evaluation*, § 1.4.2.2), merges overlapping datasets to remove duplicate information (*merging*, § 1.4.2.3) and finally generates a set of structures *via* CNS[26] (*structure calculation,* § 1.4.2.4). In the following round, these structures are used to discriminate between different assignment options for the same ambiguous cross-peak as well as to identify artefacts and noise peaks on the basis of empirical rules (*noise recognition*, § 1.4.2.5). ARIA then repeats this cycle (*iteration*) of peak assignment, restraint evaluation and CNS structure calculations several times, until satisfying convergence of the structure bundle is achieved (Figure 1.13). During the iterations, an increasing number of less-representative assignment options are rejected for each ambiguous peak[32], resulting in the unambiguous assignment of most ambiguous cross-peaks in the final iteration (§ 1.4.2.1.2). After some (usually 9) iterations, the programme returns a bundle of final, refined structures (Figure 1.14), as well as a list of (mostly unambiguously) assigned NOEs.

The systematic and ordered output of the programme facilitates the manual check of the rejected peaks, which represent an integral part of the result and which may even be submitted with the structures and the accepted restraints.

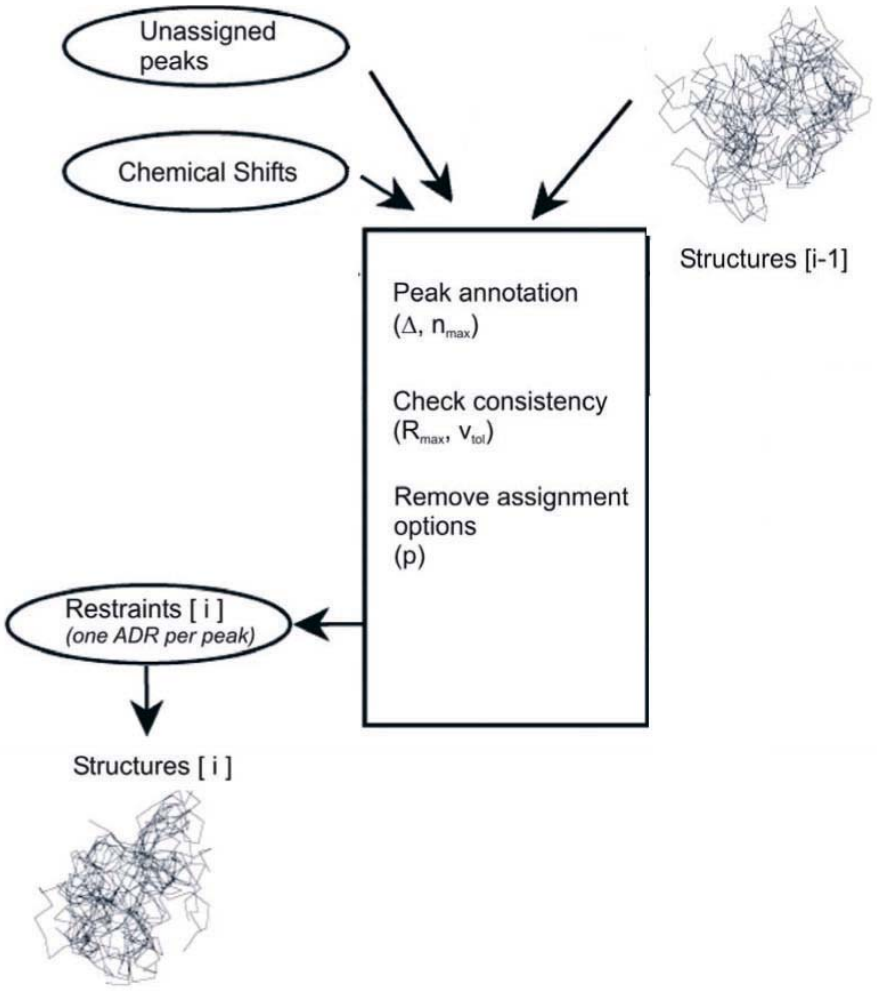

Figure 1.13 Flow-chart of the programme.

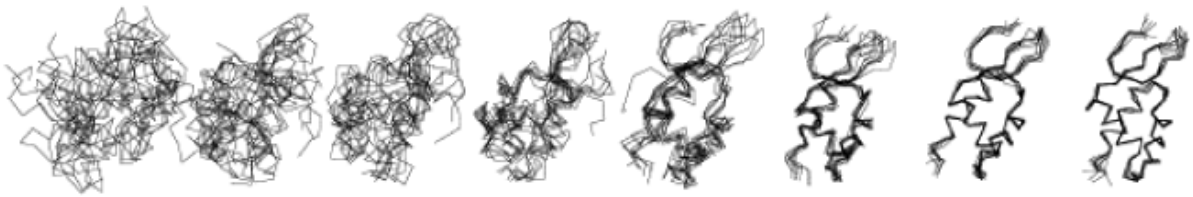

Figure 1.14 Example of convergence of the ensemble of calculated structures during ARIA iterations.

1.4.2.1 Peak annotation

Peak-picking is the procedure generating the peak list of the spectrum. Peak-picking associates every signal intensity identified as cross-peak j to a chemical shift vector

$$C_j = [(c_j^{het1}), c_j^{pro1}, (c_j^{het2}), c_j^{pro2}], \quad (1.19)$$

entry of the peak list C. The values referring to the heteronuclear dimensions of the spectrum are indicated in parentheses since they are only present in 3D and 4D spectra.

Formally, chemical shift values are affected by an intrinsic degree of uncertainty represented by the vector of digital resolutions

$$R=[(r^{het1}), r^{pro1}, (r^{het2}), r^{pro2}]. \quad (1.20)$$

In practice, changing experimental conditions in acquiring different spectra, peak overlap, heating effects etc. further contribute to the actual uncertainty that affects NOE cross-peak co-ordinates. These factors may influence every peak individually. For these reasons, it is more appropriate to use the term *actual chemical shift uncertainty*, represented with the vector

$$U_j=[(u_j^{het1}), u_j^{pro1}, (u_j^{het2}), u_j^{pro2}]. \quad (1.21)$$

It is a function of each individual peak and not a global parameter of the spectrum like the digital resolution. In order to account for the limited precision in chemical shift measurements and for the systematic experimental errors, ARIA, like other automatic methods for the assignment of NOESY spectra, globally applies a vector of chemical shift tolerances

$$\Delta = [(\delta^{het1}), \delta^{pro1}, (\delta^{het2}), \delta^{pro2}]. \quad (1.22)$$

Sufficiently large values for Δ should be chosen to compensate for all sources of inconsistencies between the list of resonance assignments and the peak lists.

During the *peak annotation*, the programme generates a list of assignment options for each peak. This procedure depends intimately on Δ. Every set of spins represented by the two couples of frequencies $A_m = [(a_m^{het1}), a_m^{pro1}]$ and $A_n = [(a_n^{het2}), a_n^{pro2}]$, contained in the list of resonance assignments A, which fulfils the conditions:

$$\begin{cases} c_j^{\alpha} - \delta^{\alpha} \le a_m^{\alpha} \le c_j^{\alpha} + \delta^{\alpha} & \alpha = (het1),\ pro1 \\ c_j^{\beta} - \delta^{\beta} \le a_n^{\beta} \le c_j^{\beta} + \delta^{\beta} & \beta = (het2),\ pro2 \end{cases} \tag{1.23}$$

is accepted as a possible assignment of the peak C_j. NOEs with no assignment options or with a number of assignment options which exceeds the cut-off parameter n_{max} (see next paragraph) are rejected. Depending on the number of assignment options according to Equation 1.23, the accepted NOEs are divided into *unambiguous* (only one assignment possibility) and *ambiguous* (several assignment possibilities).

In case of complete resonance assignment, if

$$\begin{cases} u_j^{\alpha} < \left|\delta^{\alpha}\right| & \alpha = (het1), pro1 \\ u_j^{\beta} < \left|\delta^{\beta}\right| & \beta = (het2), pro2 \end{cases}, \tag{1.24}$$

the chemical shift tolerances Δ account for the actual uncertainty affecting the peak C_j and the correct assignment is taken into account. This peak is referred to as *correctly annotated*. A peak is *incorrectly annotated* anytime one or more conditions in Equation 1.24 are not satisfied.

Prochiral groups which have not been manually stereospecifically assigned are treated according to the floating assignment approach. ARIA does not perform floating assignment by swapping atom positions in the calculated structures, but rather by swapping the chemical shift assignments during the peak annotation itself.

1.4.2.1.1 Δ and n_{max}: two parameters that strongly influence the peak annotation

The chemical shift tolerances Δ and the cut-off n_{max} for the number of assignment options of a peak influence the automated assignment of the NOESY spectrum in various ways. The effects of four different choices for Δ on the annotation of a cross-peak are shown in Figure 1.15. C is a cross-peak of a 2D spectrum with co-ordinates c^1 and c^2. Let $\mathbf{a_a}$, a_b … and a_v, a_w… be examples of resonance assignments close in frequency to c^1 and c^2, respectively. The co-ordinates ($\mathbf{a_a}$, a_x) (bold dashed lines) indicate the only correct assignment for the peak C; any other combination of resonance assignments represents an incorrect assignment option. Black dots designate the accepted assignment possibilities for the peak C. In this example, n_{max} = 20 (the default value) is chosen. In Figure 1.15a, the choice of very narrow Δ values leads to the rejection of the cross-peak due to the lack of assignment possibilities, as no frequency in the list of resonance assignments matches the range $[c^2-\delta^2, c^2+\delta^2]$. In Figure 1.15b, the choice of slightly larger Δ values leads to some assignment possibilities, hence the peak is accepted. However, since the tolerance window is too small to compensate for the actual uncertainty in chemical shift position (i.e., one of the conditions of Equation 1.24 is not satisfied), the correct frequency $\mathbf{a_a}$ lies outside the tolerance window and the assignment possibilities do not contain the correct one. As a result, the peak is accepted but is incorrectly annotated, thus it will be incorrectly assigned at the end of the calculation. In contrast, with much larger Δ values (Figure 1.15c), the correct assignment is taken into account, although together with

many more assignment possibilities. Since the total number of assignment options (20) does not exceed n_{max}, the peak is accepted and correctly annotated. However, due to its high ambiguity, the derived distance restraint will be very loose. A further increase of Δ (Figure 1.15d) leads to the removal of the peak, since the number of assignment options (25) exceeds n_{max}.

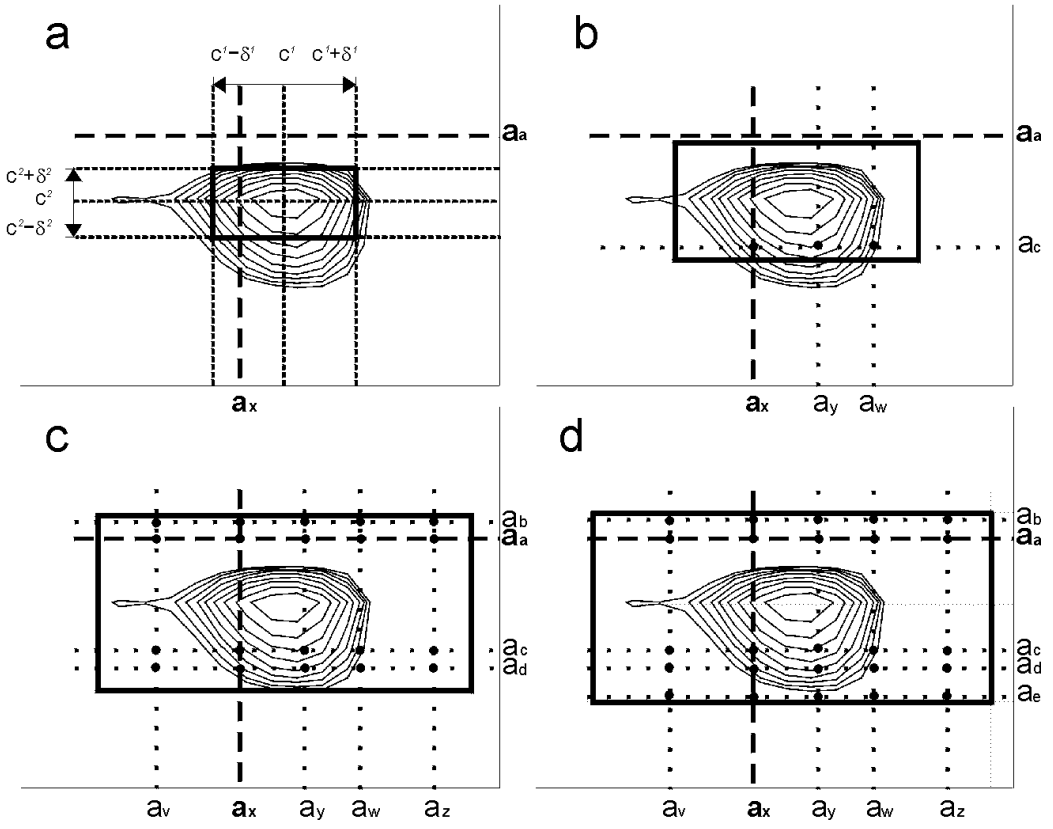

Figure 1.15 Influence of Δ and n_{max} on the annotation of a generic cross-peak. In this example, n_{max} is set to 20. The black dots indicate the assignment options for the peak, obtained by taking all combinations of resonance assignments falling inside different tolerance windows. The co-ordinates ($\mathbf{a_a}$, $\mathbf{a_x}$) (bold dashed lines) represent the only correct assignment for the peak. (a) The peak is rejected because of a lack of assignment possibilities; (b) the peak is accepted as ambiguous restraint but is incorrectly annotated, because the correct frequency $\mathbf{a_a}$ lies outside the tolerance window; (c) the peak is accepted as an ambiguous restraint and is correctly annotated: however, the large number of ambiguities makes it a very loose restraint; (d) the peak is rejected because the number of assignment possibilities exceeds n_{max}.

This example shows that, depending on the values assigned to the parameters Δ and n_{max}, a cross-peak can be accepted or rejected, and if accepted, correctly annotated or incorrectly annotated. Moreover, the example shows that the choice of Δ should not be made independently of that of n_{max}, since both parameters determine together the number of accepted NOEs, the percentage of these that are correctly annotated and the average number of assignment possibilities per peak. Thus, a strategy is required to choose Δ large enough to avoid the exclusion of the correct assignments (Figures 1.15a and 1.15b), without increasing excessively the number assignment options (Figures 1.15c and 1.15d, § 3.3.4).

1.4.2.1.2 The parameter *p*: removal of assignment options during the iterations

Peaks are annotated at the beginning of each iteration. More often than not, several assignment options are possible for the same ambiguous NOE. For each assignment option k, an ensemble-averaged distance $\hat{d}_k$ is calculated from the ensemble of S lowest-energy structures produced in the previous iteration. In the first iteration, an extended-backbone template structure is used for this purpose. Distances $\hat{d}_k$ are calculated by arithmetic average*:*

$$\hat{d}_k = \frac{1}{S}\sum_{S=1}^{S} d_{k,S} \quad . \tag{1.25}$$

Since in the first iterations the ensembles of structures tend to be quite disordered, this average is to prefer to others (e.g., $(r^{-6})^{-1/6}$ -average), which would be excessively weighted on the shortest distance.

The assignment of the spectrum is obtained by reducing, in the course of the iterations, the number of assignment options that are retained for each ambiguous peak. In order to do this,

assignment options are classified with a weight-function w_k:

$$w_k = \hat{d}_k^{-6}. \tag{1.26}$$

Once the different weightings have been normalised by:

$$\sum_{k=1}^{N_\delta} w_k = 1, \tag{1.27}$$

the peak annotation is obtained by ordering all N_δ assignment options by size, and then keeping only the N_p representative ones such that:

$$\sum_{k=1}^{N_p} w_k > p, \tag{1.28}$$

where p is the *assignment cut-off*, a real value usually between 0.8 and 1 (Table 1.1).

Iteration	0	1	2	3	4	5	6	7	8
p	1.00	1.00	0.9999	0.999	0.99	0.98	0.96	0.90	0.80

Table 1.1 Default values of the parameter *p* for all 9 ARIA iterations.

Assignment options corresponding to large distances in the ensemble of structures calculated in the previous iterations are in this way considered as much less representative and thus are more likely to be discarded.

1.4.2.2 Distance restraint evaluation

The formalism of *Ambiguous Distance Restraints*[45,46] (ADR) provides a way to derive distance restraints from ambiguous cross-peaks. In such a formalism, the volume of an ambiguous NOE is interpreted as the sum of several volumes proportional to the sixth power of the inter-atomic distances of each assignment option, i.e., the interaction of two protons far away from each other contributes much less to the total volume of the ambiguous cross-peak than the interaction between two close protons. A characteristic distance $\hat{d}$ [b] for the cross peak can be calculated as summed distance over the different assignment options. The effective distance $\hat{d}$ thus contains all contributions from all assignment options:

$$\hat{d} = \left(\sum_{k=1}^{N_p} r_k^{-6} \right)^{-1/6} \tag{1.29}$$

and is strongly weighted on the shortest of the distances r_k. Here, $r_k = \hat{d}_k$, as calculated in Equation 1.25. If the peak is unambiguous, Equation 1.29 simply reduces to Equation 1.25 and $\hat{d} = \hat{d}_k$. The presence of some incorrect options among the assignment possibilities for a cross peak does not lead to inconsistencies as long as the correct assignment is present, since the r^{-6}-weighted average distance $\hat{d}$ in Equation 1.29 is always shorter than and strongly weighted to the shortest of the distances $\hat{d}_k$. In contrast, whenever the correct assignment is not included among the assignment possibilities, the derived distance restraint is likely to be inconsistent with the others and thus potentially able to induce distortions in the structures.

[b] In the following text, the symbol ˆ will be used to indicate averaged distances.

These effective distances $\hat{d}$ are used to determine the calibration factor C for the spectrum, to directly convert peak volumes V into distances. A unique calibration factor is calculated for all NOEs, by:

$$C = \sum_{NOEs} \frac{\hat{d}^{-6}}{V} \tag{1.30}$$

Subsequently, experimental volumes V are transformed into observed distances d^{obs} directly *via* the calibration factor C. This calibration does not return upper bounds, but rather an estimation of the target distance itself:

$$d^{obs} = (CV)^{-1/6} \ . \tag{1.31}$$

Upper (U) and lower (L) limits for distance restraints are obtained *via* equations:

$$\begin{aligned} L &= d^{obs} - \Delta^{-} \\ U &= d^{obs} + \Delta^{+} \end{aligned} \tag{1.32}$$

where

$$\Delta^{-} = \Delta^{+} = \varepsilon_0 + \varepsilon_1 \, d^{obs} + \varepsilon_2 (d^{obs})^2 + \varepsilon_3 (d^{obs})^3, \tag{1.33}$$

which gives an empirical estimation of the error *via* a simple polynomial in d^{obs}. By default, $\varepsilon_0 = \varepsilon_1 = \varepsilon_3 = 0$ and $\varepsilon_2 = 0.125$, which means that the estimated errors increase with the square of the target distance d^{obs}. L and U are therefore functions of the experimental volume, as it

can be seen by substituting Equation 1.31 into Equation 1.32.

After L and U have been determined, NOE cross-peaks can be transformed into *ADRs* by imposing that the effective distance $\hat{d}$ in Equation 1.29 is contained between L and U.

$$L \leq \hat{d} \leq U \tag{1.34}$$

where now distances r_k in Equation 1.29 are the unknown to be calculated.

The formalism of ambiguous distance restraints can be easily extended to the treatment of other ambiguous data, which may include equivalent protons (Figure 1.16A), hydrogen bonds (Figure 1.16B) and disulfide bridges (Figure 1.16C).

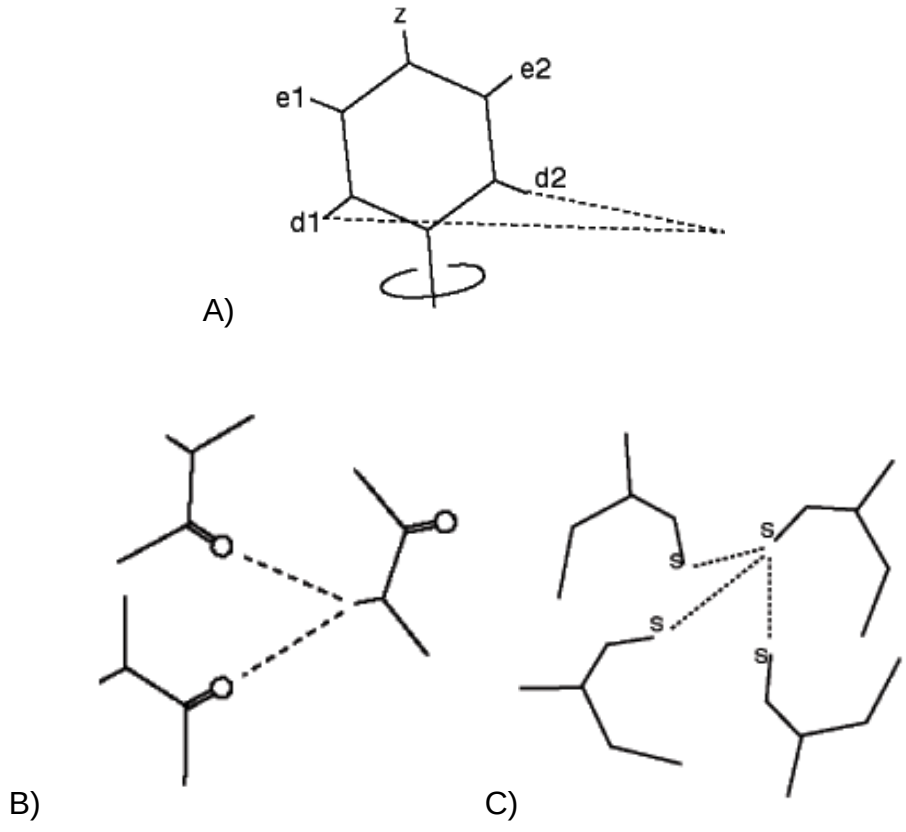

Figure 1.16 Further applications of the formalism of ambiguous distance restraints.

1.4.2.3 Merging of the peak lists

ARIA can read several input NOESY peak lists. It often happens that the same restraint occurs twice or more in different peak lists. Additionally, further duplications arise when peak-picking is done on both sides of the diagonal of a 2D spectrum. *Merging* indicates the process of creating a unique peak list (the *merged list*) from all supplied peak lists by removing duplicate peaks, arising from the same NOE interactions being present in different spectra. Whenever ARIA finds two restraints with exactly the same assignment options, only the restraint with the tightest error-bounds is kept. E.g., if two NOESY peak lists measured at different mixing times are used, distance restraints from the spectrum with lower mixing times (tighter error-bounds) are used in place of those from the spectrum with higher mixing times, if both can be satisfied.

1.4.2.4 A description of the SA protocol used in ARIA

After their evaluation, distance restraints are subjected to simulated annealing with the CNS[26] programme to generate a set of structures. In ARIA, the user can choose between Cartesian[11] and Torsion Angle Molecular Dynamics (TAMD)[47]. Both of these simulated annealing-based structure calculation strategies have been optimised for ambiguous NOE-derived restraints. TAMD was employed for all structure calculations in this work, as this generally produces an increased convergence radius and leads to better local geometries. The standard protocol for TAMD in ARIA used in the present work consists of an initial high-temperature conformational search at 2,000 K, followed by two cooling stages in which the temperature is decreased linearly from 2,000 K to 1,000 K and then again from 1,000 K to 50 K. During these three stages, the weightings w_i in Equation 1.12, defining the relative weights of the different energy terms of the target function, are varied (see Figure 1.17).

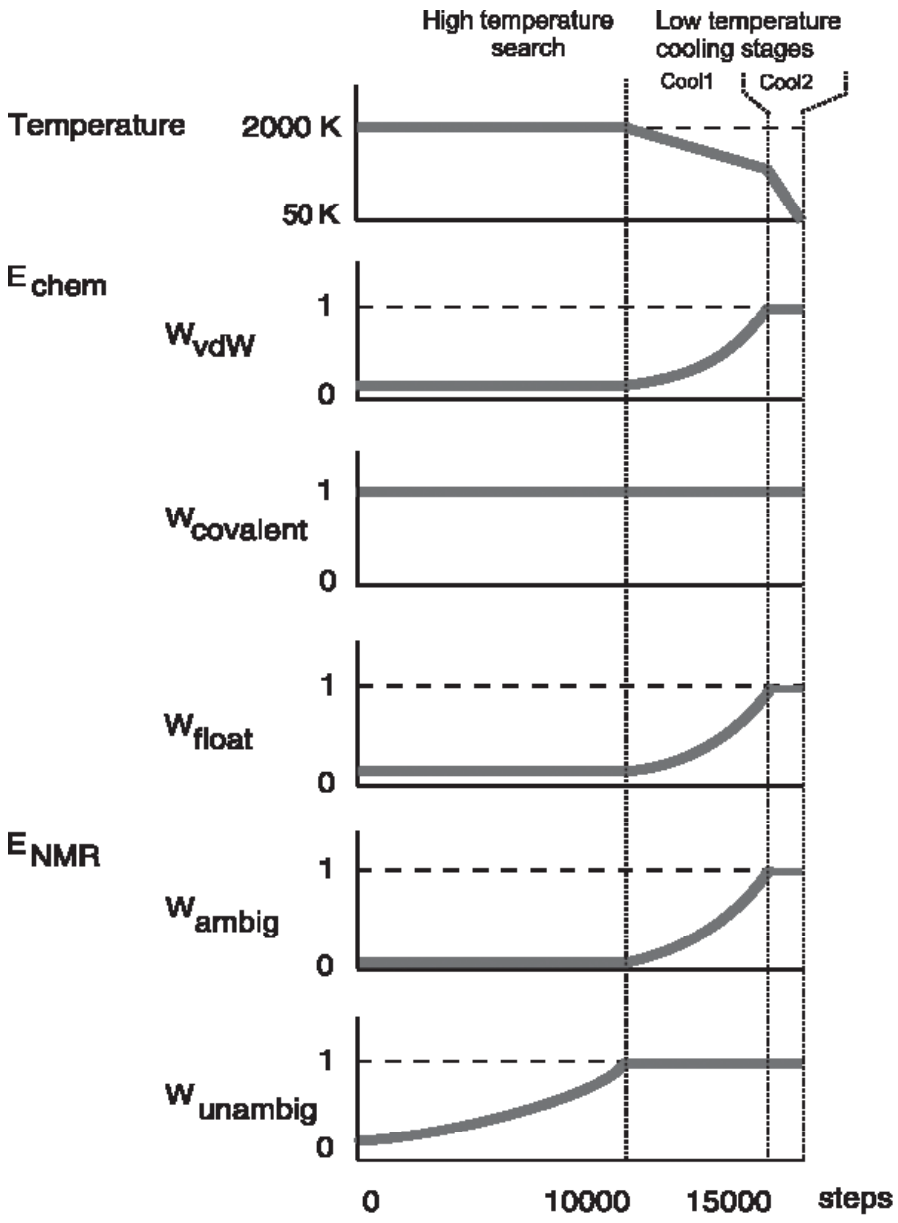

Figure 1.17 Temperature and energy constants (w_i) in a typical torsion-angle dynamics SA protocol. $w_{covalent}$ and w_{vdW} are the energy weightings for bonded and nonbonded interactions, respectively. w_{float} is the weighting of the angular energy term for chemical groups which undergo floating chirality assignment. The length of the high-temperature search and of the two low-temperature cooling stages is commonly expressed as number of MD steps. The two parameters Cool1 and Cool2 indicate, respectively, the number of MD steps for the first and the second cooling stage and determine the slope of the temperature function during simulated annealing.

The length of each of these three stages can be specified by the user prior to calculation and is expressed as number of Molecular Dynamics (MD) steps. MD steps during the two cooling

stages are commonly referred to as *cooling steps* and are indicated with Cool1 and Cool2. These two parameters are particularly relevant, since they determine the slope of the temperature function during the cooling.

1.4.2.5 Recognition of artefacts and noise-peaks

NOESY peaks lists often contain artefacts and noise-peaks. Artefacts may arise from spectral processing, impurities in the sample or incomplete suppression of the strong water signal. Additionally, there are uncertainties in peak-position due to imprecision of automated peak-picking algorithms, especially in crowded regions of the spectrum where peaks are highly overlapped, which often leads to an incorrect annotation of real data.

In general, distance restraints derived from noise-peaks or incorrectly annotated peaks will not be consistent with the three-dimensional structure: this can be used as a criterion to identify and to reject them. ARIA first annotates cross-peaks and evaluates ADRs: restraints which are systematically violated in the converged structures of the previous iteration are rejected. A violation is considered "systematic" if it exceeds a threshold v_{tol} in N_v converged structures. The threshold is decreased sensibly during the iterations. The ensemble of converged structures corresponds usually to the first third of lowest-energy structures. N_v is usually set to 50% of the converged structures. If the number of structures in which the restraint is violated (R_{vio}) exceeds the threshold R_{tol} (usually set to 0.5), than the peak is considered inconsistent. R_{vio} is calculated by:

$$R_{vio} = \frac{1}{S_{conv}} \sum_{s}^{S_{conv}} \Theta(\hat{d} - U - v_{tol}) \tag{1.35}$$

$\Theta(x)$ is the Heaviside step function, which is 1 for x>0 and 0 for $x \leq 0$; $\hat{d}$ is the effective distance for the peak (Equation 1.29) and U is the upper limit (Equation 1.31).

The upper error bounds U play a crucial role for discriminating between noise-peaks and real data. The tighter they are set, the larger the number of peaks which violate the structures. Too tight values will lead to the rejection of real data affected by problems in exact volume determination (due to imprecision in volume integration in crowded regions of the spectra, or because of spin-diffusion or dynamics effects, which notably lead to a significant deviation of the r^{-6}-dependence of the peak volume). Conversely, when the error bounds are set to generous, some artefacts will be included in the calculations, and, additionally, information content is reduced; as a result, the precision of the structures decreases.

1.5 Overview of the thesis

In these studies, new computational methods for protein structure calculation *via* automated assignment of NMR spectra were developed. In particular, the possibilities and pitfalls of using ADRs (§ 1.4) for structure calculation were extensively investigated. A detailed description of the protein datasets and of the parameters used for the calculations can be found in **Chapter 2**.

The performance of software packages for automated NOESY assignment is intimately dependent on the setting of a number of parameters for optimal performance. In **Chapter 3**, the influence of different values for Δ and n_{max} (§ 1.4.2.1.1) on the performance of programmes for automated assignment of NOESY spectra was systematically investigated (using the ARIA protocol as an example) and a strategy is presented to choose optimal values for the two parameters. In **Chapter 4**, the influence on the calculations of other two important parameters, the numbers of SA cooling steps Cool1 and Cool2 (§ 1.4.2.4) was investigated.

This work shows that large tolerance windows, and the correspondingly high levels of ambiguity, do not cause problems when appropriately high numbers of cooling steps are used in the simulated annealing protocol. In this way, high quality structures can be obtained even for proteins whose NMR spectra show great degeneracy, and where there are serious inconsistencies in peak alignment between different samples.

In **Chapter 5**, a first successful example of automated protein structure determination from solid-state NMR data is presented. The SOLARIA programme, a MAS NMR-dedicated version of ARIA, was developed to automatically assign MAS NMR peak lists and, subsequently, was then tested on completely unassigned PDSD peak lists of the α-spectrin SH3 domain. Such an automated approach allows including in the calculation also peaks in highly overlapped regions of the spectra.