Peptide Structure from NMR

Michael P. Williamson and Jonathan P. Waltho
Krebs Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield S10 2UH

1 Introduction

In recent years, NMR studies of peptides have been overshadowed by the rapid progress of NMR studies of proteins. The techniques developed for proteins, both those used for assignment and those used for structure calculation, are now being applied to peptides. It is therefore appropriate to begin this review with a brief consideration of protein NMR, and of the differences between proteins and peptides.

It is relevant to ask why protein NMR has been so successful. To a large extent, this success has come from technical advances, primarily from the development of two-dimensional NMR. However, there are special features of proteins (as compared, for example, with nucleic acids, polysaccharides, or linear peptides – as we shall discuss below) that make them especially suitable for structure determination by NMR:

(a) Because of the densely packed nature of the interior of proteins, the relative positions of atoms in the centre of proteins are largely fixed because of steric constraints.

(b) The globular nature of the structure means that there are many short internuclear distances that can be observed using nuclear Overhauser effects (NOEs) inside proteins (see below).

(c) Contacts between residues far apart in the primary sequence act as particularly powerful constraints (Figure 1).

These reasons have made it possible to develop a common methodology for protein structure determination, shown in Figure 2. Typically this methodology would involve measurement of NMR constraints (primarily the NOE, but also *J*-coupling and NH exchange); application of these constraints plus non-NMR constraints (bond lengths and angles, and other energy terms) to drive towards a structure; and repetition of the calculations with different starting structures, in order to give a spread of calculated structures. These structures represent equally accurate approaches to the time-averaged solution structure. There is currently some debate about how much the spread of calculated structures represents the spread of structures actually accessible in solution, but the weight of opinion is that they do, provided that enough constraints have been used (typically 10—12 per residue).[1]

Mike Williamson obtained his Ph.D. from the University of Cambridge in 1981 in the laboratory of Dr. Dudley Williams, and continued his research with a Junior Research Fellowship at Churchill College, Cambridge, and an SERC/NATO Overseas Postdoctoral Fellowship in the laboratory of Prof. Kurt Wüthrich at the ETH, Zürich. He was appointed as Team Leader in Bio-NMR at Roche Products in 1984 and as Senior Lecturer in the Krebs Institute of the University of Sheffield in 1990. His research interests are in the application of NMR, and particularly the NOE, to the structures of peptides and proteins.

Jon Waltho obtained his B.Sc. from the University of Durham in 1983 and his Ph.D. from the University of Cambridge in 1987 in the laboratory of Dr. Dudley Williams. After working as a Postdoctoral Fellow jointly with SmithKline and French Research plc, he moved to the Scripps Research Institute, in the laboratory of Dr. Peter Wright. He was appointed as a Lecturer in the Krebs Institute of the University of Sheffield in 1990. His research interests are in using NMR structural parameters to understand the thermodynamics of molecular interactions involving proteins.

Obtaining the structures of peptides is generally not so easy. There are far fewer packing restraints, there is a lower density of NOEs (6 NOE constraints per residue would be high), there are often no contacts between residues far apart in the sequence, and there is a much greater ratio of surface to interior. Even more important, it is rare that peptides have a single rigid conformation. An important consequence is that the time-averaged conformation may poorly reflect the conformational ensemble populated by the peptide (Figure 3), and therefore any structure produced by application of the standard protein structure calculation techniques may be limited in its usefulness;[2] this point is discussed below.

If there *is* only one structure (or at least one family of closely related structures), the standard protein approach is relevant for the appropriate regions of the peptide. But this is rare, and certainly rarer than one might expect judging by the large number of peptide 'structures' appearing in the literature. Therefore we argue below that in general a different approach is usually needed. In the remainder of this review we survey briefly what is known of the structures of peptides in solution, and then go on to look in more detail at the methods used to generate protein structures. We then consider how these methods are applied to solving the structures of rigid peptides (particularly where the methods differ from those used for proteins), and finally survey approaches to multiple conformations. The review is not intended to be comprehensive, and the references cited are largely either recent primary publications or reviews.

2 Peptides in Solution

In this section we review what is known of peptide structure, from experimental and theoretical evidence. Many of the techniques described here were developed for proteins, but it is reasonable to assume that they also describe peptides adequately. Finally, we discuss the nature of the poorly characterized 'random coil' structure.

2.1 Experimental Evidence

The techniques that have proved most informative about the structure of peptides in solution are fluorescence, circular dichroism (CD), Fourier transform infra-red (FTIR) and Raman spectroscopies, and NMR.[3] The range of information from these techniques is to a large extent complementary. The first four techniques are all capable of time-resolving most conformational exchange phenomena occurring within peptides, but each has its own drawbacks, for example, fluorescence requires specific 'reporter' groups (fluorophores) in the region of study, CD is dependent upon a cooperative effect of the amide transitions and hence is positional-averaged, and FTIR and Raman, although capable of both positional and time resolution, suffer from insufficient frequency resolution to be used reliably in a site-specific-manner without recourse to isotopic labelling. NMR, in contrast, offers considerably more attainable frequency resolution for site-specific measurements but has the disadvantage that observations are made over a timescale of milliseconds to seconds. Conformational exchange on timescales faster than this are therefore time-averaged. This includes virtually all peptide conformational exchange processes, except for proline *cis–trans* isomerization, which occurs with rates of the order of 0.01 to 0.1 s^{-1}. Peptides containing prolines therefore usually give multiple sets of signals. In linear peptides,

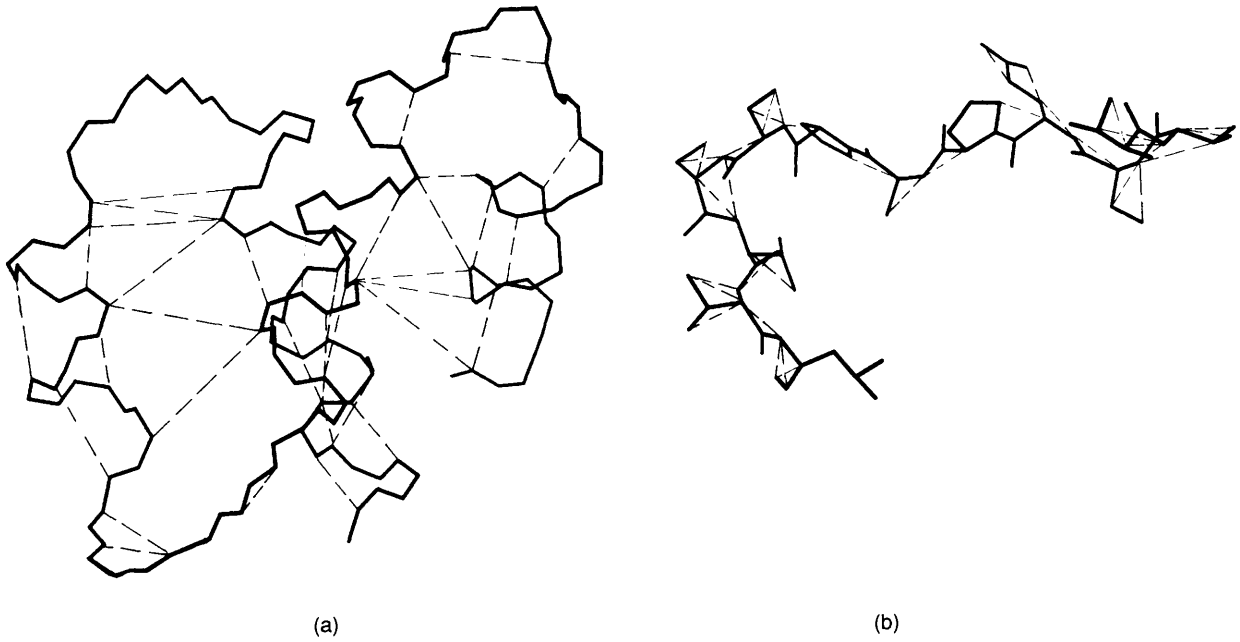

Figure 1 For a globular protein (a), the overall structure is well determined by qualitative constraints between residues far apart in the sequence. For a peptide (b), structure determination relies on accurate constraints between neighbouring residues.

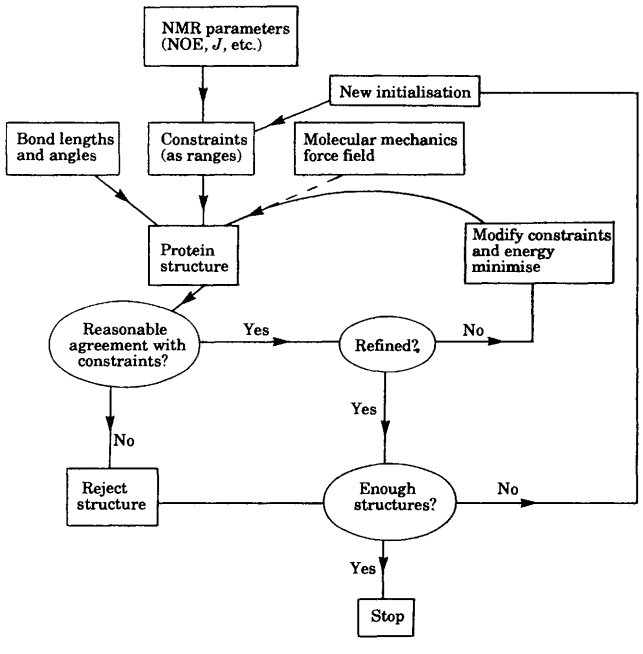

Figure 2 Typical procedure for protein structure calculation.

approximately 20% of prolines are *cis*, although the proportion depends on sequence, with Tyr-Pro and N-terminal X-Pro having particularly high *cis*-Pro content.

We shall return to the question of conformational averaging repeatedly. A major consequence is that there is no direct way from NMR to determine the relative populations of the different conformers present. However, using a combination of experimental approaches, it is possible to obtain information on conformational populations using time-resolving techniques (fluorescence, CD, FTIR/Raman) and site-resolving techniques (fluorescence, FTIR/Raman, NMR). We shall concentrate on the information content and role of NMR spectroscopy to the study of peptide conformations.

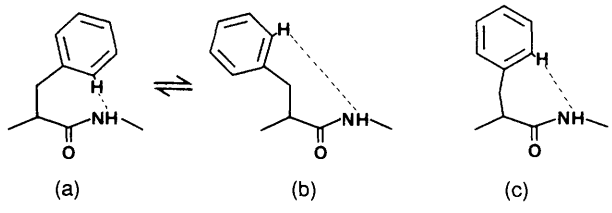

Figure 3 In a hypothetical protein, the phenylalanine sidechain is averaging between conformations (a) and (b). The resultant averaged NOE implies structure (c), which has no real existence.

Early work on linear peptides in aqueous solution, particularly using CD, suggested that they are largely unstructured (*i.e.* they consist of a large number of rapidly interconverting structures, all of low probability). By contrast, conformationally restricted peptides (*e.g.* cyclic peptides) often display some form of conformational preferences.[4] These results agreed with theoretical predictions (see below). For many years, the lone exception was a linear peptide corresponding to the N-terminal 20 residues of ribonuclease (the ribonuclease S-peptide), which under some conditions appeared to be at least partially helical.[3] More recent results, particularly using NMR, have shown significant conformational preferences for a range of peptides (as we shall discuss below).

2.2 Computational Evidence

Peptide conformation is normally described in terms of the backbone dihedral angles ϕ and ψ, and sidechain torsion angles χ_1, χ_2 *etc.* (Figure 4). The well-known Ramachandran plot, describing preferred regions for the backbone angles ϕ and ψ, used simple hard-sphere repulsions as a first-order approximation to the molecular structure (Figure 5). A range of more sophisticated techniques has also been applied, from molecular orbital calculations to the more empirical molecular mechanics force fields, but the general features of all the calculations are similar: there is a broad minimum in the β-sheet region, and a somewhat smaller (and less deep) minimum in the α-helix region, together with a much less favoured region in the right-handed α-helix region. The energy barriers between different conformations are low enough to permit rapid conformational exchange. For example, the energy barrier between α-helix and β-sheet areas is estimated to be approximately 8 kJ mol^{-1}, corresponding to virtually unrestricted interconversion.

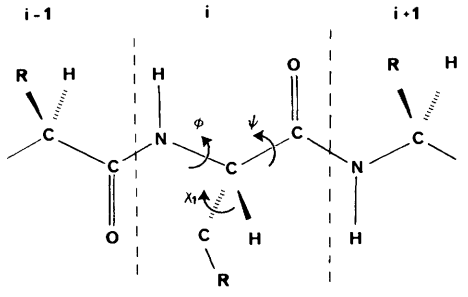

Figure 4 The angles ϕ, ψ, and χ_1.

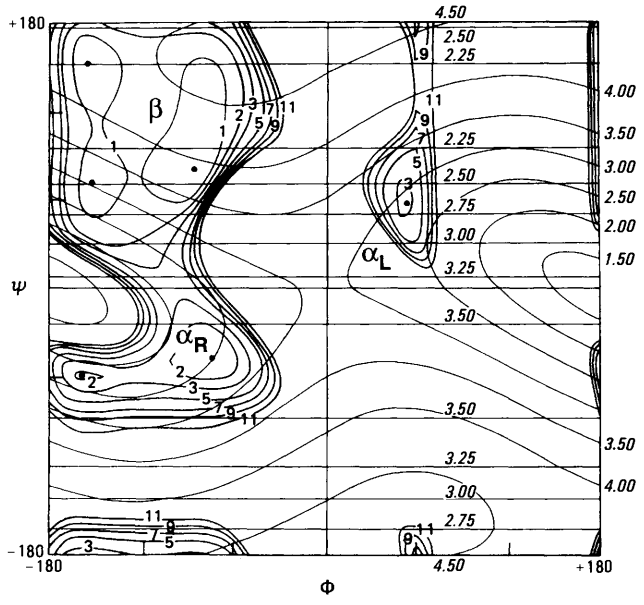

Figure 5 Conformational energy map for the alanine dipeptide using Ramachandran's parameters. Energy contours are drawn at intervals of 1 kcal mol^{-1}. The potential energy minima for β, α_R, and α_L are labelled. The dependence of the sequential distance $d_{NN}(i,i+1)$ on the dihedral angles ϕ and ψ is shown as a set of contours labelled according to interproton distance at the right of the figure. The $d_{\alpha N}(i,i+1)$ distance depends only on ψ and is shown as a set of contours parallel to the ϕ axis. These distances are defined in Figure 7.
(Reproduced with permission from reference 3.)

There is no easy way to check the accuracy of these calculations. If the backbone dihedral angles found in proteins (Figure 6) are compared to those calculated, the agreement is fair; but there is no guarantee that the angles found in proteins are necessarily constrained purely by local energetic constraints. For example, estimates of the conformational strain within proteins suggest an *average* of approximately 8 kJ mol^{-1} per residue.[5]

Overwhelmingly the most popular technique currently used for the calculation of protein energetics and dynamics is molecular mechanics/molecular dynamics. Its great popularity should not blind us to the fact that it is a very approximate and empirical method. It would be a foolhardy person who could maintain that a classic Newtonian system (and moreover one *in vacuo* in most calculations) is a close approximation to the behaviour of a highly complex quantized system in a cooperative polarizable solvent system; nevertheless, molecular mechanics is the best method routinely available.

2.3 The Random Coil

The 'random coil' is, by common usage, the conformation adopted by 'unstructured' peptides. What does this mean? If it means anything, it means that the peptide has a range of

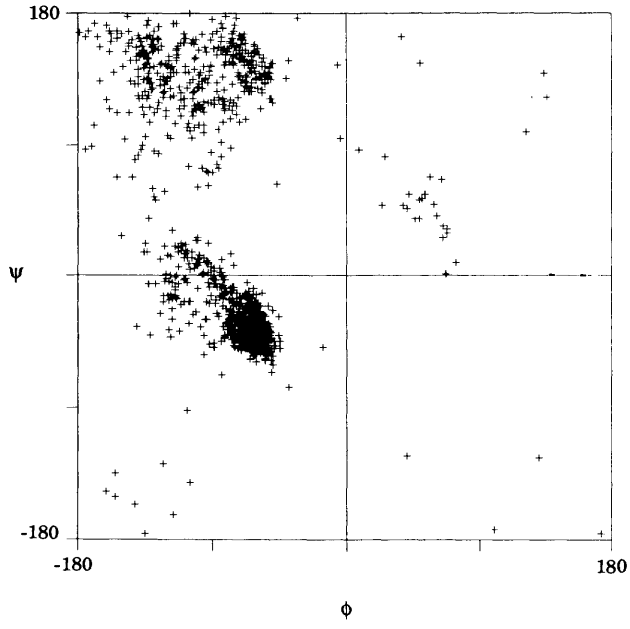

Figure 6 Plot of ϕ and ψ for approximately 1500 non-glycine residues from thirteen of the highest resolution crystal structures of different proteins in the Brookhaven data bank.

conformational states available to it that are of similar free energy, and little or no barrier (compared to kT) to interconversion between them. If Ramachandran-type calculations are correct, this would mean that any individual amino acid residue in a random coil peptide would spend up to 95% of its time in the broad β-sheet region, and almost all of the rest in the α-helix region.[6] The available NMR evidence is in agreement with this description, although this does not constitute a proof of its validity: for example, the three-bond coupling constant between NH and C$^{\alpha}$H ($^3J_{HN\alpha}$) in most linear peptides is around 7 Hz, much closer to the typical β-sheet value (around 8.5 Hz) than the α-helix value (around 4.5 Hz). Likewise, the NOEs seen in random coil peptides are typified by large sequential $d_{\alpha N}$ connectivities, as found in β-sheet, and small or non-existent d_{NN} connectivities, as found in α-helix (Figure 5; these distances are defined in Figure 7). Techniques that can measure over longer distance ranges, such as fluorescence quenching, indicate that peptides are largely extended in solution, again consistent with their existing largely in the β-sheet region. The use of the word 'coil' for this structure is therefore misleading.

3 Methods used for the Calculation of Protein Structure

3.1 NMR Parameters

As mentioned in the Introduction, the methods used for deriving peptide structures are based closely on those used for deriving protein structures. We therefore review these techniques briefly, with a special emphasis on the effects of motional averaging.

3.1.1 The NOE

By far the most useful parameter is the nuclear Overhauser effect (NOE). The NOE is a phenomenon arising from cross-relaxation between two nuclei, and is manifested in one-dimensional (1D) spectra as a change in intensity of one signal on saturation of another, arising from a nucleus close in space. In the two-dimensional experiment, known as NOESY, the NOE is seen as a cross peak, and builds up during the mixing time at an initial rate proportional to r^{-6}; however, the buildup rapidly becomes non-linear, particularly when there are other protons close by. To remain in the linear region, it is necessary to use very short

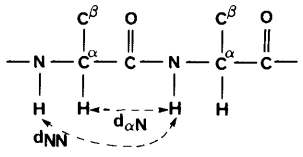

Figure 7 Definition of the sequential distances d_{NN} and $d_{\alpha N}$.

mixing times, which creates major problems of artifacts and poor signal-to-noise. The alternative is to recognize that the NOE buildup is affected by the presence of other protons, and calculate the expected NOEs based on the calculated protein structure. These so-called back-calculated NOE intensities can then be compared to the experimental values, and the difference used in an iterative structure refinement procedure (Figure 2).[7]

A second serious problem with NOEs, much more so with peptides than with proteins, is motion. If intramolecular rearrangements are occurring at a rate fast compared to the overall tumbling rate, the NOE is averaged as $<r^{-3}>$. This leads to a strong biasing of the apparent NOE, such that minor conformations can give rise to very large NOEs if they contain short internuclear distances. We have attempted to illustrate this phenomenon in Figure 3, by drawing the 'average' conformation as having a shorter distance than the simple arithmetic mean of the extremes. This is the probable explanation for the appearance of d_{NN} NOEs in peptides: a 10% population of conformers with d_{NN} of 2.6 Å (typical of tight β-bends) in a population of otherwise extended conformers ($d_{NN} = 4.3$ Å) gives an NOE 5% of the $d_{\alpha N}$ NOE, corresponding to an apparent distance of 3.8 Å. Slower intramolecular motion averages the NOE as $<r^{-6}>$, which gives an even more biased picture. The NOE is therefore very powerful, but tends to overemphasize structural tendencies. This is no doubt one reason why NMR has been so successful (compared with other techniques) at 'detecting' the presence of folded peptide conformations in solution.

A further consequence of peptide motion that causes difficulties in NOE interpretation is the dependence of the buildup rate on the correlation time of the inter-proton vector that gives rise to the NOE. Thus, for accurate quantitation of NOEs, it is necessary to know not only the inter-proton distances in each significantly populated conformation, and the population, but also the motional properties of each conformation. Correlation times in peptides are often in the range that gives rise to small NOEs (Figure 8). The rotating frame NOE (ROE) experiment is designed to overcome this problem, as the ROE is always negative (or always positive in 1D experiments). However, this solution is not free of difficulties, as it is more prone than the NOE to giving rise to misleading peak intensity (positive or negative) arising from spin–spin coupling pathways.[8]

3.1.2 Spin–Spin Coupling

Coupling constants are dependent on the angle between the protons. The angular dependence for $^3J_{HN\alpha}$ has been calibrated on several proteins; the current best estimate is[9]

$$^3J_{HN\alpha} = 6.7\cos^2(\phi - 60) - 1.3\cos(\phi - 60) + 1.5$$

and is illustrated in Figure 9. The coefficients for $^3J_{\alpha\beta}$ are reasonably well established, although they vary with the electronegativity of the β-carbon substituents; a generally applicable set of values is[10]

$$^3J_{\alpha\beta} = 9.4\cos^2\theta - 1.4\cos\theta + 1.6$$

Angular dependencies for heteronuclear couplings are less well characterized, but with the advent of isotopic protein labelling, this situation is certain to change rapidly.

For both the NH–α and α–β couplings, the coupling constants

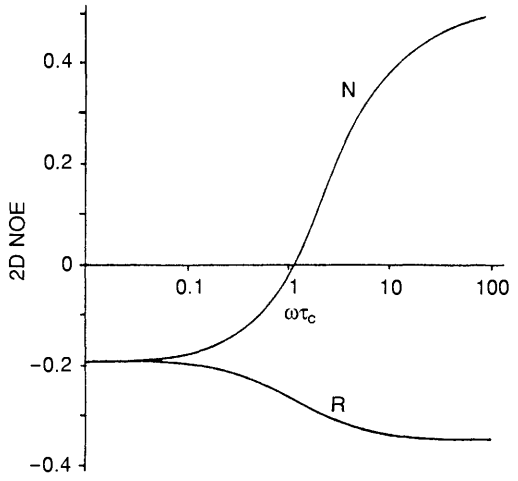

Figure 8 Dependence of maximum 2D NOE intensity on $\omega\tau_c$ for longitudinal NOE (N) and transverse (rotating frame) NOE (R). ω is the spectrometer frequency (in rad s^{-1}) and τ_c is the rotational correlation time, the time constant for rotation of the interproton vector. When $\omega\tau_c$ is close to unity (as is typical for small to medium sized peptides) the longitudinal (normal) NOE is close to zero. (Reproduced with permission from reference 2.)

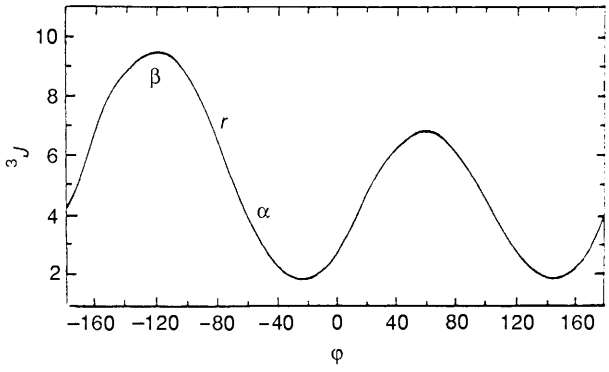

Figure 9 Dependence of $^3J_{HN\alpha}$ on ϕ. Also indicated, by α, β, and r respectively, are the values typically found in α-helix and β-sheet, and the averaged value commonly seen in 'random coil' peptides in solution.

corresponding to energy minima usually happen to be at the extreme ends of the range. Thus, any conformational averaging inevitably brings the splittings towards an averaged value, which cannot be used for deriving conformational information (Figure 9). The contrast between this behaviour and that for the NOE is striking.

3.1.3 Other Parameters

The only other information routinely used in the determination of protein structures by NMR concerns the location of hydrogen bonds. The NH end of the hydrogen bond is indicated by slow NH exchange rates, but there is no direct evidence to indicate the CO end; this end is deduced by presupposing the nature of the local secondary structure. This can be a more or less circular argument. For example, an α-helix may be indicated by a string of slowly exchanging amide protons; the helix is then entered as a constraint by including hydrogen bonding distances in the distance constraint list; these constraints produce a well-formed α-helix, which is then used to justify the inclusion of the exchange rates as hydrogen bonding constraints. The circularity is avoided by justifying the secondary structure purely by consideration of the NOE and J constraints (which is not always straightforward, especially in distinguishing between different types of helical structure), and is the reason why some authors

are reluctant to use hydrogen bonds as a structural constraint. Similar arguments could also be made for peptides, with the possible exception of small cyclic peptides, where in some cases only one hydrogen bonding partner is geometrically feasible.

In peptides, NH exchange rates are seldom measured; the preference is to use the temperature dependence of the NH shift for the same purpose. There are two reasons for this, firstly that many peptide NMR studies are carried out in dimethylsulfoxide (DMSO), where no exchange occurs, and secondly that for many peptides in water or methanol, exchange is so rapid that it is only measurable by saturation transfer experiments, which are technically harder than simple observation of exchange. There is little firm evidence as to whether either NH exchange or amide temperature dependence gives direct information on NH hydrogen bonding. Acid-catalysed NH exchange in model systems appears to occur by protonation of the amide oxygen rather than the nitrogen, and thus may depend on solvent exposure or hydrogen bonding of the oxygen as well as the nitrogen. However, in proteins, amide exchange rates do correlate well with crystallographically determined NH hydrogen bonding patterns,[11] while in peptides, amide temperature dependence has been stated to be applicable to the determination of turn populations in a diagnostic manner.[12]

In peptides, the solvent dependence of the NH chemical shift has also been used as an indicator of amide hydrogen bonding, with the more solvent-shielded amides showing a lower change on change of solvent.[2] Solvent dependence studies are useful in a wider sense, as comparison of spectral parameters (*e.g.* J and NOE) in different solvent compositions indicates how sensitive the peptide conformation is to a change in solvent, and therefore how 'unique' the peptide conformation is. If the peptide is present as a conformational ensemble, then one would anticipate that changing the solvent would affect the population distribution within the ensemble, and therefore alter the spectral parameters; conversely, if one conformation dominates, then changing the solvent would not greatly affect the spectral parameters.[2]

One drawback to the structural interpretation of amide hydrogen bonding is hydrogen bonding from sidechains. Aspartate and glutamate sidechains hydrogen bond to their own amides and to adjacent amides, and lower their temperature coefficients. Great care must be taken in analysing peptides in such circumstances.[3]

Chemical shift deviations from random coil values have been used to show that peptides are not random coil, but unfortunately they give very little other information, except in rare cases an indication as to the orientation of aromatic rings. Relaxation parameters, such as ^{13}C T_1, T_2, and NOE, give very valuable information on molecular motions,[13] and will undoubtedly become more popular as the techniques become more widely available.

Peptide sidechains are seldom constrained sufficiently to be limited to a single torsion angle. Hence, the rapidly interconverting conformations are often modelled as an equilibrium between the three possible staggered rotamers. The relative proportions of the rotamers can often be analysed by consideration of the coupling constants.[10]

3.2 Calculational Methods

The two main methods developed for calculation of protein structures are distance geometry and molecular dynamics (Figure 10). Distance geometry (DG) comes in two different flavours. The first, represented by the programs DISGEO, DSPACE, and DG-II, is a 'true' DG program, in that it works initially purely in distance space, calculating a consistent matrix of distance bounds between all atoms, starting from the limited set of distances (NOEs and covalent distances) input to it. It then embeds a subset of these distances into Cartesian space and refines the resultant structure in various ways, to improve the fit of the structure to the distance matrix. It is in principle the least biased way of generating a representative set of structures

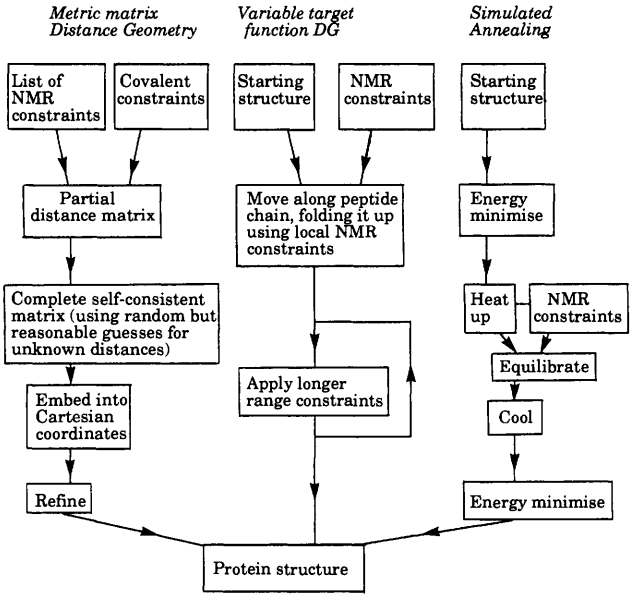

Figure 10 Diagrammatic representation of the two types of distance geometry calculation and simulated annealing. Frequently, distance geometry will be used to provide starting structures for refinement using simulated annealing.

consistent with the initial distance bounds, since it does not start from any initial structure. The second type of DG, represented by the programs DISMAN, DIANA, and DGEOM, works in angle space. Starting from an initial structure, *e.g.* an extended chain, it folds the peptide so that it satisfies the distance constraints. It uses a variable target function, *i.e.* it starts by applying intra-residue constraints one residue at a time along the chain, then adds $(i, i+1)$ constraints and operates on dipeptides, working on each peptide pair along the sequence, then uses $(i, i+2)$ constraints, and so on. This procedure leads to increased computational efficiency. Both kinds of DG (particularly the second) have the problem that they can get into local minima in which the peptide chain is tangled up incorrectly. It is therefore necessary to calculate a large number of structures, many of which must be subsequently rejected as being poorly folded, on the basis of a poor fit to the NOE constraints used as input.

The second method for structure calculation is restrained molecular dynamics (RMD), in which the standard forces are supplemented by others representing NOE constraints, angle restraints, *etc.* It is increasingly common for the molecule to be 'heated up' to get over conformational barriers, and then cooled, and for some of the standard forces to be omitted (throughout most of the calculation at least), because of inaccurate biases that they may introduce and in order to speed up the calculation; the method is then known as simulated annealing (SA). The energy terms used in RMD and SA penalise atomic configurations with, for example, non-standard bond lengths and angles, and hence drive the structures towards more standard geometries. These techniques are often used after DG to 'refine' DG structures, and indeed often reduce the NMR constraint violations in the family of DG structures. It has been suggested (though not demonstrated) that RMD and SA are less good at sampling conformational space than DG. Their sampling is of course biased by the imposed non-NMR geometric constraint potentials. This is another reason for the tendency to do DG first, and then 'tidy up' the result with RMD. It should also be said that RMD is more demanding on computer time than DG, although this is not true of SA.

Although the general features of all calculations are similar, the detailed results are different. There are a variety of force fields used in RMD and SA, including CHARMM, XPLOR (developed from and very similar to the CHARMM force field),

AMBER, SYBYL (which uses a force field developed from COSMIC), and GROMOS. Their parametrization is in all cases empirical, and usually designed for molecules other than proteins. The protocols used for simulated annealing also vary widely. This makes it hard to compare results from different calculations, and hard to interpret whether a given structural feature is meaningful. As we shall see, this has important consequences for the determination of peptide structure by NMR.

The way in which electrostatics is handled is especially important for peptides, since the majority of the component amino acids are normally solvent exposed. There is no standard method for handling charges; sometimes they are ignored, sometimes reduced, and sometimes included in full but with large dielectric constants. It is becoming increasingly common to include solvent in calculations, despite the greater computing effort required. The reason for this is that, if calculations are carried out in the absence of solvent, the lack of competition for hydrogen bonding partners leads to a strong tendency to artifactual intramolecular hydrogen bonding.

Distance geometry calculations use only the information that comes from NMR, plus covalent distances. RMD calculations insert a host of other information, much of it empirical. The results of any RMD-based calculation should therefore be treated with some caution. If the calculation shows a hydrogen bond or a particular conformation of a sidechain, is there any experimental evidence for it, or could it be a calculational artifact? The number of unseen parameters going into an RMD calculation is large and often unappreciated by the user (and even more so by the uninitiated reader): it should be a responsibility of all researchers to check that their detailed structural features are justified by the data.

4 Applications to Structures of Rigid Peptides

How do we know if our peptide is rigid? This question should be addressed *before* starting any calculations. Rigidity is a measure of the overwhelming population of a single conformer (in this case a single torsional angle) and hence a high barrier to rotation to or through other conformational states. The necessary depth and narrowness of the free energy well (in torsional space) to define as the cut-off for 'rigidity' is a debatable point. However, as discussed above, small changes in inter-proton distances brought about by torsional angle fluctuations may significantly influence the observed NOE intensities. It should also be noted that rigidity is a property of each torsional angle and of course need not be uniform throughout the molecule.

If, for example, all the peptide torsional angles were rigid, the peptide could be said to have a single conformation (ignoring possible sidechain torsional angle heterogeneity). In this situation, it is likely that the local geometry will be sequence-dependent, and there should be significant variations of J and the temperature dependence (or exchange rate) of amides along the sequence.[14] For reasons given above, the NOE tends to exaggerate structural tendencies; it is therefore highly likely that there will be some kind of sequence-dependent NOE variation. It is certainly necessary that all NOEs be consistent with any proposed structure, although this is not a particularly good guide. As mentioned above, if there is only one structure present, then changes in the solvent composition should have relatively little effect on the structural parameters.

The clearest indication of rigidity is that all backbone atoms should have the same correlation time. In the past, it has been very hard to measure relaxation parameters from heteroatoms, but with the increasing use of proton-detected ^{13}C and ^{15}N, and of biosynthetically directed isotopic labelling, this approach is becoming more possible.

We give a few examples of recent applications.

4.1 DG Approach

Senn *et al.*[15] investigated the conformation of the cyclic hexapeptide, *cyclo*(-Pro-MeTyr-Ala-MeTyr-MeTyr-D-Ala), where MeTyr indicates *N*-methyl tyrosine. They assumed that the backbone would be essentially rigid, whereas the sidechains would be flexible. The conformation was determined following a standard protocol as used for proteins, with the program DISMAN. Thus, DISMAN was applied with a variable target function, to generate a set of structures in good agreement with the distance constraints. A restrained molecular dynamics refinement gave reduced constraint violations and better van der Waals packing; an unrestrained molecular mechanics minimization also gave similar structural changes. Within the limits of accuracy of molecular dynamics calculations, this is a good sign, and implies that the conformations generated are in a good energy minimum. Sidechain conformations were determined by a population analysis of sidechain rotamers using observed coupling constants. Cyclization was imposed by adding nine distance constraints across the peptide bond. An emphasis in this study was to establish how the cyclization restraints should be applied. The question was whether to add these first as a separate stage, or to incorporate them as part of the sixth stage of the variable target function. It turned out that both gave equivalent conformations, but the second method was more efficient.

4.2 RMD Approach

The conformation of endothelin-1 (ET-1) has been investigated in aqueous ethylene glycol using RMD.[16] ET-1 is a 21-residue peptide with disulfide bridges 1—15 and 3—11. From NOESY spectra under a variety of conditions, 137 NOE constraints and 31 non-NOE (*i.e.* lower bound only) constraints were obtained. The number of NOE restraints per residue is fairly high for a peptide, although (as noted above) low for a protein. [Non-NOE restraints should be applied with caution, because NOEs may be unobservable owing to partial saturation or exchange, or to motional effects. Non-NOEs generally do little to tighten the peptide structure, while incorrectly inserting non-NOEs as long lower distance bounds can have major deleterious effects on the calculated structure.] Starting from at least four quite different starting structures, RMD was applied using fairly standard protocols, followed by a steepest descent energy minimization procedure. The interesting feature of this work is that three different RMD packages were used, namely CONGEN, DISCOVER, and XPLOR. The encouraging result is that all three packages produced a similar spread of structures for the more structured core region. For the N- and C-terminal ends, the packages gave somewhat different results. Before minimization, XPLOR produced a much better fit to the NOE constraints than the other two; however, it only did this by accommodating larger distortions of bond angles and large violations of van der Waals repulsions. After minimization, there is very little difference between the three. This is promising evidence that the choice of RMD protocol does not significantly bias the range of structures satisfying the NMR evidence, provided that there are enough NMR constraints, and the RMD is applied intelligently.

4.2 Cyclosporin

The potent immunosuppressor cyclosporin A (Figure 11) is a cyclic undecapeptide. It has received much study over recent years, and there is a crystal structure as well as several different NMR structures. They have recently been compared.[17] The first NMR structure[18] was obtained from data obtained in $CDCl_3$ and C_6D_6, and used NOEs, coupling constants, and amide temperature dependence, as well as heteronuclear $^{13}C\{^1H\}$ NOEs, and the chemical shifts of ^{13}CO and ^{15}N. These data were used to produce a structure by model building. The structure was later refined using RMD. Another structure was obtained by carrying out NMR-constrained RMD, but starting with the crystal structure. Other structures were obtained using DISMAN followed by RMD refinement. All structures were similar. Interestingly, the best DISMAN structure (*i.e.* the one

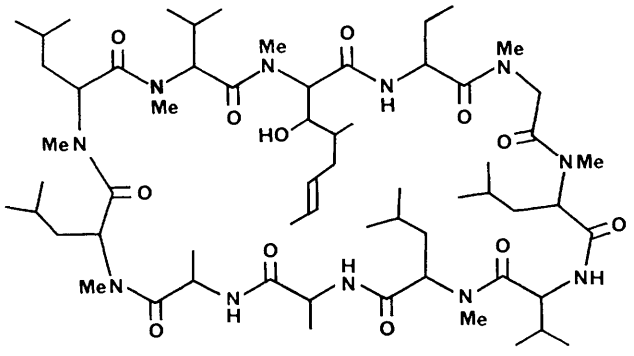

Figure 11 Cyclosporin A.

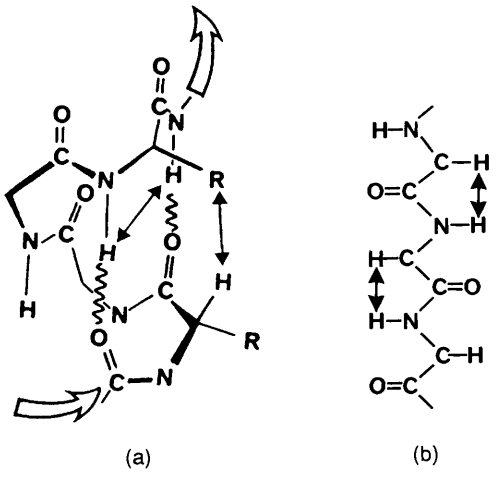

Figure 12 (a) A helical peptide. Two NOEs are shown by double-headed arrows; the right-hand NOE is a $d_{\alpha\beta}(i,i+3)$ NOE diagnostic for helices, and left-hand one is a d_{NN} NOE characteristic of helices, but also of turns. (b) An extended strand, showing the strong $d_{\alpha N}(i,i+1)$ NOE. This NOE is much weaker in helices.

that fitted the NMR constraints best) produced one of the worst fits to the experimental data after RMD refinement; this demonstrates the need for caution in the over-interpretation of 'good' structures, as well as the complexity of the search for the global minimum. It is also noteworthy that RMD could only move two out of the nine DG structures into the 'correct' fold. This illustrates the limitations of RMD in moving out of a local minimum (especially for the more difficult case of a cyclic peptide), and the need for calculating a large number of structures. The authors expressed a strong preference for DG followed by RMD as the best way for generating good structures.[19]

Other structures for cyclosporin have been calculated using a systematic conformational search method, and using a novel method that does an iterative calculation of the position and the variance in the position of each atom, thus providing an explicit quantitation of the error in each atomic position.[17] In terms of the backbone dihedral angles ϕ and ψ, the different structures vary by up to 50°. However, when the structures are compared they look very similar overall, and also very similar to the crystal structure. The greatest differences between structures occur in the sidechain conformations.

5 Fitting Multiple Conformations

In the general case, a peptide populates N structures s_i, each with a probability w_i. The ensemble present in solution is then given by $\Sigma s_i w_i$ (with $\Sigma w_i = 1$), so that for a full description of the peptide conformation one needs a knowledge of all the s and w. The easiest situation to handle is if only one w is non-zero, *i.e.* the rigid peptide described above. *This is rare*, and much rarer than it looks from the literature. For small peptides, rigidity is only likely for a few types of cyclic system. As stated above, NMR will not time-resolve the individual conformations s_i; the interconversion rate between conformers almost always leads to an averaging of NMR parameters. Further, the number of parameters available from NMR is very small, and hence a multiple conformation system is grossly under determined experimentally. Several approaches to describing multiple conformer systems are under investigation, utilizing different simplifying assumptions. We consider below two special cases, and finally a more general case.

5.1 One Distinct Conformation within a Random Coil Ensemble

If a multi-conformer system populates conformations that give rise to characteristic unique NMR parameters, then NMR may be used as a diagnostic technique to detect such populations. Quantitation of this population is not so straightforward (see below). As an example, it is possible to detect the population of helical conformations in some linear peptides in aqueous solution. In all small peptides for which this phenomenon is observed, rapidly interconverting conformational ensembles differing in backbone torsional angles are present. In the case of the myoglobin H-peptide[20] (a peptide corresponding to the H-helix of myoglobin), a range of $d_{\alpha\beta}(i,i+3)$ NOEs is observed throughout the length of the peptide. These medium range NOEs are diagnostic of helical character, in that the $(i,i+4)$ hydrogen bond of a regular α-helix causes these groups to be by far the closest of non-sequential residues. The NOE spectra also contain extensive stretches of $d_{NN}(i,i+1)$ NOEs that are typically found in peptides occupying the α-region of ϕ, ψ space, but which by themselves are not diagnostic for the existence of helical structure. Also observed is an equally intense series of $d_{\alpha N}(i,i+1)$ NOEs – the NOEs that dominate the spectra of peptides in extended conformations (*i.e.* peptides that populate the β-region of ϕ, ψ space) (Figure 12). These latter NOEs are considerably more intense than would be expected in a 100% populated rigid α-helix. The $^3J_{HN\alpha}$ couplings of myoglobin H-peptide are marginally lower than those observed in putative 'random coil' peptides. The peptide therefore contains helical turns, in equilibrium with more extended conformers.

It is not clear how large is the torsional angle range of the conformations that give rise to the medium-range NOEs, or whether the helical turns are present simultaneously all the way along the peptide. The first question is just one of the pitfalls to using NMR to quantitate the population of helical conformations – there may be considerable variation in the $d_{\alpha\beta}(i,i+3)$ distance whilst the peptide still populates helical conformations. A second pitfall is that if parameters are common to both the state to be quantified and other populated states, the average behaviour of all other conformations needs to be known accurately. As for the second question, the simultaneous presence of many turns in these peptides is supported by thermodynamic studies of the helix–coil transition and, in the case of the myoglobin H-peptide, by CD studies (the circular dichroism response having a substantial dependence on the length of the chromophore in helices of this size). CD spectra indicative of helices are not always unambiguously observable for peptides that show medium range NOEs; for example, if the myoglobin H-peptide is shortened in length (myoglobin fragment 1 peptide) the helicity as probed by CD becomes unconvincing, but $d_{\alpha\beta}(i,i+3)$ NOEs are still observable. These studies highlighted a further factor that is crucial to all studies of conformational preferences within peptides – the influence of intermolecular association on the population of conformations and the probes by which they are detected. Specifically, it was shown that as the myoglobin H-peptide associated to a tetrameric state with increasing concentration, the increase of the buildup rate of $d_{\alpha\beta}(i,i+3)$ NOEs from those observed in the monomeric state

significantly preceded the bulk of the chemical shift changes, owing to the strong influence of the change in correlation time in the associated state.

Semi-quantitative attempts have been made to measure helicity in multi-conformer peptides using NMR parameters, making the first-order approximations of uniform correlation times, idealized helical geometry, and isotropic motion. This approach included a description of the NOEs expected for peptides populating a wide range of ϕ, ψ angles in the β-region of ϕ, ψ angle space, again subject to the above assumptions.[21] An important point to come from this and other similar studies of protein fragments, is that the population of secondary structure elements by these peptides is far greater than would be expected given a random distribution through the ϕ, ψ torsional space available in a Ramachandran map. This translates to a considerable conformational free energy in favour of helical states *versus* all other states for these peptides compared with random sequences. This free energy advantage has been implicated in the recognition of native protein conformations by the corresponding anti-peptide antibody, and the non-random distribution of conformations present in the 'unfolded' states of proteins that form on sub-millisecond timescales from the onset of refolding.

Many peptides have been encouraged to adopt helical conformations by adding methanol or trifluoroethanol (TFE) to aqueous solutions. A common reason for doing this is that the peptides in question act in membranes. The co-solvent simulates the lipid environment and reduces the water activity, thus stimulating the peptide to take up its *in vivo* conformation. The method comes as a poor second best to the much more difficult alternative of putting the peptide into a micelle,[22] but is often surprisingly successful. However, there has been a report that helicity in TFE is not necessarily related to the *in vivo* structure of the protein.[23]

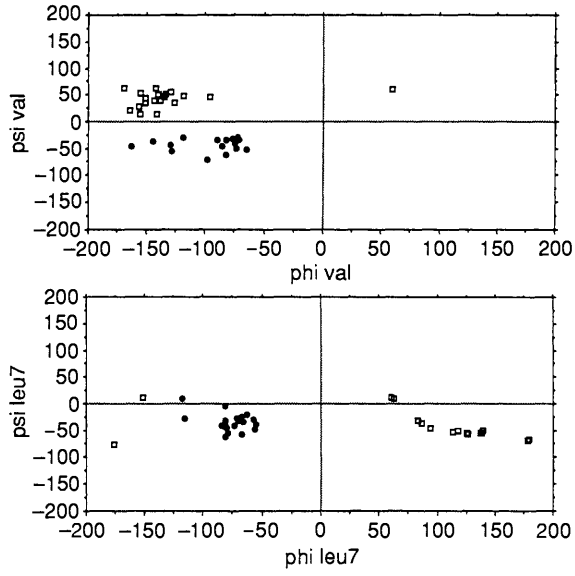

Figure 13 Calculated (ϕ,ψ) plots for residues Val5 and Leu7 of the cyclic heptapeptide *cyclo*(-Ser-Phe-Leu-Pro-Val-Asn-Leu). 500 cyclic conformations were subjected to NOE-constrained RMD, of which the 36 shown here satisfied the NOE constraints best and had the lowest energy. These 36 were then grouped according to whether they correctly predicted the intensities of the NOEs seen. Only the conformers represented by filled symbols met these additional conditions. (Reproduced with permission from reference 24.)

5.2 Limited Number of Conformations in Fast Exchange

This is probably the normal situation for cyclic peptides of 5—8 residues. Already the problem is too complex for NMR alone to provide a complete description; instead, *NMR must be used to restrict the sampling, rather than directly to drive the conformation towards a single 'best' answer* (or group of answers), as was done in Section 4. We note that this approach is not applicable for proteins, because of the enormous range of possible conformations that would have to be generated, although it is certainly applicable to isolated loops in proteins.

One example of this approach is provided by Peishoff *et al.*,[24] who studied the cyclic heptapeptide evolidine, *cyclo*(-Ser-Phe-Leu-Pro-Val-Asn-Leu). They assumed that all the structures would be reasonably similar, and allowed for a range of conformations by specifying NOE constraints as ranges rather than distances. (This is an interesting approach and appears to work well here, but its generality remains to be proven.) An initial set of 500 conformers was generated by DG in torsion angle space, constrained by 21 NMR constraints. Each conformer was then subjected to RMD, and conformers were accepted as good if they had low overall energy and low NOE violations. At this point the conformations generated were put through a further screen, by using additional qualitative NMR constraints that were not used in the original structure calculation. Structures were only accepted if they passed the criteria:

(a) are there no NOEs that should be observed, but are not,
(b) are there no NOEs that should not be observed, but are, and
(c) are the relative intensities of predicted and observed NOEs consistent?

Most of these criteria were included in the original constraint set, but some were not, usually because of the difficulty of including them as constraints. These criteria were able to place further limits on which structures were acceptable (Figure 13), resulting in a much tighter grouping of final structures. This demonstrates the importance of using the experimental data at each step of the analysis to ensure that agreement with the data has not been lost. Encouragingly, further application of *un*restrained MD did not produce any significant change in the conformation.

A slightly different approach is provided by Williamson *et al.*,[25] who studied the cyclic pentapeptide *cyclo*(-Arg-Gly-Asp-Ser-Lys). The conformer set was produced by a geometrical calculation, followed by energy minimization. From each calculated conformer, the expected NOEs and J values were calculated, and were then compared with the experimental NMR parameters. No single structure provided a good fit, but the observed parameters could be well reproduced by assuming rapid exchange between three of the low energy conformers. This should not be taken to mean that the three low energy conformers 'represent' the solution structure, but rather that the solution structure is in rapid exchange between a number of structures, many of which resemble the three low energy conformers. Alternatively, a good fit could be obtained by using a single structure generated by RMD refinement of the best structure. However, the NMR parameters are highly suggestive of a dynamic structure. The single structure therefore probably represents a misleading 'averaged' structure with no physical reality.[2]

5.3 Multiple Conformations

This case may be expected to be the normal one for linear peptides, of which the situation described in Section 5.1 is a special case. When there are many conformers in fast exchange, NMR cannot be used to generate the structures s_i; this must be done by some non-NMR method. All NMR can do is to place limits on the possible combinations of w_i, preferably in combination with results from other techniques. It should be noted that any approach of this kind is clearly heavily dependent on the choice of initial structures s_i. As noted above, the calculation of low energy structures is one of the least satisfactory parts of the structure calculation process at present. A promising approach is being developed by Nikiforovich *et al.*[26] A set of possible structures s_i is generated by dividing up conformational space into large regions; typically a peptide would have 4^{n-1} possible structures s_i, corresponding to the four quadrants of ϕ,ψ space for each of the n residues. A search is then made for sets of

w_i that give results consistent with the experimental observations. Monte Carlo methods are used to pick individual structures from within each conformational region s_i. The result is a large number of sets of structures s_i with their associated weights w_i, all of which would generate the observed NMR results. It is found that some conformations s_i have non-zero w_i in all sets; in other words, they are essential to a description of the conformation. Others have w_i very close to zero in all sets, and so can be definitely excluded (Figure 14). The method was applied to Leu-enkephalin in DMSO,[27] where it was found that the central region of the peptide could be represented by only two conformations, the others having values of w_i close to zero. This result represents a much more reliable and 'scientific' proof of a conclusion that would otherwise be made simply by inspection. A conceptually similar method has been proposed independently by other workers.[28]

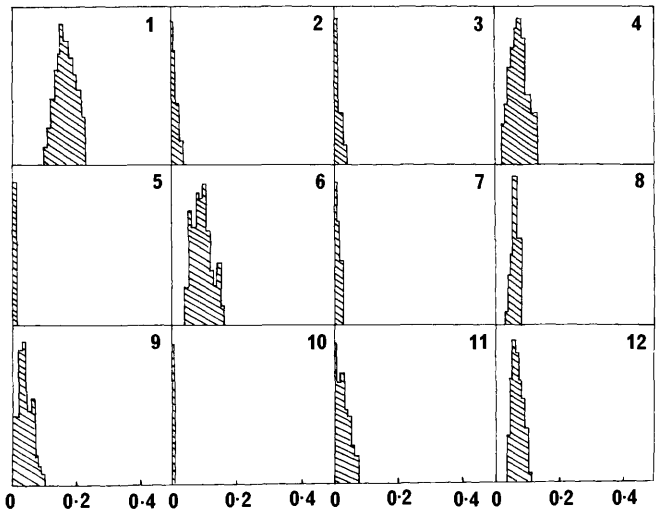

Figure 14 Conformational frequencies calculated for different conformers of a spin-labelled derivative of angiotensin. Panels 1 to 12 represent 12 families of conformers (s_1 to s_{12}). Using Monte Carlo methods, a large number of sets of probabilities w_i were calculated, used to weight the probability of the different conformers ($w_i s_i$), and tested for consistency with experimental observations. The histograms depict the w_i found for each family in all successful calculations. Families 1, 4, 6, 8, and 12 have non-zero probability in all successful sets, and are therefore indispensible – from the calculations, these families are the minimum necessary for a description of the conformation of the peptide that agrees with experimental data.
(Adapted from reference 26.)

6 Future Prospects

For systems that can be described as essentially rigid, existing techniques now being applied to proteins will be extended further, such as stereospecific assignment of prochiral methylene protons and methyl groups. Back calculation of NOEs will also become important for establishing the accuracy with which the sets of structures correspond to the NMR spectrum. Improvements in RMD will be essential, in particular to account for electrostatic effects. Explicit inclusion of solvent will undoubtedly become popular, especially given the rapidly increasing power of computers available. It will become increasingly important to be able to establish the conformational flexibility of a peptide, in order to provide a proper interpretation of the results. Much more use will be made of proton-detected heteronuclear relaxation parameters, as well as heteronuclear coupling constants.

For mobile systems, there is an urgent need for new methodology. The problem is not merely that of the small number of NMR parameters available, it includes the much wider question of how one represents the 'structure' of a mobile molecule. Perhaps a structure should be represented not by a 'best' molecule, or even by a superimposition of a group of molecules, but by a movie? Such questions are important for proteins as well as peptides, since the absence of rigidity is a widespread phenomenon, and loops and turns in proteins appear to have similar dynamic properties to peptides.

Finally, we come to the thorny question of the relevance of peptide structure in solution to the biologically important conformation, which is its structure bound to a receptor. There is little direct evidence on the question, since there are few crystal structures of receptor-bound peptides. Some evidence comes from structure–activity studies on peptide hormones and the use of rigid non-peptide analogues. This evidence is fragmentary and equivocal. While it can be interpreted to show that the active (bound) conformations of many peptides are close to those found in solution, it is also clear that the bound conformations of some peptides are different from any seen for the peptide in solution. One example is given below; another is the (Arg-Gly-Asp) sequence, which has been shown largely to populate β-turn conformations in solution, but is almost certainly extended in the active conformation.[29]

There has long been interest in the question of what is the best solvent to use to mimic the binding site in a receptor. Theoretical studies suggest that receptor binding sites are hydrophobic; in addition, many peptide hormones bind to membrane-bound receptors. This would suggest that solvents with low dielectric constants should be the most realistic. However, although a receptor may be generally hydrophobic, it has polar groups placed at the appropriate places to hydrogen bond with the peptide. Thus, for similar reasons as *in vacuo* molecular dynamics calculations lead to structures that are too highly internally hydrogen bonded, conformational studies in apolar solvents may do the same. However, it has been reported that a series of enkephalin amides fold in $CDCl_3$ into conformations that are consistent with known receptor models, whereas in DMSO they are 'random coil'.[30]

The best option is clearly to study the conformation of the bound peptide directly. This implies an increased use of the transferred NOE, which makes use of the fact that the NOE builds up much more rapidly in the bound state than in the free peptide. Thus, in a mixture of bound peptide with 5—20-fold excess of free peptide, NOEs are characteristic of the bound state, provided that the dissociation rate is fast enough.[7] The technique is applicable even to very large receptors.

An even more elegant way of studying bound conformation is to isotopically label the peptide and study the complex directly, using heteronuclear editing techniques. Because the NMR observations are made on the peptide in the bound state, the problems of linewidth and spin diffusion inherent to studies of large proteins are still present. This limits the method to relatively small, soluble receptors. It is, however, an extremely powerful technique. It has been used recently to study the conformation of cyclosporin bound to cyclophilin.[31,32] Two studies were carried out independently, both using cyclosporin biosynthetically labelled with ^{13}C (and ^{15}N in one case). Both studies used isotope-edited 2D methods, one using additional 3D methods. Both arrived at very similar bound conformations. The most striking point to come out of these studies is that the bound conformation of cyclosporin is totally different from the solution conformation: it has all-*trans* peptide bonds and no internal hydrogen bonds, whereas the solution structure has one *cis* peptide bond and four hydrogen bonds (Figure 15). This result is all the more striking because of the similarities in structure between the NMR and crystal structures, which give no suggestion of any other accessible conformation. This implies that we still have a long way to go before we can fairly claim to understand peptide structures in solution.

7 References

1 K. Wüthrich, *J. Biol. Chem.*, 1990, **265**, 22059.
2 M. P. Williamson, in 'Methods in Molecular Biology', ed. C. Jones, B. Mulloy, and A. H. Thomas, Humana Press, New Jersey, 1992, Vol. 7, in press.

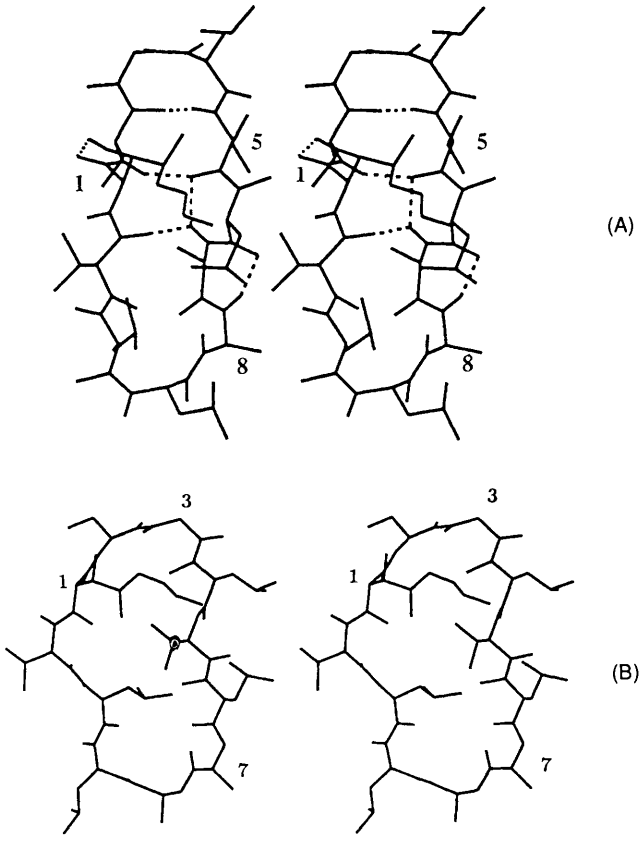

Figure 15 Cyclosporin conformations in solution obtained from NMR data, shown as stereoviews. Hydrogen bonds are shown by broken lines. (A) Conformation in chloroform. (B) Conformation bound to cyclophilin.
(Reproduced with permission from reference 31.)

3 H. J. Dyson and P. E. Wright, *Ann. Rev. Biophys. Biophys. Chem.*, 1991, **20**, 519.
4 H. Kessler, *Angew. Chem., Int. Ed. Engl.*, 1982, **21**, 512.
5 T. E. Creighton, 'Proteins', W. H. Freeman and Company, New York, 1983, p.326.
6 J. Feeney, G. C. K. Roberts, J. P. Brown, A. S. V. Burgen, and H. Gregory, *J. Chem. Soc., Perkin Trans.* 2, 1972, 601.
7 R. Boelens, T. M. G. Koning, G. A. van der Marel, J. H. van Boom, and R. Kaptein, *J. Magn. Reson.*, 1989, **82**, 290.
8 D. Neuhaus and M. P. Williamson, 'The Nuclear Overhauser Effect in Structural and Conformational Analysis', VCH, New York, 1989.
9 S. Ludvigsen, K. V. Andersen, and F. M. Poulsen, *J. Mol. Biol.*, 1991, **217**, 731.
10 V. F. Bystrov, *Progr. Nucl. Magn. Reson. Spectrosc.*, 1976, **10**, 41.
11 T. G. Pedersen, B. W. Sigurskjold, K. V. Andersen, M. Kjær, F. M. Poulsen, C. M. Dobson, and C. Redfield, *J. Mol. Biol.*, 1991, **218**, 413.
12 H. J. Dyson, M. Rance, R. A. Houghten, R. A. Lerner, and P. E. Wright, *J. Mol. Biol.*, 1988, **201**, 161.
13 A. G. Palmer, M. Rance, and P. E. Wright, *J. Am. Chem. Soc.*, 1991, **113**, 4371.
14 H. Kessler and W. Bermel, in 'Applications of NMR Spectroscopy to Problems in Stereochemical and Conformational Analysis', ed. Y. Takeuchi and A. P. Marchand, VCH, Florida, 1986, p.179.
15 H. Senn, H.-R. Loosli, M. Sanner, and W. Braun, *Biopolymers*, 1990, **29**, 1387.
16 S. R. Krystek, D. A. Bassolino, J. Novotny, C. Chen, T. M. Marschner, and N. H. Andersen, *FEBS Lett.*, 1991, **281**, 212.
17 R. Pachter, R. B. Altman, J. Czaplicki, and O. Jardetzky, *J. Magn. Reson.*, 1991, **92**, 468.
18 H. R. Loosli, H. Kessler, H. Oschkinat, H.-P. Weber, T. J. Petcher, and A. Widmer, *Helv. Chim. Acta*, 1985, **68**, 682.
19 J. Lautz, H. Kessler, J. M. Blaney, R. M. Scheek, and W. F. van Gunsteren, *Int. J. Pept. Protein Res.*, 1989, **33**, 281.
20 J. P. Waltho, V. A. Feher, H. J. Dyson, and P. E. Wright, *Biochemistry*, in press.
21 E. K. Bradley, J. F. Thomason, F. E. Cohen, P. A. Kosen, and I. D. Kuntz, *J. Mol. Biol.*, 1990, **215**, 607.
22 W. Braun, G. Wider, K. H. Lee, and K. Wüthrich, *J. Mol. Biol.*, 1983, **169**, 921.
23 S. R. Lehrman, J. L. Tuls, and M. Lunall, *Biochemistry*, 1990, **29**, 5590.
24 C. E. Peishoff, J. W. Bean, and K. D. Kopple, *J. Am. Chem. Soc.*, 1991, **113**, 4416.
25 M. P. Williamson, J. S. Davies, and W. A. Thomas, *J. Chem. Soc., Perkin Trans. 2*, 1991, 601.
26 G. V. Nikiforovich, B. Vesterman, J. Betins, and L. Polins, *J. Biomol. Struct. Dynamics*, 1987, **4**, 1119.
27 B. Vesterman, J. Saulitis, J. Betins, E. Liepins, and G. V. Nikiforovich, *Biochim. Biophys. Acta*, 1989, **998**, 204.
28 C. Landis and V. S. Allured, *J. Am. Chem. Soc.*, 1991, **113**, 9493.
29 H. Mickos, J. Bahr, and B. Lüning, *Acta Chem. Scand.*, 1990, **44**, 161; and M. H. Hann, personal communication.
30 P. A. Temussi, T. Tancredi, A. Pastore, and M. A. Castiglione-Morelli, *Biochemistry*, 1987, **26**, 7856.
31 C. Weber, G. Wider, B. von Freyberg, R. Traber, W. Braun, H. Widmer, and K. Wüthrich, *Biochemistry*, 1991, **30**, 6563.
32 S. W. Fesik, R. T. Gampe, H. L. Eaton, G. Gemmecker, E. T. Olejniczak, P. Neri, T. F. Holzman, D. A. Egan, R. Edalji, R. Simmer, R. Helfrich, J. Hochlowski, and M. Jackson, *Biochemistry*, 1991, **30**, 6574.