PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL PHYSICAL & ENGINEERING SCIENCES

Protein nuclear magnetic resonance spectroscopy in the new millennium

Pfuhl M.

Phil. Trans. R. Soc. Lond. A 2000 **358**, 513-545
doi: 10.1098/rsta.2000.0544

THE ROYAL SOCIETY

Protein nuclear magnetic resonance spectroscopy in the new millennium

By M. Pfuhl and P. C. Driscoll

Department of Biochemistry and Molecular Biology, University College London, Gower Street, London WC1E 6BT, UK

Solution-state nuclear magnetic resonance (NMR) spectroscopy is a rich source of information that can be exploited to elicit the three-dimensional structure of proteins, the nature of their interactions with other molecules, as well as biological function and dynamic properties. Even though NMR was established in the field of chemistry by the early 1950s it was not until the early 1980s that the first three-dimensional solution structure of a small protein was determined. From that time on, however, NMR has come to play a major role in the field of structure–function research on proteins and other biological macromolecules. It would indeed be difficult to imagine that some of the latest developments in this field, for instance the rapid establishment of many larger proteins as mosaic multi-domain assemblies of independent folding units or our recent understanding of protein folding pathways, without the insights provided by NMR spectroscopy. Despite the substantial impact already contributed by the application of NMR to solve biological problems, it is perhaps still arguable that only a fraction of the experimental parameters that can be derived from NMR spectroscopic examination of proteins have so far been fully exploited.

In the last decade, NMR spectroscopy has been boosted by enormous technical improvements, which strive to bypass the classical bottlenecks of structure–function studies of proteins. As a result of these new developments, a greater number of experimental NMR parameters can now be interpreted in a meaningful way, while others have recently become accessible for the first time. The turn of the century therefore appeared poised to witness a new spurt in both the development of new NMR techniques and the expansion of their routine application in protein research.

The problems that have been plaguing protein NMR spectroscopists for many years—the bewildering complexity of overcrowded spectra, which can be impossible to analyse, fast nuclear relaxation in large molecules (molecular weight greater than 20 000) leading to low sensitivity, the relative paucity of experimental constraints in the calculation of three-dimensional molecular structures, for example—appear to have been overcome within a few years by the cooperative effect of technological and methodological innovations. These developments include the extension of isotope labelling from ^{15}N to ^{13}C and ^{2}H, the introduction of highly stable superconducting magnets with ever-increasing homogeneous magnetic-field strengths of 20 T (corresponding to a proton NMR frequency of 800 MHz) and higher, and the exploitation of the experimental consequences of newly rediscovered physical phenomena, such as the partial alignment in solution of proteins in strong magnetic fields or liquid crystals, and the interference effects of different mechanisms contributing to nuclear relaxation.

It is therefore anticipated that the current pace in the development of NMR spectroscopy into a yet more powerful tool will speed up in the new millennium rather than slow down.

In this paper, we will describe the basic principles behind the most important of the recent developments in protein NMR spectroscopy, which include aspects of spectrometer hardware and software, NMR experiments, isotope labelling and data analysis. These facets will then be discussed in terms of sample applications to illustrate their use as practical tools in addressing biological and biophysical phenomena at the molecular level.

Keywords: NMR; protein; structure; relaxation; isotopes; dynamics

1. Introduction

Spectroscopy is the science of the interaction of matter with electromagnetic radiation, and is typically characterized by the presentation of the pattern of intensity of absorption by the target sample for a range of applied frequency (approximately 1/wavelength) of the applied radiation. NMR spectroscopy describes the phenomenon of the absorption of radiofrequency radiation that leads to excitation of the nuclear spin-states in the target molecules (see below), and is a method that has found great prominence in the chemical analysis of materials, particularly in the solution state. The major practical requirement for the application of NMR spectroscopy is that the sample be placed in as high a magnetic field as possible, as only then are the nuclear energy levels sufficiently differentiated to lead to a detectable NMR response (see equation (2.1)).

The analysis of protein structure and function by NMR spectroscopy has several advantages compared with other approaches. In principle, all NMR measurements can be performed in aqueous solution under conditions identical to those used in biochemical assays, and, thus, can be chosen to be arbitrarily close to the physiological *in vivo* case. Compared with other spectroscopic methods, NMR spectroscopy does not rely on specific reporter groups, e.g. aromatic side chains of tyrosine or tryptophan residues, or artificially attached dyes to yield a signal. Instead, essentially every single atom can become observable via its resonance line in the NMR spectrum. Thus, the entire protein can be monitored in a direct manner at atomic resolution without substantial intervening calculations (as in X-ray crystallography). The further advantage of NMR over X-ray crystallography is the circumvention of the problems involving the crystallization of proteins. Firstly, not all proteins can be coaxed into crystallization, and, secondly, a number of detailed features of a protein structure obtained in the crystal state can be distorted by crystal packing interactions and the presence of high concentrations of cosolvents, required to induce crystallization. In contrast, there is a substantial array of parameters that can be extracted from the analysis of resonance lines in the NMR spectrum of a protein, and these can be used to probe a vast range of structural and functional features. These properties range from the simple identification and characterization of ligand interactions via the determination of binding and acidity constant (pK_a) values, the mapping of ligand-binding surfaces, and the elucidation of polypeptide folding pathways, up to the determination of the three-dimensional molecular solution structure together with a description of its dynamic properties on time-scales varying from picoseconds to hours.

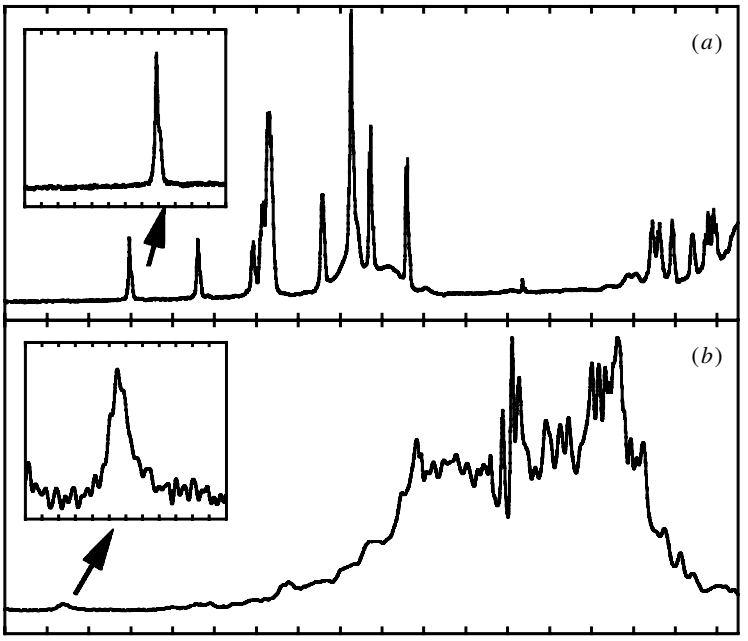

Figure 1. Parts of a one-dimensional ^{1}H (proton) NMR spectra of an 11-residue peptide (*a*) and of a 406-residue protein (*b*). Both spectra were measured at a ^{1}H resonance frequency of 600 MHz (with a magnetic field equal to 14 T) and a temperature of 25 °C. The spectra were scaled to compensate the different sample concentrations. The portions shown cover a region with a spectral width of 6000 Hz, usually populated by resonance lines of amide protons. The leftmost resonance line in each spectrum is shown magnified in an inset for each spectrum that covers 600 Hz. Note that the resonance line in the case of the large protein is very broad with a high noise level, while the resonance line in the peptide is sharp with little noise.

It goes almost without saying that since the initial discovery of the NMR phenomenon in the mid-1940s, NMR has had a powerful and indispensable impact on the practice of synthetic and analytical chemistry, both of small molecules in solution and of solid powders, crystals, polymers and glasses. In the medical field, the uses made of the NMR phenomenon to produce two- and three-dimensional magnetic resonance images (MRIs) are equally impressive. These applications of NMR are not discussed further here, since, in many respects, NMR in these fields has bedded down into disciplines with rather separate technical and practical aspects, which are beyond the scope of the structural biologist.

Despite these evident advantages of the technique, the impact of the investigation of proteins over the first decades of NMR spectroscopy was somewhat limited. Two major problems were responsible for its limited applicability despite its theoretical potential. First, the fact that almost every single atom in a protein has the potential to generate an NMR signal in a spectrum means that the NMR spectroscopist has the difficult task of finding any given signal of interest amongst what can be a very large number—a 100 amino acid polypeptide chain yields around 1000 proton NMR lines, for example—giving rise to a situation that is not unlike trying to find the proverbial needle in a haystack (or, better, one needle in a needlestack). Second, the complexity is further exacerbated by the limited 'space' available in the spec-

trum, which limits peak resolution: the spectral bandwidth available for the NMR signals of a given nuclear isotope is restricted, and for the very complex cases of proteins this leads, necessarily, to resonance overlap and eventual overcrowding of the spectrum.

An example of this overlap and overcrowding is given in figure 1, where the one-dimensional ^{1}H NMR spectra of a small peptide (11 amino acids) and of a much larger globular protein (406 amino acids) are compared. In these spectra, a peak of positive intensity corresponds to the specific absorption of radiofrequency radiation by a single proton or a chemically and magnetically equivalent group of protons (e.g. a methyl group). At first glance, it is clear which is the spectrum of the small peptide and which is the spectrum of the larger protein: for the peptide spectrum, the vast majority of resonance lines are resolved. For the protein, the spectrum consists essentially of broad bands of unresolved intensity comprising many overlapping resonances. A few resonance lines at the left edge of the spectrum are not overlapped, but here another characteristic becomes evident. The typical width of the resonance lines is much larger for the protein than for the peptide. An increased width for the protein resonances goes along with a reduction in the resonance peak height (to first order, the area of the signals is a constant). Therefore, not only does a greater width of a resonance line increase the resonance overlap, it also affects the apparent sensitivity of an NMR experiment. The sensitivity of any experimental investigation is given by the signal-to-noise (S/N) ratio and is typically specified as the ratio of the signal height (rather than area or volume) to the level of the noise generated in the experiment, most of which derives from the apparatus that is used to perform the measurement, in this case an NMR spectrometer (see figure 10).

In technical terms, the spectroscopic lines detected in NMR spectra of proteins are broad (up to tens of hertz) and the typical spectral dispersion is relatively limited (for hydrogen *ca.* 8000 Hz at available magnetic-field strengths). The NMR signal of a nucleus within the protein will be present in the spectrum, but, more likely than not, it will be overlapped by many others. Therefore, simple observation of a particular NMR signal may already be a challenge, and this leaves aside the problem of being able to 'assign' which resonance line in a spectrum belongs to which atom in the target protein in the first place. A secondary issue is to find NMR parameters that can be measured with a sufficient degree of precision and accuracy to be useful for addressing the particular biological question at hand, for example, to describe the three-dimensional solution structure of the molecule or to attribute an acidity constant (pK_a value) to a specific amino acid side chain. Suffice it to say that many of these challenges for the application of NMR to the study of ever larger protein molecules have been met with considerable success, to the point where NMR is one of the methods of choice for studies of protein structure and biochemistry.

The introduction of multidimensional NMR spectroscopic methods in the 1980s and isotopic enrichment of proteins with ^{15}N and ^{13}C in the 1990s has produced a quantum leap in the impact of NMR spectroscopy on structural biology. Let us take the determination of three-dimensional structures of proteins as an example. As recently as 1990 there were only six entries based on NMR data in the protein data bank (PDB): a repository of information describing the three-dimensional structures of proteins and their complexes with ligands. By 1994 this number had increased to 179, and by the beginning of 1999 a total of 1459 NMR solution structures had been deposited.

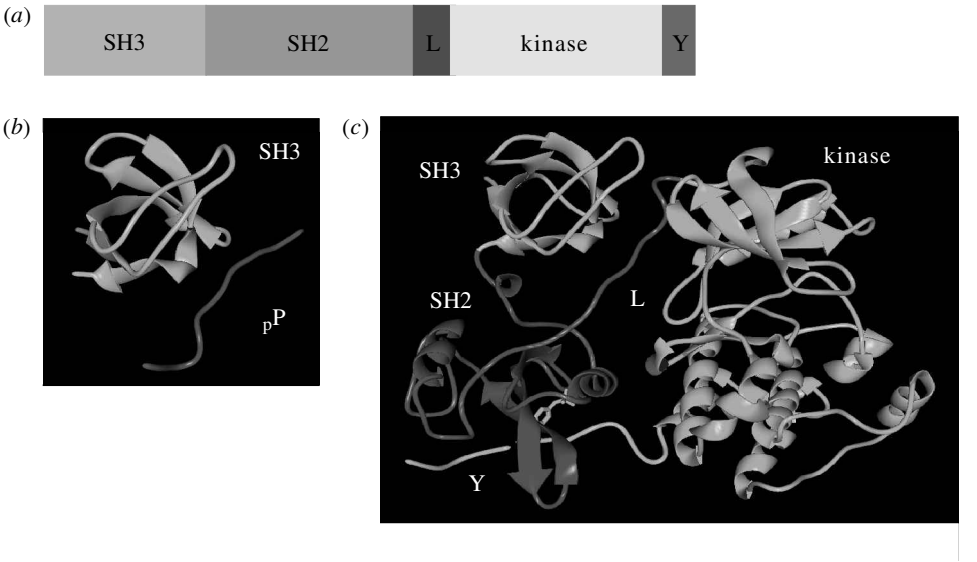

Figure 2. Overview depicting the aspects of structural studies on protein–protein interactions in different molecular environments. (*a*) Cartoon representation of the structure of the *src* tyrosine kinase, depicting the types of globular domains present in the linear amino acid sequence. (*b*) Crystal structure of an SH3 domain from fyn complexed to a proline-rich peptide (pP) from the phosphoinositide 3-kinase. (*c*) Crystal structure of the autoinhibited form of *src*. Note that the type of interaction of the SH3 domain in the more complex system of *src* kinase with the segment labelled L is very similar to the interaction found in a situation where the isolated SH3 domain has been studied. The linker peptide (L) is not a proline-rich sequence and yet is able to bind to the SH3 domain as if it were a proline-rich sequence. Such an interaction, which is highly important to biological function, could only be identified in the intact protein.

Of these, however, almost half (709) contain fewer than 50 amino acids, roughly a third (539) have between 50 and 100 residues, 221 have between 100 and 150 residues and only 77 are of proteins larger than 150 residues. By comparison, the example of *src* tyrosine kinase in figure 2 has 531 residues. The size distribution of other protein NMR data in the literature (resonance assignments, dynamics parameters) is very similar, reflecting the problem that despite the technological advances in the field over the last 20 years a substantial obstacle towards a uniform application of NMR spectroscopy to the study of protein structure and function lies in the limits imposed by the molecular weight. To continue to provide answers to current topics of research in biology, there is natural pressure for NMR spectroscopy to push at the boundaries imposed by the molecular weight.

Until relatively recently, many structural biologists could content themselves with the pursuit of important structural information on small isolated protein domains. An example of this approach is the family of *src* homology 3 (SH3) domains, which have been the focus of substantial effort on the part of both NMR spectroscopists and X-ray crystallographers (see figure 2). The main questions that could be addressed were the nature of the polypeptide fold of these domains, their evolutionary relationships, and principles of their function, not least the strong interactions with proline-rich peptides. Ironically, it was the molecular-weight limitations of NMR spectroscopy

that contributed to the rapid establishment over the past decade of the concept of 'mosaic' or 'modular' proteins, composed of small autonomously folded domains (Campbell & Baron 1991). Being limited to the study of small fragments of larger proteins, NMR spectroscopists have been forced to probe the structure of larger proteins by biochemical excision of stably folded modules. Such 'divide and conquer' approaches to the examination of mosaic protein structures have met with a considerable degree of success. However, along with this progress has come a realization that a complete picture of the biological activity can only be obtained when protein modules are studied in the context of the intact protein, or at least in complexes with relevant ligand molecules. Again taking up the example of the SH3 domain, this concept implies that a full understanding of its functional role would best be derived by the investigation of the intact proteins from which the individual SH3 and proline-rich segments are derived. The substantial potential of this type of approach was recently demonstrated by the determination of crystal structures of auto-inhibited protein kinases (Sicherl *et al.* 1997; Xu *et al.* 1997).

These ideas serve only to increase the imperative for the scope of NMR spectroscopy to be expanded to ever-larger protein molecules. Fortunately, the last few years have seen a number of completely new approaches to overcome some of the difficulties associated with the study of larger proteins. Of these, we have chosen to highlight the most exciting and qualitatively different aspects that are also expected to play a substantial role in years to come.

2. Deuteration of proteins

(*a*) *Some essentials about nuclei, spins and spectra*

The most essential nucleus used in protein NMR spectroscopy is the proton (from the element hydrogen). The isotope ^{1}H, which makes up 99.985% of all hydrogen in nature and is thus present in sufficient amounts in any protein, happens to be one of the most sensitive nuclei available for NMR investigations. The other elements in proteins that are suitable for NMR are nitrogen (the NMR-active ^{15}N isotope is present at a level of only 0.37% at natural abundance, which leads to the requirement for artificial enrichment to make NMR experiments possible) and carbon (the NMR-active ^{13}C isotope is present at a level of only *ca.* 1.1% in nature, thus also requiring enrichment to make most modern NMR experiments feasible). The nuclei of these three isotopes are all characterized by a spin quantum number of 1/2. NMR is fundamentally a quantum-mechanical phenomenon, and the fullest treatments of the theory are somewhat complex. From the classical standpoint, a 'spinning' (i.e. rotating) charge has an associated magnetic field, best described as a magnetic dipole as in a simple bar magnet. It therefore possesses, like any piece of magnetic material, a north pole and a south pole. In complete analogy to a compass, which only works because it is in the Earth's magnetic field, the magnetism of a nuclear spin would go by completely unnoticed unless brought into contact with another magnetic field. As the needle of a compass will reorient, so will the magnetic dipole of a nucleus orient itself with respect to a strong magnetic field. Since the nuclear magnetic field is so weak, quantum laws apply. In contrast with a compass, which is allowed only one orientation with respect to the Earth's magnetic field, a spin is allowed two: parallel or antiparallel to the external field, which results in two quantized energy levels. Quantum numbers for these states are $+1/2$ or $-1/2$ (usually referred to as the α

or β spin states). The energy difference ΔE between these two states, even in the highest artificial magnetic fields, is still actually only very small, so that, according to the Boltzmann law (equation (2.2)), both states are almost equally populated (h is Planck's constant; γ is the gyromagnetic ratio of a nucleus—a measure of the strength of its magnetic field—and B_0 is the strength of the applied magnetic field; N_0 and N are the populations of the ground and excited states, respectively):

$$\Delta E = \frac{h}{2\pi}\gamma B_0, \tag{2.1}$$

$$\frac{N}{N_0} = \exp\left(\frac{-\Delta E}{RT}\right). \tag{2.2}$$

At room temperature ($T = 295\,\mathrm{K}$) on a 600 MHz spectrometer (magnetic-field strength $B_0 = 14\,\mathrm{T}$; for comparison, the strength of the Earth's magnetic field is 0.001 T), the energy difference for a hydrogen leads to a ratio N/N_0 of approximately 1×10^{-5}, i.e. only 1 out of 100 000 molecules in the sample will be able to interact with externally applied electromagnetic radiation. The consequence is that, compared with most other spectroscopic methods, any protein sample used in an NMR experiment appears to be 'diluted' by a factor of 100 000. In NMR, the situation is completely dissimilar to optical spectroscopy, where, essentially, all molecules are in the ground state and are thus available to excitation. As indicated in equation (2.1), the energy difference depends on the strength of the magnetic field B_0, so that the energy gap and, thus, the N/N_0 ratio, will increase with increasing B_0 (as a rule of thumb, the signal-to-noise ratio in NMR spectroscopy scales with $B_0^{1.75}$). The consequence of these properties is that a premium is placed on the sensitivity of NMR experiments, and much of the emphasis to develop ever-higher magnetic-field strengths is driven by the desire to improve upon what is a very weak spectroscopic phenomenon.

(*b*) *Nuclear relaxation and its importance in NMR spectroscopy*

The investigation of proteins by NMR spectroscopy takes advantage not only of the individual responses of different nuclei (which give rise to specific resonance frequencies, usually named chemical shifts), but also targets the different types of magnetic interactions that arise between these nuclei. The network of covalent bonds that make up the chemical structure of the protein—as illustrated in figure 3 for the polypeptide chain of a protein—mediates one mechanism of cross-talk between spins, usually referred to as scalar coupling. The strength of this cross-talk—which can be transferred over up to four, or sometimes five, bonds—is given by the scalar coupling constants, denoted J, as illustrated for the example of an H_N–N–C_α fragment in figure 3. Using specific NMR experiments designed to exploit the scalar coupling interactions, connections between the NMR signals of different nuclei can thus be recognized in fragments, which will ultimately allow the identification of all individual spins. This is a fundamentally important procedure, central to the analysis of NMR spectra, which has come to be known as the process called resonance assignment. The example shown in figure 3 illustrates, in a schematic way, a multidimensional NMR spectrum that has one spectral axis for the amide hydrogen signals, another axis for the amide nitrogen signals, and a third axis for the α-carbon signals. By performing a number of different experiments of this type, essentially all ^{15}N, ^{13}C and ^{1}H nuclei in a protein spectrum can be unambiguously identified, as indicated in figure 7.

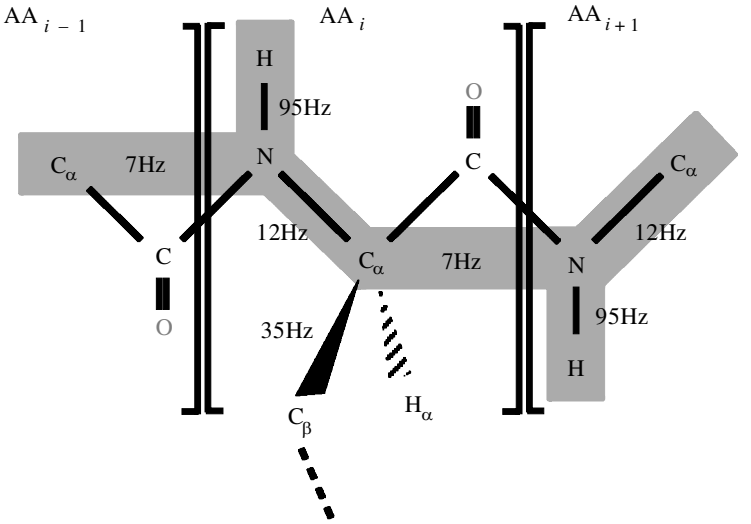

Figure 3. Schematic representation of the correlation of various spins in a segment of the backbone of a $^{13}C/^{15}N$ isotope-enriched protein. The correlation of the three nuclei is indicated by the shaded bars. The size of the coupling constants is given next to the corresponding correlation. The corresponding spectrum can be seen in figure 7.

In the implementation of these 'correlation' experiments, an excited state of a single set of nuclei, e.g. the backbone ^{15}N, is first produced. It is only from this excited state that the cross-talk between neighbouring spins takes place. The NMR experiments take advantage of this cross-talk between spins to transfer the excitation from the originating nucleus to another site, e.g. the $^{13}C_\alpha$ nucleus. The simple-minded extension of this idea, that, in principle, one could engineer the excitation to 'hop' from nucleus to nucleus through an entire amino acid, or from amino acid to amino acid, thereby defining the correlation pattern of a very large number of NMR signals corresponding to distant parts of the molecular framework, is appealing. Unfortunately, the correlation of nuclear spins is only one of the processes that can occur in the course of an NMR experiment. A competing process that must always be taken into account is the finite rate with which nuclei simply leave the excited state and return back to the ground state. This is an important and universal characteristic of all types of spectroscopy, usually referred to as relaxation. The build-up of a spin–spin correlation takes place over a time period that is inversely proportional to the magnitude of the scalar coupling constant, and during which competing processes divert the signal into unproductive pathways including nuclear relaxation. It turns out that while nuclear relaxation is generally a rather slow process, compared with relaxation processes in other spectroscopies, it nevertheless provides one of the fundamental limitations to the scope of NMR spectroscopy.

The main limitation of the extent to which coherence transfer can be accomplished is, therefore, the lifetime of the excited state involved. In some sense, spectroscopists would prefer that excited spin states would ideally never relax back to the ground state, or at least decay sufficiently slowly that all required manipulation can be achieved with high efficiency. Alas, nuclear relaxation and the strong dependence of the relaxation rate on molecular size are facts of life, and NMR spectroscopists have

a vested interest in understanding and exploiting the nature of their origins. Again, the underlying physics of nuclear relaxation is complex, and only a brief outline is given below. A variety of mechanisms contribute to the decay of an excited nuclear state. For proteins there are two dominant sources of nuclear relaxation: dipole–dipole interactions and chemical shift anisotropy (CSA). A detailed knowledge of the physical principles underlying these two mechanisms and the modes by which they mutually interact with each other has formed the basis of the recent developments to attenuate nuclear relaxation to improve the quality of NMR spectra in general and to exploit these mechanisms to extract information about protein structure, function and dynamics.

(*c*) *Dipolar spin–spin relaxation*

The energy inherent in an excited state can be exchanged through space between any pair of nuclei by the magnetic field of one nuclear dipole that is sensed at the position of the other nuclear dipole, and vice versa. This is the essence of the so-called dipole–dipole interaction that contributes to magnetization exchange and nuclear relaxation. Let us imagine a 'protein' molecule with only two atoms that are fixed in space. As a consequence of the mutual interaction, the total magnetic field at the position of one spin is slightly increased or decreased, depending on the relative orientation of the axis connecting the two nuclei with respect to the direction of the magnetic field. As a consequence, the resonance line will be shifted away from the position in the absence of a neighbouring spin. Each spin will see the other in the α state in one half of the molecules and in the β state in the other half, with α and β being antiparallel. The fields have opposite signs, thus shifting the resonance line in opposite directions in each half of the molecule. This gives the impression that the original resonance line is split into two new lines, each with half intensity. The distance between the split lines is called the dipolar coupling constant, D. In this hypothetical static system, the dipolar coupling has a fixed magnitude.

In order to contribute to nuclear relaxation, a variation of the dipole–dipole interaction has to take place, driven by reorientation of the axis connecting the pairs of dipoles. Rotational tumbling of the protein in solution will lead to fluctuating variations in the field around one spin as a result of the presence of the other one. If the fluctuations occur at frequencies close to the resonance frequency of the nucleus in question, magnetization can be exchanged between the two spins. The process is analogous to the effect that radio waves created by a radio transmitter have on the antenna in a distant radio receiver. The transmitting antenna on its own is a dipole, but that does not have any effect on a radio. The transmitting antenna only comes to life once it is connected to an oscillating voltage leading to the creation of fluctuating magnetic fields. It is only these fluctuating magnetic fields that are picked up by the radio receivers.

A simplified form of the equations that describe the relaxation rate (R) for dipolar nuclear interactions is given in equation (2.3) (μ_0 is the magnetic susceptibility of the vacuum; h is Planck's constant; r_{IS} is the distance between the two nuclei; γ_{I} and γ_{S} are the magnetogyric ratios of the two spins usually labelled I and S):

$$R \approx d_{\mathrm{DIP}} \sum_{i=1}^{n} J(\omega_i), \tag{2.3}$$

$$d_{\mathrm{DIP}} = \frac{\mu_0^2 h^2 \gamma_{\mathrm{I}}^2 \gamma_{\mathrm{S}}^2}{(16\pi)^2 r_{\mathrm{IS}}^6}, \tag{2.4}$$

$$J(\omega) \approx \frac{\tau_m}{1 + \omega^2 \tau_m^2}. \tag{2.5}$$

The dynamic nature of the nuclear relaxation process is introduced in expression (2.3) via the so-called spectral density function $J(\omega)$, as defined in equation (2.5). $J(\omega)$ provides a measure of the power inherent in molecular movements that are occurring for any given frequency ω present in the random motions of a protein characterized by the tumbling time τ_m, available to drive NMR spectroscopic transitions. The tumbling time, τ_m, of a protein is reasonably well approximated by the Einstein–Stokes formula, $\tau_m = MV\eta/RT$, so that τ_m increases in proportion to molecular size and decreases with increasing temperature (assuming constant viscosity). For biological macromolecules and at typical resonance frequencies used in protein NMR, $J(\omega)$ will also increase with τ_m. To summarize, the larger a protein molecule is, the slower it will tumble in solution, leading to faster relaxation of excited nuclear spin states. Faster relaxation means bigger linewidths, as demonstrated in figure 1, and thus a reduced S/N.

Because of the inverse sixth-power contribution of the distance between the dipoles (see equation (2.4)), it is only rather short-range dipolar interactions that make a substantial contribution to the overall relaxation rate. The other important parameter to consider is the gyromagnetic ratio γ. Nuclei with a low γ value contribute less to the relaxation than nuclei with a high γ value.

The flip-side of the dipolar interaction phenomenon is that, aside from being detrimental to the quality of the NMR spectrum by effecting the loss of signal with time, if appropriately addressed by appropriate NMR measurements, it can also be a very useful phenomenon for protein NMR. By measuring the rate at which the signal that is lost on one spin arrives at another spin, it is actually possible to estimate distances between pairs of spins. Again, the strong dependence of the magnitude of the dipolar interaction (in this context giving rise to the so-called nuclear Overhauser effect, or NOE) upon the separation distance restricts the useful application to a very short range. In practice, this is usually less than 5 Å, which is very short compared with, for example, the length of a compact protein of 100 amino acids, which is *ca.* 50 Å. Nevertheless, it is this exploitation of the dipolar interaction that has provided for the vast majority of the experimental input for solution structure determination of protein molecules (see § 2 *c*).

(*d*) *Effects of deuteration*

In order to have a chance to attenuate the deleterious nuclear relaxation that arises for large molecules, we have little choice but to manipulate some of the adjustable parameters given in equation (2.3). Among the parameters contained in the coefficient d_{DIP}, only the gyromagnetic ratios γ_{I} and γ_{S} are accessible to any manipulation by the NMR spectroscopist, through selection of the nuclear isotopes present in the molecule. The distances between the nuclei are either fixed by covalent bonds or are themselves the actual target of the NMR investigation. Fortuitously, nature has provided us with a number of alternative isotopes for the most important elements in protein NMR: hydrogen, carbon and nitrogen are all available as different 'versions' (isotopes), each with a distinct γ value. With respect to nuclear relaxation,

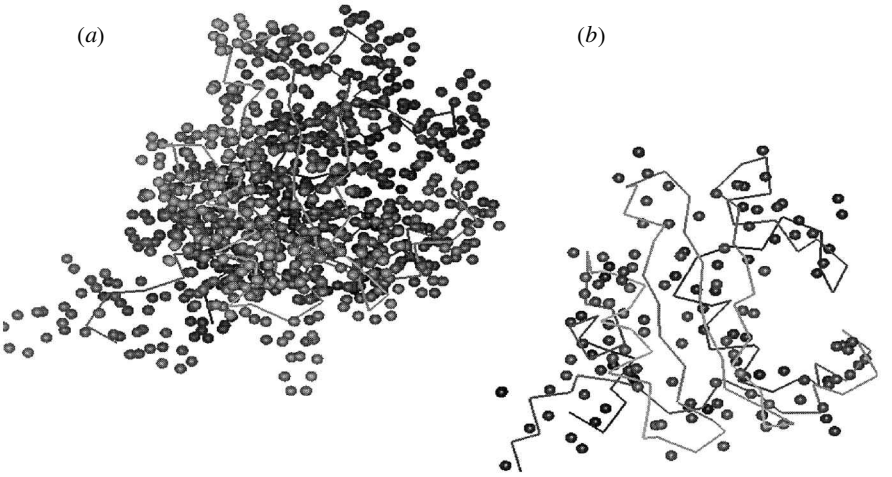

Figure 4. Illustration of the effect of deuteration on the density of protons in a protein. An *src* homology 2 (SH2) domain from phosphoinositide 3-kinase. (*a*) The protein with all hydrogen atoms shown, a state that represents a protein that is fully protonated. (*b*) The same protein as in (*a*) but with only the amide protons retained, a state in which all aliphatic protons are replaced by deuterons. The coloured balls indicate the position of the hydrogen atoms in each case.

the isotope most worthwhile to consider replacing is the proton ^{1}H: firstly, it has the largest γ value (ten times greater than ^{15}N and four times greater than ^{13}C); secondly, the γ of its heavier isotope ^{2}H (deuteron) is seven times smaller, giving the (theoretical) maximal reduction of d_{DIP} by a factor of 49.

Experimental replacement of the hydrogen isotope ^{1}H (proton) with ^{2}H (deuteron) dates back several decades, when deuteration was used to simplify proton spectra by removing ^{1}H signals from a spectrum. The applications of such selective deuteration techniques were later replaced by uniform labelling techniques, based on the over-expression of recombinant proteins in bacteria growing in heavy water ($^{2}H_2O$, often denoted D_2O) (Torchia *et al.* 1988; LeMaster & Richards 1988). Samples produced in such a way are perdeuterated, but, by taking advantage of the fact that amide hydrogens (NH) exchange with hydrogens from solvent, proteins can be prepared that are deuterated at aliphatic and aromatic positions (i.e. bound to carbon), but essentially fully protonated on the amide positions, as depicted in figure 4. In proton NMR spectra of samples prepared in such a manner, the aliphatic and aromatic regions are essentially empty, while the region of the amides is fully represented, as can be seen in figure 5. The amide region—the left half of the spectrum—is of similar intensity in the two spectra, while the aliphatic region—the right half of the spectrum—is much weaker in the deuterated protein. Having lost the majority of immediate neighbouring protons, relaxation of the amide hydrogens becomes substantially damped. Relaxation rates are typically reduced by a factor of about 3 in proteins deuterated at the level of *ca.* 85% of their aliphatic/aromatic sites (Markus *et al.* 1994). A direct consequence of the reduced relaxation rates is the narrowing of the width of the remaining ^{1}H resonance lines. As shown in the inset of figure 5, the resonance lines of amide hydrogens are narrower in the deuterated than in the proto-

nated sample. The measurement of NMR spectra for the determination of distances between nuclei is then facilitated in several ways. The reduced linewidth not only reduces the overcrowding of the spectrum, but also helps to improve the apparent S/N ratio.

In addition, the application of deuteration to attenuate nuclear relaxation helps to reveal a further aspect that arises particularly for larger molecules. For normal proteins containing large numbers of hydrogen nuclei all close to each other (see figure 4), magnetization is exchanged by dipolar interactions in complex pathways. In the analysis of relaxation experiments for the determination of internuclear distances, it is generally presumed that the magnetization transfer occurs directly through space without any interventions. In the densely packed ensemble of nuclei in the interior of a protein, however, the transfer between a given pair of sites very often occurs via a relay nucleus. Because of the r^{-6} dependence of dipolar relaxation (see equation (2.4)), such a relayed transfer can be faster than the direct transfer covering the same distance. Converting a relaxation rate for such relayed transfer pathways into an internuclear distance will inevitably produce an underestimate. In addition, the high density of nuclei will prevent efficient direct transfer over long distances: there is simply too great a chance that an intervening relay nucleus will play its role.

Figure 4 gives a clue to the effect that perdeuteration has on the problem of spin diffusion. The reduction in the number of nuclei available for relaxation pathways also removes potential relay stations in the indirect transfer routes for magnetization exchange (Torchia *et al.* 1988). Therefore, for the remaining NH protons, this approach not only makes the measurement of internuclear distances more reliable, but it also allows measurement over greater distances, perhaps as far as 8 Å (Mal *et al.* 1998). The regular elements of secondary structure in proteins are characterized by close proximity of sequentially connected backbone amide groups: *ca.* 2.5 Å in an α-helix; *ca.* 4.2 Å in β-sheets. The latter distance would be on the limit of the measurement in a fully protonated protein, but does not pose a problem in a moderately deuterated protein. Once extracted from the NOE spectra of a deuterated protein, the large number of amide–amide distances can be used to accurately determine the elements of secondary structure. In the case of β-strands, the topology (i.e. the arrangement of several β-strands into β-sheets) might also be recognized from this type of data. For a highly detailed structure determination, the practical utility of these long distances (Venters *et al.* 1995) is, however, limited, because a very high degree of deuteration has to be achieved (greater than 95%). With only one backbone NH per amino acid, an insufficient number of distances can be obtained with this method alone to determine the structure of the protein. Not always, however, is a very detailed three-dimensional structure the principal aim of an NMR study. Often it is sufficient to delineate the overall backbone fold of a protein. For such a limited task, the sparse set of distances based only on amide hydrogens might indeed be sufficient (Mal *et al.* 1998). In the context of the current drive towards the systematic analysis of genomes at the level of protein structures ('structural genomics'), the experimental NMR data could be combined with sequence alignments and other theoretical methods from homology model-building up to *ab initio* prediction to improve the assignment of fold and function. Important conclusions could thus be obtained as the substrate specificity of a hitherto unidentified enzyme, even before the determination of a detailed three-dimensional structure and the mapping of ligand binding sites is completed (or even started).

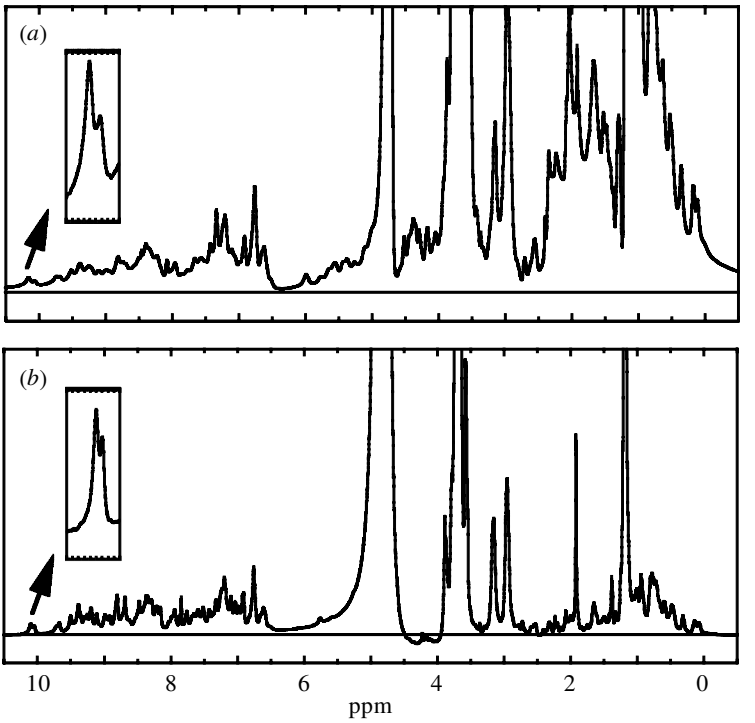

Figure 5. One-dimensional ^{1}H (proton) NMR spectra of a dimeric fragment (dimer molecular mass 24 kDa) of the regulatory subunit of phosphoinositide 3-kinase measured at 600 MHz proton frequency (a magnetic field of 14 T) and a temperature of 25 °C. (*b*) Sample produced in *E. coli* grown on minimal medium with ^{15}N ammonium sulphate, normal glucose and heavy water. (*a*) A similar sample to that in (*b*) but produced with normal water instead of heavy water. While the signals in the region between 5 and 11 ppm (this is the region of the amide protons) are essentially unchanged, a substantial reduction of signal intensity in the aliphatic region (0–6 ppm) is seen. Note the absence in (*b*) of the well-resolved peak at 6 ppm in (*a*) as well as a number of other α-proton peaks between 5 and 4 ppm. On the other hand, some aliphatic signals, most notably the sharp line at *ca.* 2 ppm, have almost the same intensity. The insets show the two leftmost peaks in both spectra to illustrate the reduced linewidth in the case of the deuterated protein.

NOE spectra of perdeuterated (above 95% deuteration) proteins can perhaps more importantly contribute when used for resonance assignment purposes, as was recently demonstrated for the human immunodeficiency virus (HIV-1) *Nef* protein (Grzesiek *et al.* 1995). Using the improved spectral resolution resulting from the line-narrowing effect from the attenuation of relaxation pathways, it proved very straightforward to obtain the assignment by the identification of a nearly uninterrupted chain of sequential distances between amide protons.

The problem of substantial overcrowding in the spectra of *Nef* was, in part, caused by unfolded segments of the polypeptide chain. It was, however, possible to identify which amino acids were unfolded using the above-mentioned strategy. It was also observed that the stretches of amino acids preceding and following this unfolded segment formed a β-hairpin (i.e. two β-strands consecutive in the amino acid sequence connected by hydrogen bonds). It was, therefore, clear that the beginning and the

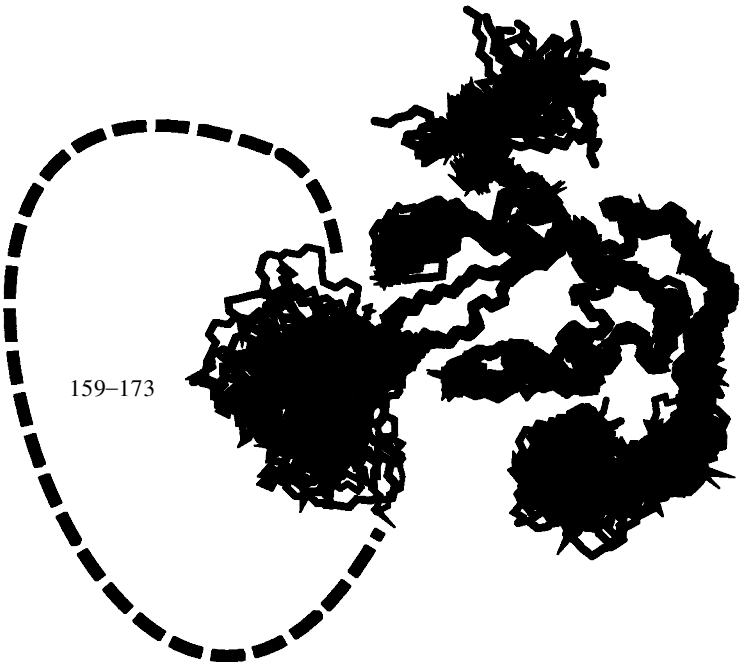

Figure 6. Solution NMR structure of HIV-1 *Nef* (PDB entry 2nef). The construct used to determine this structure had the N-terminal 39 amino acids deleted, as well as a stretch of 30 amino acids in the middle of the sequence (159–173). The ensemble of structures typical for structure calculations from NMR data is shown. Well-defined regions of the structure show little deviation within the ensemble, while ill-defined segments (e.g. the loops and the termini) show large deviations within the ensemble.

end of the unfolded segment must be very close in space. Consequently, it was decided to remove most of the amino acids in the unfolded segment. The structure of the modified protein is shown in figure 6. As predicted, the unfolded loop—indicated by the dotted line—is attached to a β-hairpin in the core of the structure. Most of the remaining residues connecting the ends of the hairpin are ill-defined in the structure (Grzesiek *et al.* 1997), indicating their mobility. It was only the preliminary experiments using deuterated samples that allowed the design of a more accessible protein. Not only the structure was determined but also some of its biological functions—binding to CD4 and SH3 domains—could be characterized by NMR spectroscopy (Grzesiek *et al.* 1996*a*, *b*).

Even more impressive effects can be realized for the use of perdeuterated protein samples in heteronuclear correlation (scalar coupling) experiments applied to carbon-13 enriched proteins. Dipolar relaxation by the directly attached hydrogen is the dominant relaxation pathway for carbon-13 nuclei in proteins. The shorter bond length (*ca.* 1 Å for an H–C bond, compared with *ca.* 1.5 Å for the aliphatic C–C bond) and the larger gyromagnetic ratio ($\gamma_{\mathrm{H}}/\gamma_{\mathrm{C}} = 4$) make the relaxation of a carbon by its covalently bonded hydrogen theoretically 180 times faster than by a directly bound carbon-13 atom. This is of particular importance for a number of heteronuclear resonance experiments that rely on the C_α atom as a relay for both connections along the polypeptide chain (as indicated in figure 3), as well as for

mapping backbone–side chain correlations. It is the C_α that usually exhibits the fastest relaxation rates of all carbon atoms in any amino acid. One of the major benefits of deuteration is the attenuation of the relaxation rate for the C_α nucleus in larger molecules, as demonstrated in figure 7.

(*e*) *Methods for protein deuteration*

Currently, the major source of proteins for structural study is recombinant ('genetically engineered') bacteria, which can be made to over-express the target molecules in a defined culture medium. The main source of hydrogen atoms in the bacterial synthesis of proteins is the bulk solvent water. Replacing 'normal' water with 'heavy' water (^2H_2O or D_2O) in the culture medium is, therefore, a simple mechanism for obtaining highly deuterated proteins. Bacteria are sufficiently robust to be able to grow in such a medium, albeit at slightly slower rates than normal. In addition to the water, a proportion of hydrogen atoms can also be derived from the carbon- and nitrogen-containing nutrients. For the production of proteins for NMR spectroscopy, bacteria are usually grown in a 'minimal' medium. This contains a number of minerals and vitamins at very low concentrations, as well as ammonium salts enriched in ^{15}N and glucose enriched in ^{13}C. In such a medium, degrees of enrichment higher than 97% are routinely achieved for ^{13}C, and labelling with ^{15}N is often obtained with higher than 99% efficiency. To reach similar degrees of enrichment in ^2H, besides the heavy water one requires a carbon source (often glucose) that has itself been chemically perdeuterated. For heteronuclear resonance applications, the ^2H glucose must also be ^{13}C labelled. The benefits of achieving very high levels of deuteration are perhaps not yet so widely appreciated. The relatively high costs associated with doubly isotope-enriched chemicals makes this approach uneconomic for many applications, though increased demand and the economies of scale may provide for more cost-effective use of such materials in the future. An inspection of the metabolic pathway charts of *E. coli* reveals, however, that high degrees of C_α position enrichment in ^2H can readily be obtained using a D_2O medium containing glucose only enriched in ^{13}C (i.e. a protonated carbon source). The reason this works satisfactorily is that the α position in all amino acids receives the hydrogen only from water. Such a cost-effective enrichment strategy has been shown to yield essentially 99% enrichment of the H_α position with *ca.* 85% overall deuteration.

A rather different approach, which involves partial deuteration of proteins, is based on the selective incorporation of fully protonated hydrogen positions, e.g. methyl groups (Gardner & Kay 1997; Rosen *et al.* 1996), into an otherwise perdeuterated protein using biosynthetic methods. The idea behind this strategy is that methyl-group NMR signals relax relatively slowly (due to fast internal rotation), even in very large proteins. The deuteration of all other C–H sites further reduces methyl-group nuclear relaxation rates, so that high-quality spectra can be measured that allow the measurement of more distances than can be obtained from NH groups alone (see above) (Zwahlen *et al.* 1998*a*, *b*). In addition, the methyl-group-containing amino acids—mainly valine, leucine and isoleucine—tend to be involved in making up the hydrophobic core of proteins. The distances derived from methyl protonated samples can, thus, give information about the packing in the interior of the protein, making them complementary to the NH-based distances, and they should add considerably to the aim of defining folds of unknown proteins in a rapid and reliable fashion.

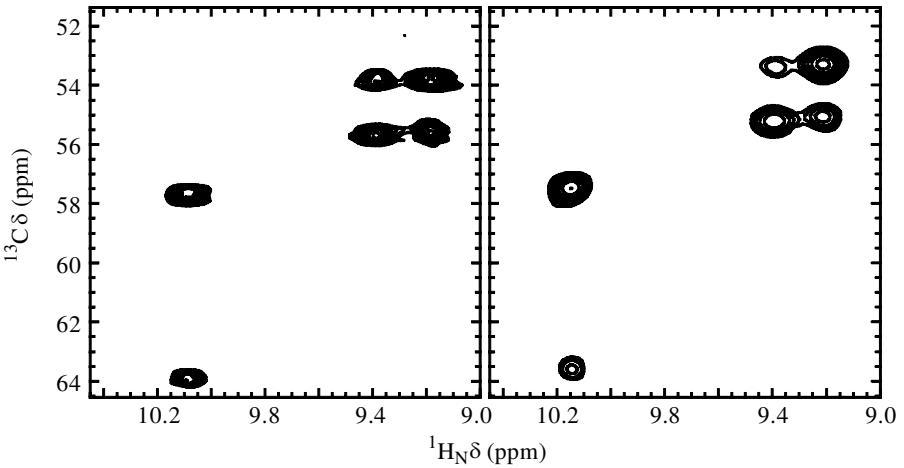

Figure 7. Two-dimensional cross-sections from a three-dimensional HNCA spectrum (producing the correlation of nuclei as described in figure 3). Note that there are two carbon correlations per amide proton chemical shift, corresponding to the α carbon of the own amino acid and the α-carbon of the preceding amino acid, of a dimeric fragment (dimer molecular weight 24 kD) of the regulatory subunit of phosphoinositide 3-kinase, measured at 500 MHz proton frequency (a magnetic field of 11 T) and a temperature of 25 °C. On the left-hand side is shown a small part of the spectrum measured on a ^{2}H/^{15}N/^{13}C-labelled sample; on the right-hand side is shown the corresponding part of the spectrum measured on a ^{15}N/^{13}C-labelled protein. The narrower linewidth and the substantially higher peak intensities are apparent. Note that the resolution of the spectrum on the deuterated sample is so good that the splitting caused by the non-decoupled ^{1}J scalar coupling between C_α and C_β has become evident.

Assignments and calculation of protein folds based on this strategy have already been demonstrated for proteins in the molecular mass range of 40–50 kDa (Gardner *et al*. 1997, 1998).

A number of proteins and protein complexes involving nucleic acids and carbohydrates have been successfully characterized using heteronuclear resonance experiments applied to samples uniformly enriched in ^{13}C, ^{15}N and ^{2}H (Shan *et al.* 1996; Gardner *et al.* 1998; Venters *et al.* 1996; Caffrey *et al.* 1997). Molecular masses routinely covered by backbone resonance assignments are now in the range of up to 40–50 kDa, while the maximum molecular weight attained with this strategy is approaching the 70 kDa mark.

3. Interference of relaxation mechanisms

(*a*) *Nuclear relaxation arising from chemical shift anisotropy*

In the field of NMR spectroscopy, the strictly accurate term ‘resonance frequency’ for the position of a resonance line in a spectrum is rarely used. Instead, NMR spectroscopists usually refer to ‘chemical shifts’. For the chemist or the biochemist the resonance frequency as a property of a nucleus is fairly dispensable. The more interesting feature is the effect different *chemical* environments have on the resonance frequency, hence the term ‘chemical shift’. For any given nucleus giving rise to an NMR signal, the pattern of covalent bonds to attached atoms, the different types of

non-bonding contacts and varying molecular charge densities and electron densities will give rise to a particular 'shift' of the NMR resonance frequency away from the frequency it would exhibit in an entirely isolated state.

The underlying nature of the chemical shift (resonance frequency) of an NMR signal is, therefore, somewhat misrepresented by the simple (scalar) number suggested by the presence of one resonance line per nucleus in a high-resolution solution NMR spectrum. Instead, the chemical shift is best represented by a tensor whose components reflect the three-dimensional nature of the shielding effects provided by the surrounding molecular structure. Since the chemical shift of any nucleus is influenced considerably by the distributions of surrounding electronic and magnetic fields, the anisotropy of the environment around a given nucleus is particularly evident for the atoms in a folded protein. Depending on the details of the electronic structure, the magnetic field at the position of the nucleus will be more or less reduced (or 'shielded') compared with that for a bare nucleus. The distribution of electrons around a nucleus can be highly anisotropic, depending, in particular, on the covalent bonding network. The effective field at the position of the nucleus thus depends on the orientation of the molecule with respect to the external magnetic field. In solution, the fast reorientation ('tumbling') of a protein averages out the variation in chemical shift, so that, indeed, only one resonance line is seen per nucleus. However, the anisotropy of the chemical shift combined with the diffusive rotational molecular motion leads to fluctuations in the local magnetic field with an influence similar to the presence of another nearby nuclear dipole (as described in § 2 *c*). Consequently, this fluctuation also contributes to the relaxation of a nucleus from an excited state towards its ground state.

The functional form of the contribution to the nuclear relaxation rates from CSA is shown in equation (3.1). Here, $\sigma_{\|}$ and $\sigma_{\perp}$ are the axial and the perpendicular components of the chemical shift tensor, assuming that the CSA tensor is axially symmetric. Usually, the axial component is closely aligned with a covalent bond axis, while the perpendicular component is much smaller and lies at right angles to the bond:

$$R \approx d_{\mathrm{CSA}} \sum_{i=1}^{n} J(\omega_i), \tag{3.1}$$

$$d_{\mathrm{CSA}} = \tfrac{1}{3}(\sigma_{\|} - \sigma_{\perp})\omega_{\mathrm{I}}^2 = \tfrac{1}{3}(\sigma_{\|} - \sigma_{\perp})\gamma_{\mathrm{I}}^2 B_0^2. \tag{3.2}$$

So far, remedies against fast transverse relaxation in proteins have been sought almost exclusively in the analysis of the physics of dipolar relaxation (as described in § 2 *c*). Not so much can be achieved by consideration of the relaxation via the CSA alone, since, as equations (3.1) and (3.2) reveal, the main parameters $\sigma_{\|}$ and $\sigma_{\perp}$ are features of the covalent structure of the molecule, while the spectral density terms $J(\omega)$ are subject to the same molecular characteristics as apply to dipolar relaxation.

(*b*) *Interference between dipolar and CSA relaxation*

An important aspect of the influence of CSA upon nuclear relaxation has, however, recently been rediscovered and exploited in a series of NMR experiments ideally suited for application to larger proteins (Pervushin *et al.* 1997). While it might be expected that the relaxation rates caused by different processes simply sum to give a

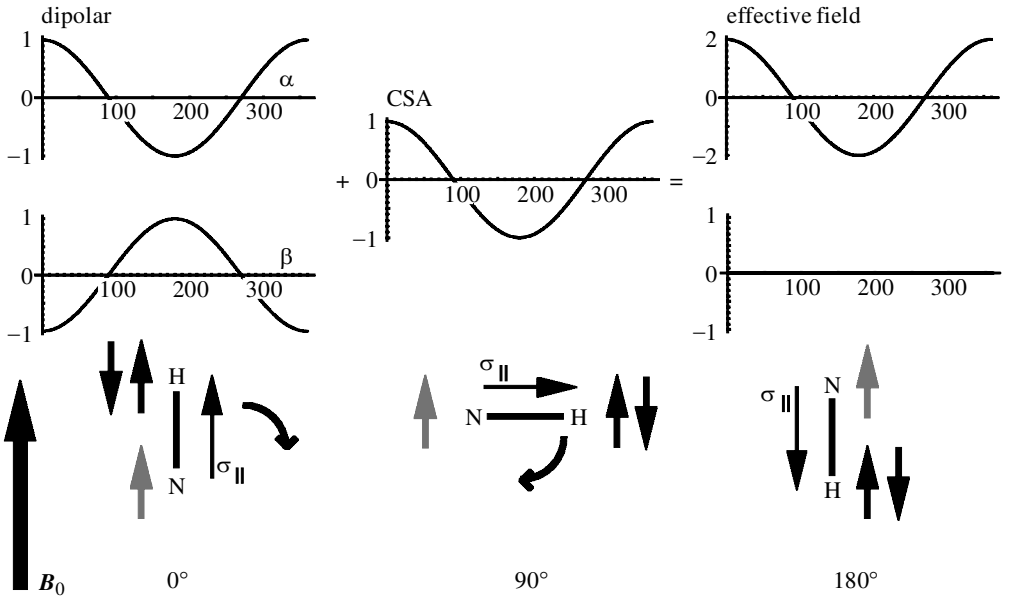

Figure 8. Schematic of interference of ^{15}N–^{1}H dipole–dipole interaction and ^{15}N CSA for the relaxation of ^{15}N in an N–H spin pair. Some orientations of a backbone N–H group relative to the external field are depicted in the bottom half of the figure, while the fluctuations corresponding to these orientations are shown in the top half. The two possible orientations of the ^{1}H spin are represented by the antiparallel pair of black arrows in the bottom half and by the two sine oscillations, which are 180° out of phase, on the top. The CSA of the ^{15}N nucleus is indicated by the arrow labelled $\sigma_{\|}$, since it is assumed for simplicity that $\sigma_{\|} \gg \sigma_{\perp}$.

total effective rate, it is known that certain mechanisms cannot be treated in such an independent manner. Instead, it is said such effects interfere with one another. The interference of relaxation by CSA and dipolar interactions can be best understood by considering the physical principles of each mechanism. Both rely on the creation of fluctuating magnetic fields at the position of the excited nucleus. Even though the source of fluctuating fields—in the case of dipolar relaxation, the magnetic dipole of another nucleus; in the case of CSA relaxation, the variation in the chemical shielding resulting from the immediate bonding structure—is different, the nucleus that relaxes senses the overall combination of static and fluctuating fields. The time dependence of the dipolar and CSA fields is similar for each mechanism, since both are mainly driven by the stochastic rotational tumbling of the molecule. The inflexibility of the bond means that both influences are subject to the same dynamic fluctuations.

A simple *Gedankenexperiment* can then reveal the fundamental nature of the interference. Let us take simple one-dimensional sinusoidal oscillations as highly simplified versions of the fluctuating dipolar and CSA fields arising from random tumbling of the protein in solution. The resulting total field can be simply constructed by knowing the amplitude and phase of each oscillation. If the phase difference is zero, both waves will simply add in a constructive manner (as indicated in the upper pathway of figure 8). But when both oscillations are 180° out of phase, a substantial damping of the total effective field will occur. The maximal interference effect is a complete cancellation of both oscillations if the two amplitudes are identical (as indicated in

the lower pathway of figure 8). In such an idealized situation, where CSA and dipolar interaction are the only relaxation mechanisms present, a nucleus would not be subject to relaxation effects at all.

Moving to the situation pertaining to real samples, the two fields from dipolar and CSA interactions fluctuate in three-dimensional space. The maximal cancellation effects arising from interference are thus strongly influenced by the relative orientation of the principal component of the chemical shift tensor ($\sigma_{\|}$ in the axially symmetric case) and the vector that connects the two dipolar coupled nuclei. The orientation of $\sigma_{\|}$ is strongly dependent on the precise nature of the covalent bond structure, and thus tends to be different for different bonded atom pairs (e.g. N–H versus C–H, etc.).

The backbone N–H pair of atoms is particularly suited to the experimental exploitation of interference effects, since ^{15}N $\sigma_{\|}$ is almost parallel to the N–H bond direction. Only a parallel orientation of $\sigma_{\|}$ will allow the theoretical maximum cancellation effect. As depicted in figure 8, the oscillations of the dipolar and the CSA fields have their respective maxima and minima at the same orientation. The orientation of $\sigma_{\|}$ from the ^{15}N CSA tensor is approximately coincident with the N–H bond direction. Thus, the instantaneous chemical shift will reach a maximum when the N–H bond is parallel to the external field, and a minimum when the N–H bond is antiparallel to the external field. The same is true for the dipolar N–H interaction, which is maximal when both dipoles are 'one on top of another': that is, when the N–H bond is either parallel or antiparallel to the external field. In the picture of a simple one-dimensional oscillation, both fields in the N–H pair fluctuate perfectly in phase, supported by the fixed geometry of the N–H unit. A 180° phase shift can then be produced by a simple inversion of a spin from $\alpha \rightarrow \beta$ or $\beta \rightarrow \alpha$. As pointed out in equations (2.1) and (2.2), the populations of α and β (i.e. parallel or antiparallel to the external field) spins are almost equal. Thus, for an N–H group, in 50% of the molecules the ^{15}N nucleus will sense the ^{1}H nucleus in the α state. For the other 50%, the ^{15}N nucleus will sense the ^{1}H nucleus in the β state. Consequently, the sign of the fluctuations caused by the dipolar interactions is inverted for one-half of the molecules in the sample. In contrast, the sense of the CSA relaxation field for the ^{15}N nucleus is always the same, because it does not depend on the spin states of the bonded ^{1}H nucleus. Therefore, for a given N–H group, one-half of the molecules in the sample are subjected to the combination of fluctuating dipolar and CSA relaxation fields, corresponding to the 180° phase shift required to achieve maximal interference of the two relaxation pathways. The final consideration required to estimate the potential extent of the attenuation due to interference effects is the relative amplitude of the two oscillating fields. The relaxation of the excited state of the ^{15}N nucleus for an N–H group is described in the following approximation,

$$R \approx (\sqrt{d_{\mathrm{DIP}}} \pm \sqrt{d_{\mathrm{CSA}}})^2 \sum_{i=1}^{n} J(\omega_i), \tag{3.3}$$

which is essentially a combination of equations (2.3)–(3.2). The sum and difference in the first bracket correspond to the values obtained for the two possible spin states—α or β—of the covalently bonded proton. For one-half of the molecules (corresponding to the difference case), the relaxation is slowest when $d_{\mathrm{DIP}} \approx d_{\mathrm{CSA}}$. For the other half (corresponding to the sum case), the relaxation rate is accelerated. For a

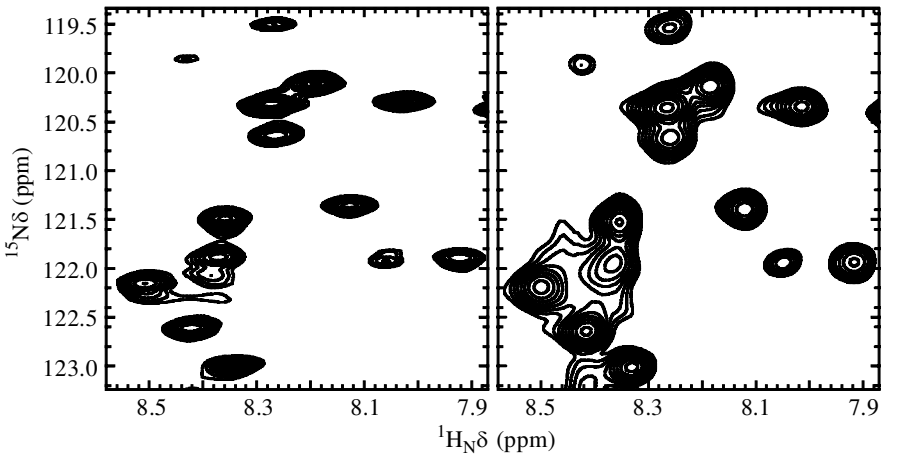

Figure 9. Comparison of a selected region from two-dimensional ^{1}H/^{15}N correlation (right) and TROSY (left) spectra for a 24 kD fragment of phosphoinositide 3-kinase recorded at 600 MHz and 25 °C. Note that the lines are sharper in the vertical direction corresponding to relaxation interference for the ^{15}N resonance. The relaxation interference effect for the ^{1}H resonances (horizontal direction) is less obvious, because of dipolar interactions between protons. Peaks that are partly overlapping in the correlation spectrum (right) are resolved in the TROSY spectrum (left).

given N–H group, most of the terms defined in equations (2.4) and (3.2), are fixed by the molecular characteristics, with the sole exception of the magnitude of the external magnetic field. It has been calculated that maximal interference cancellation effects can be expected for N–H groups at magnetic-field strengths of *ca.* 23.5 T (corresponding to a proton NMR frequency of *ca.* 1 GHz). Currently, the highest field on a commercially available spectrometer suitable for applications on proteins sits at *ca.* 18.7 T (proton frequency 800 MHz). The use of higher magnetic fields, therefore, not only improves the quality of NMR spectra by increasing the S/N ratio (see equation (2.1)), but should also contribute to the improvements offered by specific experiments that exploit the relaxation interference phenomena.

Similar consideration of dipolar/CSA relaxation interference can be applied to other combinations of atoms. For example, relaxation interference also applies to the relaxation of the ^{1}H nucleus in the NH group, though here the dominant relaxation contribution is not the dipolar interaction with the directly bonded ^{15}N nucleus, but through-space dipolar interaction with other ^{1}H nuclei (see § 2 *c*). If these other protons are diluted out, as, for example, in a perdeuterated sample, then the full benefit of relaxation interference can be obtained for this nucleus. The interference effects, though detectable, are not as large as for ^{15}N. For ^{13}C, similar relaxation interference effects are found only in aromatic ^{13}C–^{1}H pairs (Pervushin *et al.* 1998*a*). The relaxation of the attached ^{1}H, however, is much less affected because of an unfavourable orientation of the ^{1}H CSA tensor. For aromatic ^{13}C atoms, the theoretical maximal attenuation can be achieved at a field corresponding to a proton resonance frequency of *ca.* 600 MHz, well within the range of currently commercially available spectrometers.

The practical implementation into NMR experiments of methods to exploit interference effects, named transverse relaxation optimized spectroscopy (TROSY for

short) (Pervushin *et al.* 1997), is based on the selection of the NMR signal from only that half of molecules that relax with an attenuated relaxation rate, and the simultaneous suppression of the NMR signal from the other half of the molecules. Before the realization that the relaxation interference can yield benefits, NMR spectroscopists traditionally combined the signals from all the molecules by employing procedures that rapidly invert the heteronuclear spin orientation. This 'spin decoupling' removes the induced splitting and improves the apparent S/N ratio. It turns out that, for sufficiently large molecules, the benefit of exploiting the relaxation interference effect can outweigh the apparent loss of signal implied by TROSY selection of only that part of the total signal that gives rise to a narrow (i.e. slowly relaxing) resonance.

In typical applications of the TROSY method applied to N–H groups (see figure 9), a variety of recently proposed NMR pulse selection schemes to exploit relaxation interference for both ^{1}H and ^{15}N nuclei can be applied so that only the slowest relaxing part of the four-component multiplet is retained (Andersson *et al.* 1998; Pervushin *et al.* 1998*b*). Apart from producing simple two-dimensional correlation spectra, the TROSY selection elements can also be incorporated into more complex experiments, so that many important three- and four-dimensional heteronuclear experiments are likely to benefit from attenuated relaxation of N–H groups (Salzmann *et al.* 1998; Yang & Kay 1999). A further expansion of this concept is the application to perdeuterated proteins. By reducing the relaxation of amide protons by through-space dipolar interactions with other protons, the effects of the interference are enhanced, because, for the exchangeable protons left behind, the dipolar interaction with the directly bound nitrogen atom assumes a dominant role.

(*c*) *The other side of signal-to-noise*

As described in § 2 *a*, the sensitivity of an NMR experiment is expressed as the signal-to-noise (S/N) ratio. The inherent problem of the poor S/N of NMR spectroscopy has commonly been tackled by methods such as increasing the magnetic-field strength (with respect to equation (2.1)), maximizing the sample concentration or sample volume, or, more recently, via line-narrowing tricks, such as the use of deuteration and relaxation interference techniques, described in §§ 2 *d* and 3 *b*. An alternative way to tackle the problem 'from the other side', would be to attempt to reduce the noise component of the S/N ratio. Since the noise is generated in the components of the NMR spectrometer itself, NMR spectroscopists can usefully address this issue by appropriate adoption of new technologies in the fields of electronics and materials science. As we write this review, there are hopeful indications that substantial strides in this direction should be possible in the near future.

The dominant source of noise generated in an NMR spectrometer arises as the result of thermal electronic fluctuations in the signal detection circuitry of the instrument, principally the detection coil and signal amplifiers. For a number of years it has been considered that, conceptually at least, it should be possible to dramatically improve the S/N ratio for NMR by straightforwardly lowering the temperature of the entire part of the spectrometer that is responsible for the initial detection and amplification of the NMR signal. The technical problems associated with this simple-minded strategy have been severe, but have recently been overcome, in most part,

Figure 10. Photograph of a magnet giving the currently highest field (18 T, 800 MHz ^{1}H resonance frequency) suitable for magnetic resonance spectroscopy of proteins. Courtesy of Oxford Instruments Ltd.

by the design and introduction of amplifiers and receiver coils that are cooled to the temperature of liquid helium (4 K, or −269 °C). Yet more adventurous developments in this area aim at further improvements by utilization of components fabricated from high-temperature superconducting materials (Styles *et al.* 1984).

Recent demonstrations by the major NMR instrument manufacturers have shown that it is possible to increase the S/N ratio in the region of threefold to fourfold in normal protein applications, and there is an immediate prospect that such set-ups will become commercially available. A practical consequence of this technology means that it should become possible to measure high-quality NMR spectra of proteins at concentrations in the region of 50–100 μM (as opposed the current typical use of concentrations of *ca.* 0.5–2.0 mM. On the other hand, larger proteins, which tend to yield an intrinsically lower S/N ratio than smaller proteins (see § 2 *a*), could be investigated at more feasible protein concentrations.

In principal, the substantial increase in S/N ratio associated with cooled probe designs should pay handsome dividends in the systematic screening of potential pharmaceuticals in high-throughput screening strategies that use NMR. An important example of this is the structure–activity relationship-by-NMR (SAR-by-NMR) approach to drug discovery, championed by the NMR group at Abbot Laboratories (Shuker *et al.* 1996). Proteins that are enriched with ^{15}N are titrated with a series of simple chemical compounds from a 'fragment' library. The resonance lines of the protein can be followed selectively because the chemical compounds are not ^{15}N enriched. If one of the chemical fragments binds, this shifts the resonance line of one or more amide groups defining the interaction site. Discovery of two or more such protein–chemical fragment interactions can then lead very elegantly into a structure-directed chemical synthesis of a small number of tethered compounds that have a high potential to be tight-binding inhibitors of the protein function. Such studies are very cost intensive because of the need for large quantities of isotopically enriched proteins and substantial amounts of measuring time. In the recent discovery of high-affinity inhibitors to the metalloproteinase stromelysin, implicated in rheumatic diseases, protein samples were used in an SAR-by-NMR screen at concentrations of 0.3–0.5 mM (Olejniczak *et al.* 1997; Hajduk *et al.* 1997). Using a spectrometer equipped with cooled detection circuitry it was shown to be possible to reduce the concentration of the target protein of an SAR-by-NMR programme down to 50 μM. Compared to the stromelysin study, arguably only one-sixth to one-tenth of the protein would have been necessary for the successful completion of the screen. With the projected requirement for less and less protein per NMR sample, the high-throughput screening approach should become more economically accessible. On the other hand, a much better S/N ratio for the same sample concentration has the additional potential to allow the considerable shortening of the measurement time for many NMR experiments.

4. Orientation-dependent NMR of proteins

(*a*) *How to calculate a protein structure from NMR data*

The main source of information for the calculation of protein solution structures from NMR data comes in the form of internuclear-distance estimates derived from NOE spectra. The distances are extracted from monitoring the flow of magnetization when the excited state of a nucleus relaxes via dipolar interactions (see § 2 *c*), as defined in equation (2.4). It is generally accepted that an average of about ten experimental distance restraints per amino acid is sufficient for the determination of a low-resolution structural model. To obtain the high resolution that is possible, comparable perhaps to a 2.0 Å resolution X-ray crystal structure, more than 20 experimental distance restraints per residue are typically required.

In the earliest attempts at solution structure determination, the distance restraints were incorporated 'by hand' into computer-modelling procedures (Kaptein *et al.* 1985). Nowadays, based on the rapid improvement of computer performance, systematic calculations are performed to identify models of the target protein structure that represents the best agreement with experimental restraint data. Even with a large number of distance and other structural restraints that are available from NMR studies, this is generally insufficient to uniquely define the conformation of the protein chain. Instead, the strategy that is usually adopted is to sample a whole ensemble

of structural models, selecting out all those that agree with the experimental data within a given tolerance. In this way, the solution of the structure determination problem yields an ensemble of structures ('conformers'), each of which is slightly different in detail, but is in agreement with the experimental restraints. The results of this exercise are typically represented as a bundle of superposed conformers. In the parts of the structure that are well-defined by the experimental distance restraints, the conformers exhibit a close superposition and provide a 'tight' part of the bundle. In regions less well determined by the data, the superposition is worse and yields a 'loose' part of the bundle (see figure 6).

Because, in the mathematical sense, protein structure calculation using NMR data is an underdetermined problem, NMR practitioners are always striving to find ways to extract new structurally relevant parameters from the spectra. This problem stands in stark contrast to the situation that pertains for protein structure calculation based on X-ray diffraction data, where the number of experimental parameters is typically a few times *more* than that necessary to unambiguously define the coordinates of all the atoms in a protein. Indeed, a portion of the diffraction data is usually set aside and not used in the calculation at all. The resulting structure is then compared with the unused data to assess its quality in a procedure known as cross-validation. No bias is present in this measure—called the 'free R value'—because the calculated structure is not based on the data used for the cross-checking. Such cross-validation of NMR protein structures is not commonly performed, simply because the structures are very badly degraded by the exclusion of any sizeable proportion of the experimental restraints. Proper validation of protein solution structures remains a difficult issue for NMR spectroscopists. Nevertheless, very recent developments in the practice of biomolecular NMR give promise not only for improving the number and type of experimental restraint types that can be gathered, but may also provide means for simple cross-validation of the resulting structures.

(*b*) *Partial alignment of proteins*

As described above, the rapid tumbling of the protein molecules in solution has the important consequence of averaging both the splittings arising from internuclear dipolar couplings and the variation of chemical shifts with molecular orientation (CSA) in the magnetic field. However, it has recently been noted that small deviations from the ideal 'isotropic' rotational averaging make it possible to extract information of a type that would normally only be accessible in solid-state NMR spectroscopy; for example, the anisotropy of the chemical shift (Ottiger *et al.* 1997; Tjandra & Bax 1997*b*) and dipolar couplings (Tjandra *et al.* 1996). Apart from a small class of highly symmetric examples, all proteins exhibit asymmetry in terms of the overall molecular shape and the distribution of charged chemical groups and magnetic dipoles, e.g. unpaired electrons (Tolman *et al.* 1995), aromatic rings (Tjandra *et al.* 1997), and double bonds. The presence of these asymmetries leads to a trajectory of molecular tumbling in the magnetic field that is not completely random. Instead, the tumbling is, to a very slight extent, biased towards a preferential orientation with respect to the strong external magnetic field ('anisotropic'). A direct consequence of the anisotropy of rotational diffusion is that neither the internuclear dipolar couplings nor the CSA effects are completely averaged out. The incomplete rotational averaging, or 'partial molecular alignment', can yield telltale signs in the NMR spectrum that,

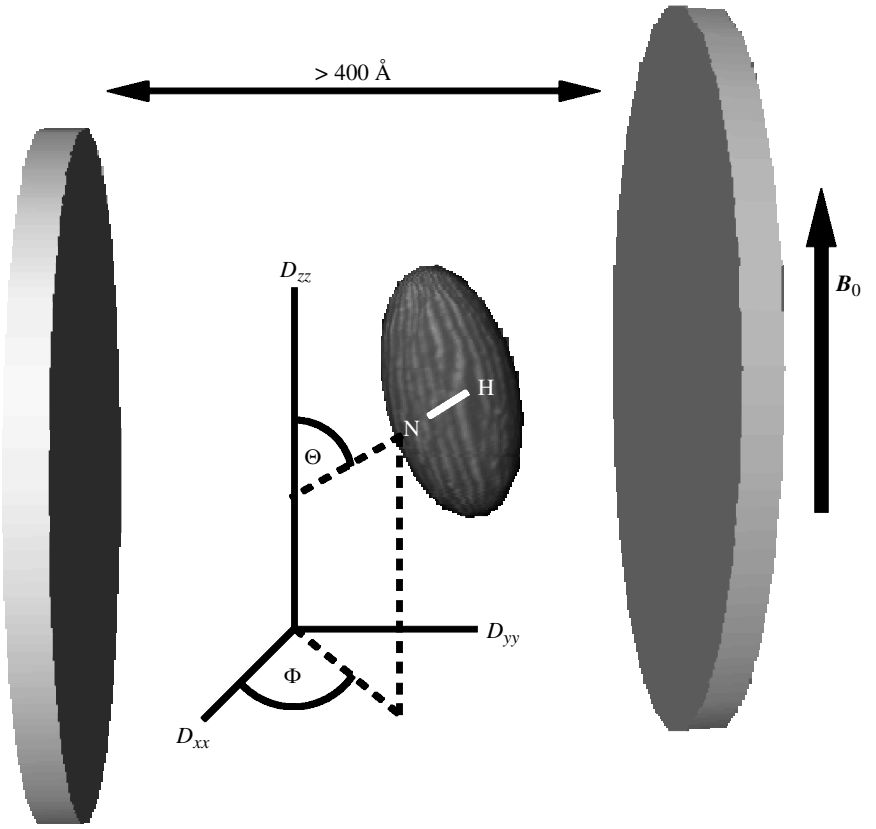

Figure 11. Schematic of the orientation of proteins by the presence of dilute liquid-crystal solutions made from bicelles. The alignment tensor is indicated together with the polar coordinate angles that connect it to the internuclear vectors in the molecular frame. Note that the parameters defining the orientation and magnitude of the alignment tensor are often very close, but not identical, to those of the rotational diffusion tensor, which, in essence, defines the molecular shape.

if appropriately targeted for measurement, provide new information to contribute to the description of the molecular structure. In practice, these effects lead to the addition of so-called residual dipolar couplings to the normal scalar couplings, and to the variation of chemical shifts with the magnetic-field strength. Both effects are normally very small (Tjandra *et al.* 1996) and observable only at the largest available magnetic-field strengths. Even then, apart from some specialized case where the molecular asymmetry properties are very strong (samples including double-stranded DNA (Tjandra *et al.* 1997), or delocalized unpaired electrons (Tolman *et al.* 1995)), the effects are typically too small for reliable measurement.

The promise of partial alignment of molecules for the generation of new types of structural information has received an enormous boost by the demonstration that it is possible to tune molecular alignment with appropriate conditioning of the NMR sample. In general, this takes the form of preparing the protein in a dilute liquid-crystal phase. Recently, it was shown that by addition to certain types of dilute lipid-mixture-based liquid crystals (so-called bilayer micelles or 'bicelles'), proteins

could be induced to partial alignment to a much greater extent than for the molecules on their own (Tjandra & Bax 1997*a*; Bax & Tjandra 1997). In this situation, specific interactions of the protein molecules with the bicelles are unwelcome: rather, the bicelles provide an anisotropic environment for the protein molecules to tumble in. Alternative methods of inducing partial molecular alignment are the addition of highly anisotropic, soluble and inert macromolecular assemblies, such as certain filamentous bacteriophage particles, tobacco mosaic virus, bacterial flagella or F-actin or purple membranes (Clore *et al.* 1998). The application of electrical fields—potentially an excellent method by which to partly align a protein on the basis of the asymmetric charge distribution of the molecules—is not suitable in aqueous media because the strength of the electric field required would lead to electrophoretic effects (Sears & Hahn 1966). It remains to be seen whether optical molecular alignment techniques will prove valuable.

(*c*) *Residual dipolar couplings*

As described above, the effect of the dipole interaction of one nucleus on another in a hypothetical, fixed (i.e. non-tumbling) protein is simply a small perturbation of the effective magnetic field, which leads to a splitting of the resonance line. The extent of the splitting (dipolar coupling) is a function of the distance between the two nuclei and the relative orientation of the vector connecting the two nuclei to the external magnetic field (or, more accurately, the alignment tensor **A**; see figure 11). For a given type of dipolar coupling, for example that between the nitrogen and hydrogen nuclei in a backbone N–H unit, the internuclear separation is fixed and essentially uniform throughout the protein chain. The overall orientation of the protein relative to the external magnetic field (the alignment tensor) applies identically to all atom pairs in the protein. This leaves the relative orientation of the vector connecting the two atoms relative to the protein coordinate frame as the only variable. As a result, in principle, the angles that this vector makes with the x-, y- and z-axes of the coordinate system can be extracted from the measurement of the dipolar coupling.

What makes measurements of the one-bond residual dipolar couplings ^1D dissimilar to the estimation of interproton distances from NOE spectra is the fact that the magnitudes reflect long-range structural order: residual dipolar couplings result from the alignment of the protein molecule as a whole. Furthermore, in principle, these measurements can be obtained for any pair of nuclei: H–N, H–C, N–C, C_α–C_β, etc. Thus, measurements of this type are not limited to pairs of ^{1}H spins, as in the case of short-range distance estimates. This makes residual dipolar couplings doubly suited for the structural study of larger proteins. As residual dipolar couplings are not based on short-range interactions, they allow the determination of, for example, the relative orientation of secondary structure features or even whole domains in a modular protein. Furthermore, since they do not entirely depend on protons for the measurement, residual dipolar couplings can be extracted from NMR experiments measured with highly deuterated protein samples.

The experimental measurement of residual dipolar couplings is straightforward and has been demonstrated already for a number of different spin pairs, e.g. H_N–N, H_α–C_α, N–CO and C_α–CO (Ottiger & Bax 1998). In a manner that is similar to that discussed in the context of the TROSY experiment (§ 3*b*), NMR spectra are recorded such that splittings due to scalar couplings are retained. The resonance line

splittings Δ observed in the absence of spin decoupling then represent the sum of the one-bond scalar and residual dipolar couplings, $\Delta = {}^1J + {}^1D$. To extract ${}^1D$ from Δ it is necessary to make two measurements, one in the presence and one in the absence of the partial alignment influence. The reasonable assumption that ${}^1J$ does not depend on the alignment, i.e. that the structure is not influenced by the aligning force, has to be invoked. If lipid bicelles are used, the alignment switch is achieved simply by a change of sample temperature, since the liquid-crystalline phase only forms above a certain critical temperature. In the case of filamentous bacteriophages, after measurement in the aligned state, the switch can be effected by sedimentation of the phage particles by centrifugation of the sample. It is evident that a change in temperature might violate the assumption that the structure of the protein—and, hence, the ${}^1J$—are identical in the aligned and the non-aligned state. An additional problem with bicelles is that they tend to interact unfavourably with some proteins, leading to absorption and denaturation of the sample. The chemically fairly inert phage particles are reported to be better suited as a general protein alignment tool.

The measured dipolar couplings can be introduced to structure calculations to supplement distance and other short-range conformational restraints. While distance restraints are readily handled in the available software programs used in the calculation of protein structures, the introduction of residual dipolar couplings is currently less straightforward. Application of experimental ${}^1D$ values to aid structure calculation has been shown to both improve the stereochemical quality of the resulting structure when tested against standard criteria, e.g. distribution of $\phi/$ pairs in a Ramachandran plot (Tjandra *et al.* 1997), and achieve an almost twofold improvement in the precision of the calculated structure (Bewley *et al.* 1998). Future projected development of residual dipolar coupling measurements suggest the prospect of the characterization of protein solution structures by triangulation, with a massively reduced (or even eliminated) requirement for the inclusion of interproton distance restraints. This would be particularly advantageous in applications to larger proteins, where the complexities of NOE spectra are rather daunting.

(*d*) *Chemical shift anisotropy*

The influence of CSA also becomes available upon partial molecular alignment, yielding parameters that have the potential to give additional improvement in solution structure determinations in a manner similar to that which arises with residual dipolar couplings. In this case, the spectroscopist monitors the change in apparent isotropic chemical shifts between the non-aligned and partly aligned conditions. The changes, which can be particularly large for the backbone carbonyl ^{13}C resonances, can be correlated with the magnitude and orientation of the alignment tensor, and, hence, indirectly with the molecular frame (Tjandra & Bax 1997*b*; Cornilescu *et al.* 1998; Ottiger *et al.* 1997). This is possible since the magnitude and the local orientation of the CSA tensor with respect to the carbonyl bond are well-known from solid-state NMR studies. Since in these experiments the effects of the CSA tensor are capable of reflecting long-range order, it will also be a useful restraint in the calculation of solution structures of proteins. Alternatively, CSA values might be left out of the structure calculation and then used in structure validation procedures by way of providing a ready source of independent data with which to calculate an objective 'quality factor' (Cornilescu *et al.* 1998), analogous to the free R value used in X-ray crystallography.

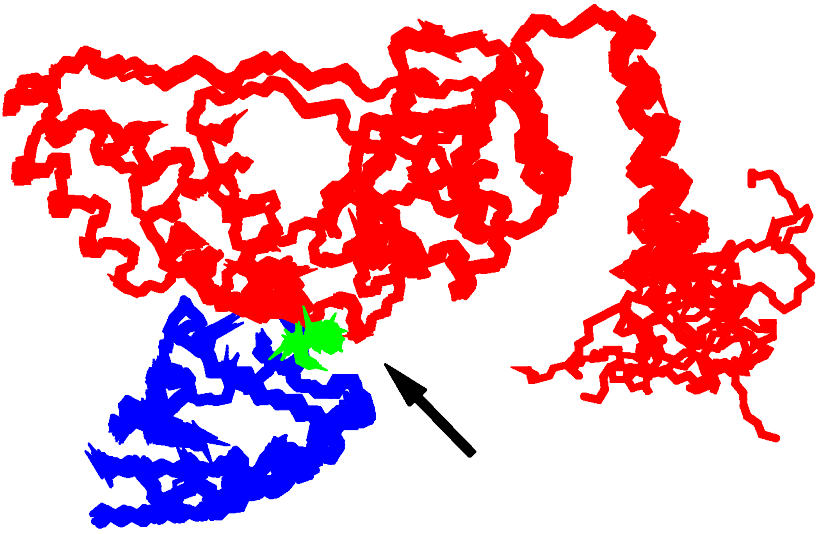

Figure 12. Solution NMR structure of the complex of the N-terminal domain of enzyme I of the *E. coli* phosphoenolpyrovate-sugar transporter (PDB entry 3ezb). An ensemble of 40 structures is displayed for both proteins. Enzyme I is shown in red, *HPr* is shown in blue. The crucial histidine 15 on *HPr* is shown in green and indicated by the arrow right in the interface between the two proteins.

5. Conclusions

(*a*) *Sensitivity and resolution*

The inherent low sensitivity of NMR spectroscopy has hampered the application of this otherwise powerful technique to proteins for decades. The considerable demands on sample concentration (where the aim has always been to maximize the signal strength) and molecular weight (to keep nuclear relaxation in check and thereby yield narrow lines) posed significant limits on the range of applications. The recent independent developments that tackle, at the same time, the problems of S/N ratio and resolution, have brought about significant advantages. With sharper resonance lines produced by the use of deuterated protein samples and clever exploitation of relaxation pathways in the TROSY experiment, enormous improvements are expected for applications of NMR to larger proteins. Already, since mid-1998, a number of NMR resonance assignments and structural studies have been reported in which the molecular weights are well above 35 kD, a limit hitherto believed to be insurmountable. The completed solution structures of the 44 kD ectodomain of SIV gp41, a symmetric trimer (Caffrey *et al.* 1998), and the structure of the 40 kD complex of *E. coli* phosphotransferase enzyme I and the small protein *HPr* from the phosphotransferase assembly (Garret *et al.* 1999) have already proved the combined power of the new methodologies.

The latter example is particularly interesting, because the small protein *HPr* has long been studied by NMR spectroscopy. It is, however, only with the investigation of the protein complex that the structural information from NMR spectroscopy really started to have a substantial impact on the understanding of the protein's function. The structure, as shown in figure 12, is remarkable for two reasons. Firstly, the degree

of definition is astounding for a complex of this size. It is hardly visible that an ensemble of 40 structures is displayed. With the exception of the C-terminal α-helix, the individual structures in the ensemble are almost identical, in marked contrast to the structure of HIV-1 *Nef*, which shows large regions that are ill-defined. The use of approximately 250 residual dipolar coupling constants (together with around 4000 distances and chemical shifts) has helped, in particular, to define the orientation of the two proteins with respect to each other. Secondly, the structure is of immediate biological significance to improve the understanding of phosphotransferase systems, since no other structure for such a complex has been solved to date. As indicated in figure 12, the phosphorylated histidine 15 of *HPr* is in the interface with enzyme I, and thus allows the identification of residues important in the transfer reaction. Mutants can be constructed and analysed based on this structure to understand the mechanism of the transferase in detail.

Other impressive examples of the application of deuteration and relaxation interference are the investigations of the 67 kD complex of a *trp* repressor protein tetramer bound to a non-palindromic DNA dodecamer, and of maltose binding protein at low temperatures, corresponding to a molecular weight of *ca.* 90 kD (Yang & Kay 1999; Shan *et al.* 1996). Even though solution structures have not yet been taken to completion in either of these cases, the complete resonance assignment of the spectrum in both cases is a promising result, particularly for the *trp* repressor–DNA complex. When bound to the non-palindromic DNA fragment, all four protein monomer subunits in the complex experience a slightly different environment. The spectra, therefore, retain much of the complexity of a monomeric species of the same size. In other words, the same protein has to be assigned four times.

(*b*) *Structure determination*

It is clear that the greater the number of independent structural restraints that can be obtained from NMR experiments, the better the precision and accuracy of the models of solution structure obtained from them (Clore *et al.* 1993). The quality of structures determined from NMR data is, therefore, expected to increase dramatically through the introduction of long-range information in the guise of CSA and residual dipolar coupling measurements. It is the long-range quality of these two properties that will make them much more than 'just another few restraints'. Not only will their use provide structure information where rather poor numbers of distance restraints can be collected (e.g. surface loop regions), but it is conceivable that residual dipolar couplings will help to close the debate on the evaluation of the quality of NMR solution structures. Certainly, the current practice of examination of how well parameters normally derived from X-ray structures are reproduced is a poor and (hopefully) temporary measure. Instead, long-range restraints could be the basis of a quality measure (Q-factor) as has recently been proposed. However, instead of choosing a special home-made parameter, the NMR community should be encouraged to adapt the 'free R-value', in a similar manner to that already successfully used in crystallography, as the only true measure of the accuracy and internal consistency of a structure. In their free R-value strategy, crystallographers typically set aside *ca.* 10% of the experimental diffraction data to provide an independent test-set with which to compare the derived structure. For solution structure determination based upon NMR measurements, leaving out *ca.* 10% of the internuclear distance data simply

proved impossible without introducing substantial distortions due to the enormous importance that certain 'key' distances can have upon the convergence of the structure calculation. Setting aside some or all the residual dipolar couplings, however, should not prove such a problem. Since these types of data contain information about the protein as a whole, as the reflections from an X-ray diffraction pattern, they are ideally suited to calculate free R-values or quality-factor values, which are likely to become prerequisites to publication and database submission for the NMR community in the future.

(*c*) *General*

The proteins to which NMR spectroscopy has been applied so successfully over the last decade have one feature in common: the majority have fewer than 150 residues and most of them are soluble and stable at low pH (below pH 7). The questions, in biology, that come into focus at the turn of the millennium now demand the analysis of less-soluble, less-stable and, most of all, higher molecular weight proteins and complexes with ligands such as other proteins, nucleic acids, carbohydrates, substrates, coenzymes and drugs. The recent developments in the field of NMR spectroscopy have not provided ultimate solutions to these aspects, but suggest promising starting points from which to tackle these challenges.

With the projected availability of ever-increasing magnetic-field strengths in combination with improved superconducting and super-cooled coils and amplifiers, we can hope that sample concentrations need not be as high as they are currently required to be. Reducing nuclear relaxation by deuteration, intelligent manipulation of interference between relaxation pathways, and other technical strategies will add to improved sensitivity. So far, many of these approaches have been demonstrated in separate fashion. A large boost in the effectiveness of protein NMR spectroscopy may be expected from future integration of some or all of these developments. For example, use of a perdeuterated protein sample to measure a TROSY experiment with a cryoprobe at 23.5 T (proton frequency 1 GHz) field strength should produce spectra of a quality and a molecular size never thought possible, even as little as a year ago. It is conceivable that resonance assignments of proteins and their complexes in the molecular weight range of up to 300 kD at sample concentrations well below 0.5 mM (the current lower limit) could, therefore, become routine. It is possible that the advancement of solution structure determinations will lag somewhat behind the ability to obtain assignments, but the scope will most certainly extend to molecules of 100 kD by the end of the next decade. It is imperative, here, to highlight that structure calculation is by no means the only useful application of protein NMR spectroscopy. However, once the resonance assignment has been accomplished, a whole wealth of NMR strategies is available to study dynamics and molecular interactions at atomic resolution. It is entirely up to NMR spectroscopists to decide if a scientific question warrants the determination of a high-quality, high-resolution three-dimensional structure or if the outline of the overall polypeptide fold is sufficient. In combination with theoretical sequence analysis and model building, low-resolution structural data from NMR spectroscopy could make a substantial contribution to the emerging field of 'structural genomics'. With the prospect of improvements in cooled NMR signal-detection circuitry, the use of high-temperature superconductors, ever-higher magnetic fields, and deuterated samples, some of the big hurdles involved

in these demanding projects, for example proteins of low solubility, could be overcome. The production-line approach inherent in a high-throughput project of this type—projected by some optimists to reach one structure determination per working day—obviously demands an experimental set-up that is ready to cope with the wide range of features that make larger proteins such a challenge.

The rapid pace of the developments of NMR protocols and technology at the turn of the last century holds the promise of a golden age of NMR in which a massively wider scope, both in the size and concentrations of the targeted molecules, and in the range of biological questions that can be posed is to be anticipated.

M.P. is a freshman Royal Society University Research Fellow. P.C.D. recently completed eight years as holder of a Royal Society University Research Fellowship. The authors acknowledge The Royal Society for its generous support.

References

Andersson, P., Annila, A. & Otting, G. 1998 *J. Magn. Reson.* **133**, 364–367.

Bax, A. & Tjandra, N. 1997 *J. Biomol. NMR* **10**, 289–292.

Bewley, C. A., Gustafson, K. R., Boyd, M. R., Covell, D. G., Bax, A., Clore, G. M. & Gronenborn, A. M. 1998 *Nature: Struct. Biol.* **5**, 571–578.

Caffrey, M., Cai, M., Kaufman, J., Stahl, S. J., Wingfield, P. T., Gronenborn, A. M. & Clore, G. M. 1997 *J. Mol. Biol.* **271**, 819–826.

Caffrey, M., Cai, M., Kaufman, J., Stahl, S. J., Wingfield, P. T., Covell, D. G., Gronenborn, A. M. & Clore, G. M. 1998 *EMBO J.* **16**, 4572–4584.

Campbell, I. D. & Baron, M. 1991 *Phil. Trans. R. Soc. Lond.* B **332**, 165–170.

Clore, G. M., Robien, M. A. & Gronenborn, A. M. 1993 *J. Mol. Biol.* **231**, 82–102.

Clore, G. M., Starich, M. R. & Gronenborn, A. M. 1998 *J. Am. Chem. Soc.* **120**, 10 571–10 572.

Cornilescu, G., Marquardt, J. L., Ottiger, M. & Bax, A. 1998 *J. Am. Chem. Soc.* **120**, 6836–6837.

Gardner, K. H. & Kay, L. E. 1997 *J. Am. Chem. Soc.* **119**, 7599–7600.

Gardner, K. H., Rosen, M. K. & Kay, L. E. 1997 *Biochem.* **36**, 1389–1401.

Gardner, K. H., Zhang, X., Gehring, K. & Kay, L. E. 1998 *J. Am. Chem. Soc.* **120**, 11 738–11 748.

Garret, D. S., Seok, Y.-J., Peterkofsky, A., Gronenborn, A. M. & Clore, G. M. 1999 *Nature: Struct. Biol.* **6**, 166–173.

Grzesiek, S., Wingfield, P., Stahl, S., Kaufman, J. D. & Bax, A. 1995 *J. Am. Chem. Soc.* **117**, 9594–9595.

Grzesiek, S., Bax, A., Clore, G. M., Gronenborn, A., Hu, J.-S., Kaufman, J., Palmer, I., Stahl, S. J. & Wingfield, P. 1996*a* *Nature: Struct. Biol.* **3**, 340–345.

Grzesiek, S., Stahl, S. J., Wingfield, P. & Bax, A. 1996*b* *Biochem.* **35**, 10 256–10 262.

Grzesiek, S., Bax, A., Hu, J.-S., Kaufman, J., Palmer, I., Stahl, S. J., Tjandra, N. & Wingfield, P. 1997 *Protein Sci.* **6**, 1248–1261.

Hajduk, P. J. (and 18 others) 1997 *J. Am. Chem. Soc.* **119**, 5818–5827.

Kaptein, R., Zuiderweg, E. R. P., Schenk, R. M., Boelens, R. & van Gunsteren, W. F. 1985 *J. Mol. Biol.* **182**, 179–182.

LeMaster, D. M. & Richards, F. M. 1988 *Biochem.* **27**, 142–150.

Mal, T. K., Matthews, S. J., Kovacs, H., Campbell, I. D. & Boyd, J. 1998 *J. Biomol. NMR* **12**, 259–276.

Markus, M. A., Davie, K. T., Matsudeira, P. & Wagner, G. 1994 *J. Magnet. Reson.* **105**, 192.

Olejniczak, E. T., Hajduk, P. J., Marcotte, P. A., Nettesheim, D. G., Meadows, R. P., Edalji, R., Holzman, T. F. & Fesik, S. W. 1997 *J. Am. Chem. Soc.* **119**, 5828–5832.

Ottiger, M. & Bax, A. 1998 *J. Am. Chem. Soc.* **120**, 12 334–12 341.

Ottiger, M., Tjandra, N. & Bax, A. 1997 *J. Am. Chem. Soc.* **119**, 9825–9830.
Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. 1997 *Proc. Natn. Acad. Sci. USA* **94**, 12 366–12 371.
Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. 1998*a* *J. Am. Chem. Soc.* **120**, 6394–6400.
Pervushin, K., Wider, G. & Wüthrich, K. 1998*b* *J. Biomol. NMR* **12**, 345–348.
Rosen, M. K., Gardner, K. H., Willis, R. C., Parris, W. E., Pawson, T. & Kay, L. E. 1996 *J. Mol. Biol.* **263**, 627–636.
Salzmann, M., Pervushin, K., Wider, G., Senn, H. & Wüthrich, K. 1998 *Proc. Natn. Acad. Sci. USA* **95**, 13 585–13 590.
Sears, R. E. J. & Hahn, E. L. 1966 *J. Chem. Phys.* **45**, 2753–2769.
Shan, X., Gardner, K. H., Murhandiram, D. R., Rao, N. S., Arrowsmith, C. H. & Kay, L. E. 1996 *J. Am. Chem. Soc.* **118**, 6570–6579.
Shuker, S. B., Hajduk, P. J., Meadows, R. P. & Fesik, S. W. 1996 *Science* **274**, 1531–1534.
Sicherl, F., Moarefi, I. & Kurlyan, J. 1997 *Nature* **385**, 602–609.
Styles, P., Soffe, N. F., Scott, C. A., Cragg, D. A., Row, F., White, D. J. & White, P. C. J. 1984 *J. Magnet. Reson.* **60**, 397–402.
Tjandra, N. & Bax, A. 1997*a* *Science* **278**, 1111–1114.
Tjandra, N. & Bax, A. 1997*b* *J. Am. Chem. Soc.* **119**, 9576–9577.
Tjandra, N., Grzesiek, S. & Bax, A. 1996 *J. Am. Chem. Soc.* **118**, 6264–6272.
Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M. & Bax, A. 1997 *Nature: Struct. Biol.* **4**, 732–738.
Tolman, J. R., Flanagan, J. M., Kennedy, M. A. & Prestegard, J. H. 1995 *Proc. Natn. Acad. Sci. USA* **92**, 9279–9282.
Torchia, D. A., Sparks, S. W. & Bax, A. 1988 *J. Am. Chem. Soc.* **110**, 2320–2321.
Venters, R. A., Metzler, W. J., Spicer, L. D., Mueller, L. & Farmer, B. T. 1995 *J. Am. Chem. Soc.* **117**, 9592–9593.
Venters, R. A., Farmer, B. T., Fierke, C. A. & Spicer, L. D. 1996 *J. Mol. Biol.* **264**, 1101–1116.
Xu, W., Harrison, S. A. & Eck, M. J. 1997 *Nature* **385**, 595–601.
Yang, D. & Kay, L. E. 1999 *J. Am. Chem. Soc.* **121**, 2571–2575.
Zwahlen, C., Gardner, K. H., Sarma, S. P., Horita, D. A., Byrd, R. A. & Kay, L. E. 1998*a* *J. Am. Chem. Soc.* **120**, 7617–7625.
Zwahlen, C., Vincent, S. J. F., Gardner, K. H. & Kay, L. E. 1998*b* *J. Am. Chem. Soc.* **120**, 4825–4831.

AUTHOR PROFILES

M. Pfuhl

Mark Pfuhl (left) is proud to say 'Ich bin ein Berliner', in which place he was born 33 years ago. He spent his whole school and university career in Berlin, where he first studied chemistry at the Technische Universität from 1985 to 1987 and then biochemistry at the Freie Unversität from 1987 to 1991. After receiving his diploma with 'sehr gut', he obtained a predoctoral fellowship at the EMBL Heidelberg. Here he worked from 1991 to 1995 in the group of Annalisa Pastore. After a short spell in Rouen he came to UCL on an EMBO long-term postdoctoral fellowship in 1996. Since 1998 he has been a Royal Society Research Fellow at the Department of Biochemistry. In his spare time he enjoys rowing on the River Thames.

P. C. Driscoll

Strictly an East End Cockney by birth (within the sound of Bow Bells), Paul Driscoll (right) was raised and educated in the southern suburbs of London. First class honours in chemistry at the University of Oxford (1985), combined with vacation employment at the Medicinal Chemistry Department at the Wellcome Foundation, led to the development of an interest in the application of NMR spectroscopy to biomolecules, large and small. After completing a DPhil in the Inorganic Chemistry Laboratory at Oxford, studying copper proteins (1997), Paul was a Fogarty Visiting Fellow at the National Institutes of Health, USA (1988–1989), where he was involved in the early applications of multidimensional heteronuclear NMR in the investigation of the structure and dynamic properties of proteins. Paul was a Royal Society University Research Fellow (1990–1998) in the Biochemistry Department first at the University of Oxford and (since 1994) at University College London, where he is now a Senior Lecturer. Paul's current research interests attempt to combine NMR analysis of protein structure with functional studies, particularly in the field of intracellular signal transduction.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES