Quarterly Reviews of Biophysics **33**, 1 (2000), pp. 29–65 Printed in the United Kingdom

© 2000 Cambridge University Press

NMR spectroscopy: a multifaceted approach to macromolecular structure

Ann E. Ferentz and Gerhard Wagner

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Cambridge, MA USA

1. Introduction 29

2. Landmarks in NMR of macromolecules 32
2.1 Protein structures and methods development 32
2.1.1 Sequential assignment method 32
2.1.2 Triple-resonance experiments 34
2.1.3 Structures of large proteins 36
2.2 Protein–nucleic acid complexes 37
2.3 RNA structures 38
2.4 Membrane-bound systems 39

3. NMR spectroscopy today 40
3.1 State-of-the-art structure determination 41
3.2 New methods 44
3.2.1 Residual dipolar couplings 44
3.2.2 Direct detection of hydrogen bonds 44
3.2.3 Spin labeling 45
3.2.4 Segmental labeling 46
3.3 Protein complexes 47
3.4 Mobility studies 50
3.5 Determination of time-dependent structures 52
3.6 Drug discovery 53

4. The future of NMR 54
4.1 The ease of structure determination 54
4.2 The ease of making recombinant protein 55
4.3 Post-translationally modified proteins 55
4.4 Approaches to large and/or membrane-bound proteins 56
4.5 NMR in structural genomics 56
4.6 Synergy of NMR and crystallography in protein structure determination 56

5. Conclusion 57

6. Acknowledgements 57

7. References 57

1. Introduction

Since the publication of the first complete solution structure of a protein in 1985 (Williamson *et al*. 1985), tremendous technological advances have brought nuclear magnetic resonance

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

Table 1. *Numbers of structures in the Protein Data Bank as of January 2000*

Type of macromolecule	NMR	X-ray diffraction & other
Proteins, peptides, viruses	1471	8488
Protein–nucleic acid complexes	62	411
Nucleic acids	286	473
Carbohydrates	4	14
Total	1823	9386

Adapted from the RCSB Protein Data Bank web site.

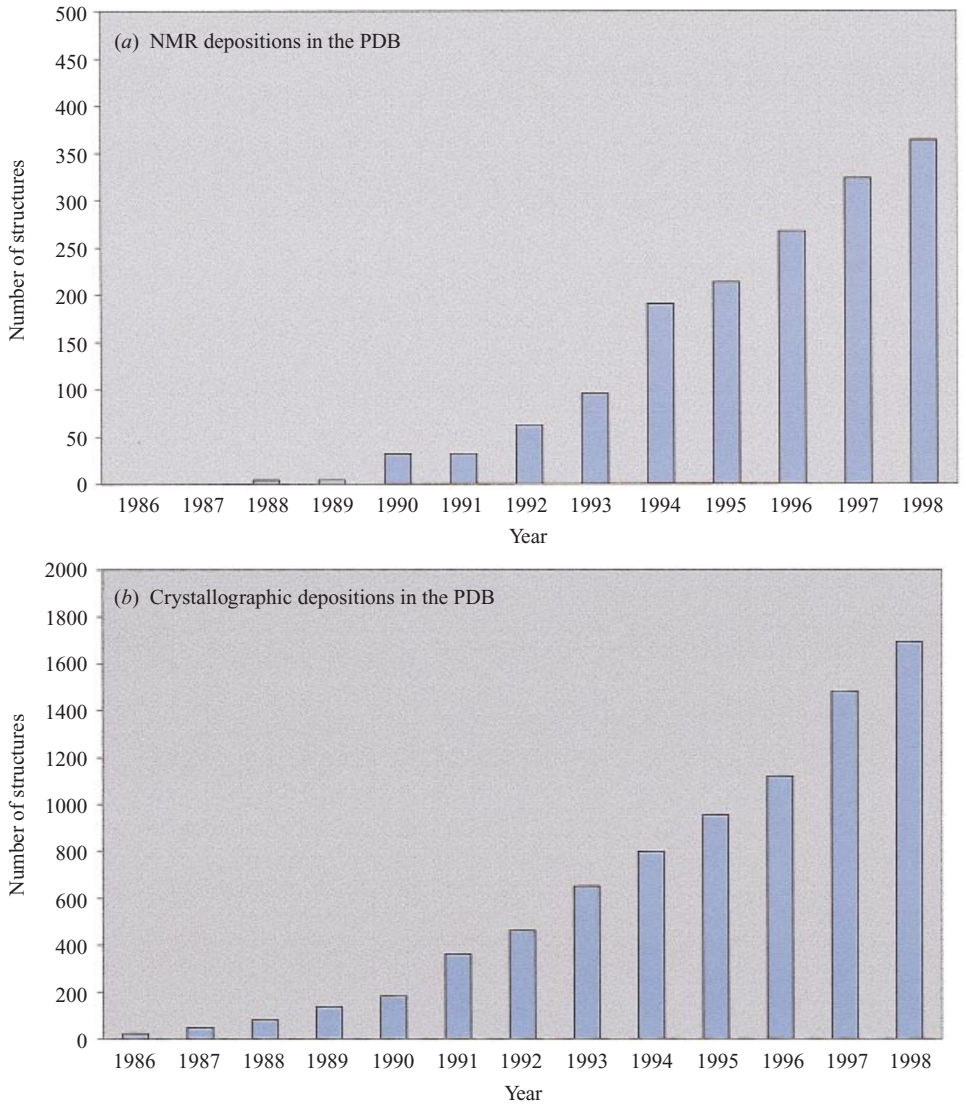


Fig. 1. Numbers of entries in the Protein Data Bank as of December 1999. (*a*) Number of depositions of NMR structures each year. (*b*) Number of X-ray crystal structures deposited each year.

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

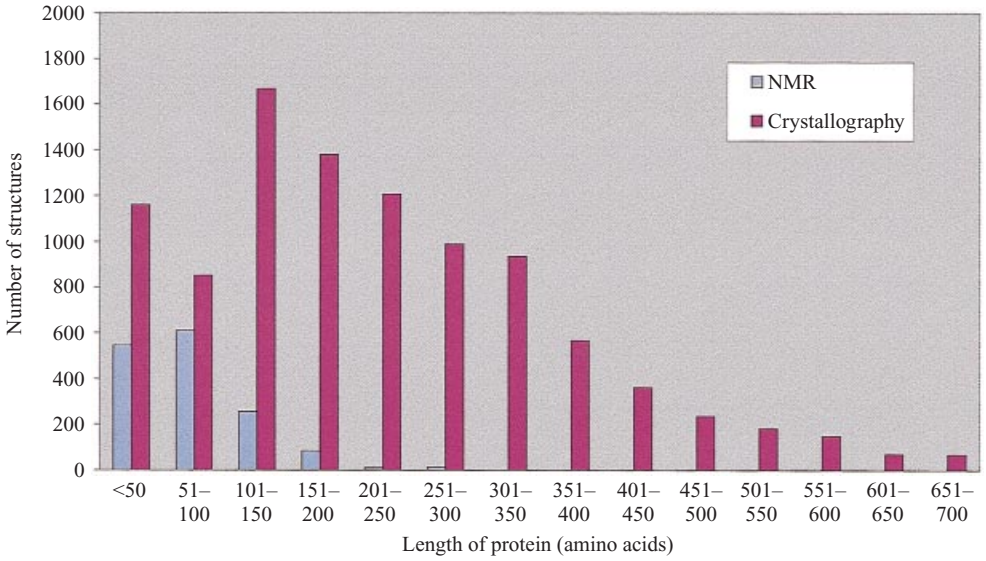


Fig. 2. Size distribution of structures in the Protein Data Bank. Size ranges represent the total number of amino acids for which coordinates are available.

spectroscopy to the forefront of structural biology. Innovations in magnet design, electronics, pulse sequences, data analysis, and computational methods have combined to make NMR an extremely powerful technique for studying biological macromolecules at atomic resolution (Clore & Gronenborn, 1998). Most recently, new labeling and pulse techniques have been developed that push the fundamental line-width limit for resolution in NMR spectroscopy, making it possible to obtain high-field spectra with better resolution than ever before (Dötsch & Wagner, 1998). These methods are facilitating the study of systems of ever-increasing complexity and molecular weight.

While ten years ago NMR structures accounted for only a few percent of depositions in the Protein Data Bank, they currently represent more than 16% of newly deposited structures. As of January 2000, there were 1823 NMR structures in the PDB, including proteins, nucleic acids, protein–nucleic acid complexes, and carbohydrates (Table 1). This is considerably fewer than the 9386 structures determined by crystallography or other techniques, but is an impressive number when one considers that the first NMR structure was not determined until a full 25 years after the first protein crystal structures (Kendrew *et al.* 1960; Perutz, 1960). Although there are still relatively few NMR spectroscopists compared to X-ray crystallographers, the number of new NMR structures has been increasing steadily over the past decade (Fig. 1(*a*)). The abundance of NMR structures of nucleic acids is particularly striking, with NMR contributing 38% of the structures in the PDB. Nevertheless, structures of proteins are by far the most prevalent, and protein NMR spectroscopy will comprise the bulk of the present discussion.

Among the coordinates in the PDB, NMR structures account for a disproportionate number of the small structures, making up 37% of structures smaller than $\sim$ 10 kDa, but only 16% of the total depositions (Fig. 2). This reflects the limitation of early solution studies to proteins having molecular weights of 10 kDa or less. Historically, great efforts have been required to determine structures larger than $\sim$ 25 kDa, but current technologies have led to the completion of structures as large as 40 kDa. It is clear that solution structures of proteins

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

and complexes in the 60 kDa range will soon be feasible (Wagner, 1997). Systems larger than this account for just 5% of crystal structures in the Protein Data Bank (Fig. 2), so once NMR can handle systems of this size, it will be able to solve all but the very largest structures typically deposited in the PDB.

To dwell solely on the number of extant NMR structures would be to overlook some of the unique contributions of NMR to structural biology. NMR alone can be used to characterize the dynamics of macromolecules in solution and to observe unfolded or partially folded proteins. In addition, weak interactions between macromolecules and ligands can be defined at atomic resolution using NMR spectroscopy, a capability that has led to a powerful method for drug discovery (Shuker *et al*. 1996). This article will review the key elements of NMR spectroscopy that have made it such a major technique for biological studies, with an emphasis on recent developments.

2. Landmarks in NMR of macromolecules

In 1957, 11 years after NMR was originally identified as an experimental technique (Bloch, 1946; Purcell *et al*. 1946), the first NMR spectrum of a protein was recorded (Saunders *et al*. 1957) and spectra of amino acids were analyzed (Takeda & Jardetzky, 1957). By that time, chemical shift had been discovered (Arnold *et al*. 1951; Knight, 1949) and coupling had been observed (Gutowsky *et al*. 1951; Hahn & Maxwell, 1951), but it was another 25 years before the technology had advanced to the point where the spin systems of an entire protein could be assigned (Wagner & Wüthrich, 1982). Complete assignment of a protein required fundamental methodological advances and substantial increases in resolution and sensitivity over the original 40 MHz spectrum of ribonuclease recorded by Saunders. Key advances included the development of Fourier transform spectroscopy, two-dimensional spectroscopy, superconducting magnets, and new computational methods. The additional insight that nuclear Overhauser effects at short mixing times could be used to obtain interproton distances within a protein was what made the determination of a complete solution structure of a protein by NMR seem feasible (Gordon & Wüthrich, 1978; Wagner & Wüthrich, 1979). However, the actual completion of a structure had to await the development of distance geometry methods that could use the distance data to derive the fold of a protein (Havel & Wüthrich, 1984). Although NMR spectroscopy had been applied to some biological studies of DNA hydration and metal chelation while still in its infancy, it began to make major contributions to biology only when it developed the capacity to solve macromolecular structures.

2.1 Protein structures and methods development

2.1.1 Sequential assignment method

The structure of the 57-residue bull seminal trypsin inhibitor reported in 1985 was the first complete solution structure of a protein (Williamson *et al*. 1985). Determination of this structure relied upon the sequential assignment method (Dubs *et al*. 1979; Wagner & Wüthrich, 1982) and distance geometry calculations (Havel & Wüthrich, 1984). Assignments were made by first identifying spins belonging to a particular amino acid type using two-dimensional through-bond correlations such as COSY, RELAY-COSY, and DOUBLE-

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

RELAY-COSY experiments (Wüthrich, 1986). Spins were then mapped onto the protein sequence by analyzing NOESY spectra to identify adjacent residues in the primary sequence. The secondary structure was established on the basis of characteristic short-range NOEs in the NOESY spectrum. Long-range NOEs and coupling constants from COSY data were then input into distance geometry calculations to determine the three-dimensional structure. Although distance geometry methods had been developed prior to this time (Crippen & Havel, 1978; Havel *et al*. 1983) they had only succeeded in determining the structures of small peptides and could not handle proteins containing hundreds of atoms. To progress from the structure of the micelle-bound peptide glucagon, which was the first peptide structure solved (Braun *et al*. 1981, 1983), to the determination of a full-sized protein solution structure required that software be developed to handle extensive data sets and fold proteins reliably. Thus it was a number of years between the time when sufficient data were available to determine a full-sized protein solution structure and when the first structure was solved. This was first achieved with a distance geometry method (Havel & Wüthrich, 1984) followed by a method for searching dihedral angle space (Braun & Go, 1985) and a molecular dynamics approach (Kaptein *et al*. 1985).

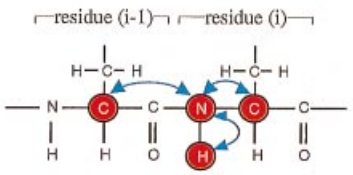
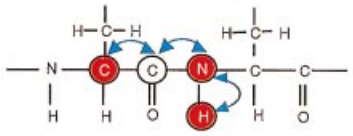
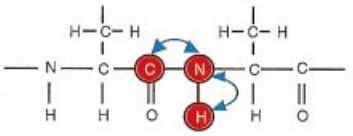
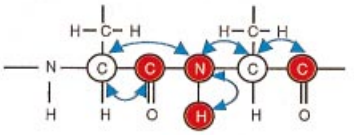
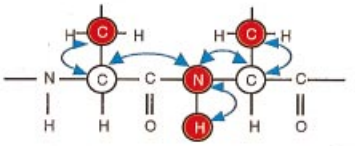
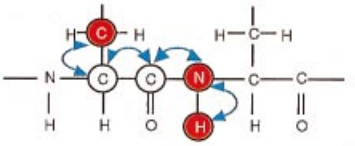
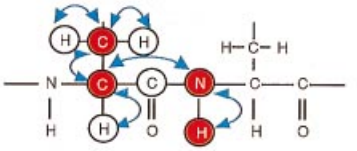
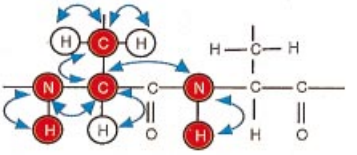
The structure of bull seminal trypsin inhibitor was initially met with some skepticism, though, because the crystal structures of two homologous proteins, the porcine secretory trypsin inhibitor (Bolognesi *et al*. 1982) and ovomucoid third domain (Papamokos *et al*. 1982) had been reported previously. NMR only became fully accepted as an independent method for complete structure determination of proteins upon publication of the NMR structure of tendamistat (74 aa) in 1986 (Kline *et al*. 1986), for which no previous structural information was available. The crystal structure was solved contemporaneously and the two structures were found to be virtually identical (Billeter *et al*. 1989; Braun *et al*. 1989; Kline *et al*. 1988; Pflugrath *et al*. 1986). This served the dual role of establishing NMR as a structural method and validating the use of X-ray diffraction to draw conclusions about the structures of proteins under biological conditions.

The first protein structures solved by NMR were necessarily small, containing fewer than 80 amino acids. Size was a severely limiting factor until the development of routine methods for isotopic labeling (LeMaster & Richards, 1985) and heteronuclear NMR techniques in the late 1980s. One of the early uses of isotopic labeling was to sort out NOESY peaks in the 17·8 kDa *Antennapedia* homeodomain–DNA complex by recording ^{15}N-filtered two-dimensional NOESY spectra (Otting *et al*. 1986, 1990). Since the protein was ^{15}N-labeled, the filtering distinguished protein NOEs (to ^{15}N-attached protons) from NOEs within the DNA. Additional use of a three-dimensional ^{15}N-edited NOESY spectrum enabled the protein to be assigned completely.

Another early structure solved using ^{15}N- and ^{13}C-edited NOESY spectra was that of interleukin 1β (Clore *et al*. 1991). This 17·4 kDa, 153-residue protein was 50 residues longer than any protein whose structure had been solved previously by NMR. During the course of assigning interleukin 1β, a three-dimensional ^{15}N-TOCSY-HMQC experiment (total correlation spectroscopy) was developed (Marion *et al*. 1989b). This experiment greatly facilitates assignment by correlating all the protons in a residue with the corresponding backbone amide group in the protein ($^{1}H^{N}$ and ^{15}N resonances). This allows immediate identification of the spin systems with much less overlap than in the two-dimensional TOCSY spectrum. Spin systems can then be linked together by the ^{15}N-NOESY-HSQC. This assignment strategy can be used for any protein small enough to give good TOCSY spectra.

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

Table 2. *NMR experiments currently used for backbone assignment*

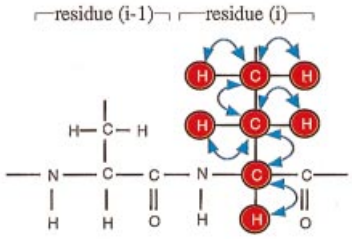
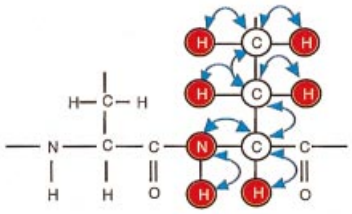
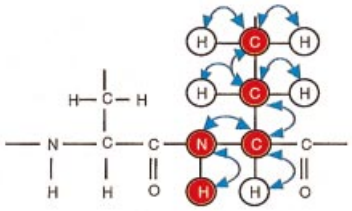
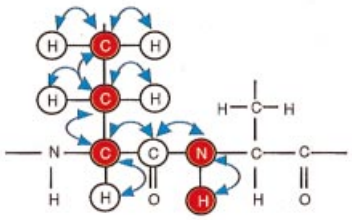
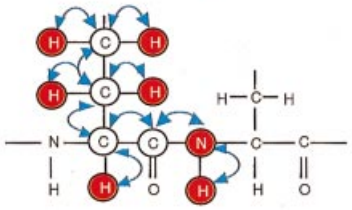
Experiment	Magnetization transfer	References
HNCA		Ikura *et al*. 1990a Ikura *et al*. 1990b Grzesiek & Bax, 1992b Yamazaki *et al*. 1994b
HN(CO)CA		Ikura *et al*. 1990a Ikura *et al*. 1990b Grzesiek & Bax, 1992b Yamazaki *et al*. 1994a
HNCO		Grzesiek & Bax, 1992b Muhandiram & Kay, 1994 Yamazaki *et al*. 1994a
HN(CA)CO		Clubb *et al*. 1992 Kay *et al*. 1994 Yamazaki *et al*. 1994a Matsuo *et al*. 1996
HN(CA)CB		Yamazaki *et al*. 1994a Muhandiram & Kay, 1994 Wittekind & Mueller, 1993
HN(COCA)CB		Grzesiek & Bax, 1992b Yamazaki *et al*. 1994a
CBCA(CO)NH		Grzesiek & Bax, 1992b Muhandiram & Kay, 1994
CBCANH		Grzesiek & Bax, 1992a

2.1.2 Triple-resonance experiments

The full benefits of isotope labeling were not recognized, though, until the development of triple-resonance techniques in the early 1990s (Ikura *et al*. 1990a; Kay *et al*. 1990; Montelione & Wagner, 1990). Experiments that allowed through-bond correlation of spins along the

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

Table 3. *NMR experiments currently used for side chain assignment*

Experiment	Magnetization transfer	References
HCCH-TOCSY		Kay *et al*. 1993
H(CC)NH-TOCSY		Lin & Wagner, 1999
(H)C(C)NH-TOCSY		Lin & Wagner, 1999
(H)C(C-CO)NH-TOCSY		Lin & Wagner, 1999 Grzesiek *et al*. 1993 Logan *et al*. 1993
H(CC-CO)NH-TOCSY		Lin & Wagner, 1999 Logan *et al*. 1993

backbone revolutionized the assignment process and paved the way for rapid assignment of larger proteins. These experiments use one-bond and two-bond couplings to correlate the backbone H^N, N, Cα, and C′ spins in ^{15}N,^{13}C-labeled proteins, and can provide continuous, unambiguous assignments of the entire backbone in proteins larger than 20 kDa. The backbone assignment is independent of any prior knowledge of spin systems, so side chain assignments are made during a later stage of the assignment process. Pairs of data sets are examined that give information about intra-residue correlations and the corresponding inter-

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

residue correlations. For example, the HNCA and HN(CO)CA (Grzesiek & Bax, 1992b; Ikura *et al*. 1990a; Kay *et al*. 1990; Yamazaki *et al*. 1994a) correlate H^N, N, and Cα spins along the entire backbone: the HNCA correlates each amide H^N and N with the intra-residue Cα, while the HN(CO)CA correlates this Cα with the H^N and N of the next residue in the protein (see Table 2). Ambiguous connectivities due to degeneracies in Cα chemical shifts can be resolved using other sets of experiments that correlate Cβ or C′ resonances with the backbone amides. Segments of connected spin systems are then mapped onto the sequence of the protein on the basis of carbon chemical shifts, which are characteristic of specific amino acids (Wishart & Sykes, 1994). The experiments most commonly used in current assignment strategies are listed in Table 2.

Side chain assignments are made on the basis of HCCH-TOCSY (Kay *et al*. 1993) and HC(C-CO)NH TOCSY-type experiments (Table 3) (Grzesiek *et al*. 1993; Lin & Wagner, 1999). The latter experiments correlate carbon and proton spins in the side chains directly with the backbone amide groups and have provided complete assignments for a protein of 20 kDa (Lin & Wagner, 1999). If backbone and Cβ chemical shifts alone were not sufficient to identify some spin systems with specific amino acids in the protein, these experiments will surely provide enough information. With assignments in hand, ^{15}N- and ^{13}C-NOESY-HSQC spectra can be assigned and the structure determined from the resulting distance constraints. Contemporary strategies for assignment and structure determination are detailed in Section 3.1 below.

2.1.3 Structures of large proteins

Three further technical developments, deuterium labeling, gradients, and the TROSY (transverse relaxation optimized spectroscopy) method, have led to refinement of these triple resonance experiments, making them suitable for studies of larger systems than ever before. It had long been realized that substituting deuterons for protons would lengthen the relaxation times of the attached nuclei. Thus proteins deuterated at the α and β positions will give substantially greater signal in triple-resonance experiments such as the HNCA and HNCACB than the corresponding protonated proteins. This difference becomes vital for systems larger than ∼ 25 kDa. Several structures of proteins over 30 kDa have been solved using deuterated samples (Caffrey *et al*. 1998; Clore & Gronenborn, 1998; Garrett *et al*. 1997, 1999; Jasanoff *et al*. 1998; Matsuo *et al*. 1997), while many more proteins of this size have been assigned. At 64 kDa, the Trp repressor/DNA complex is one of the largest systems assigned to date (Shan *et al*. 1996).

Developing methods for studying proteins of higher molecular weights continues to be an area of much activity, and the TROSY method promises to allow studies of systems larger than 100 kDa (Pervushin *et al*. 1997). The TROSY effect is particularly pronounced for ^{15}N–^{1}H spin pairs. It utilizes the correlated motion of the dipole–dipole and the chemical shift anisotropy (CSA) Hamiltonians. The dipole field at the site of one spin has two signs depending on the orientation of the neighboring spin that creates the dipolar field. The CSA is independent of the orientation of the neighboring spin and thus augments the dipole field for one doublet component and diminishes it for the other. As a consequence, one doublet component is broadened while the other is sharpened. Fortunately, the N–H dipole vector and the largest component of the nearly axially symmetric CSA tensor are almost parallel with an angle of *c*. 15° between them. Thus, the two Hamiltonians could cancel almost

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

perfectly if the largest component of the CSA tensor had the same magnitude as the dipole–dipole vector. The CSA tensor is field-dependent, and the line-narrowing effect for the N–H spin pair has been estimated to be optimal between 950 and 1050 MHz (Wider & Wüthrich, 1999). For this reason, the arrival of 900+ NMR spectrometers is eagerly awaited by the NMR community. To avoid relaxation from other protons, it is crucial that proteins be perdeuterated except for the NH groups. TROSY modules have been implemented in all the important triple resonance experiments currently used for sequential assignments (Loria *et al*. 1999; Salzmann *et al*. 1998; Salzmann *et al*. 1999a; Salzmann *et al*. 1999b). The TROSY effect can also be applied to aromatic spin systems since aromatic carbons have large CSA tensors. In this case the optimal effect is in the range of 500 to 800 MHz (Meissner & Sørensen, 1999; Pervushin *et al*. 1998b). For applications to very large proteins (those over 200 kDa), a further technique has been proposed termed CRINEPT (Riek *et al*. 1999) that uses cross-correlated relaxation-induced polarization transfer (CRIPT) with the traditional INEPT transfer.

Optimized triple-resonance experiments have vastly simplified the assignment process, and can give complete proton and carbon assignments for many sizeable proteins. Having so many assignments simplifies the task of assigning NOEs and obtaining distance constraints for structure determination using ^{15}N- and ^{13}C-NOESY-HSQC spectra. With present day advances in instrumentation and methodology, NMR is on the brink of being able to solve larger, more complex structures than ever before, as will be discussed further below.

2.2 Protein–nucleic acid complexes

Early studies of protein–nucleic acid complexes by NMR involved the determination of separate structures for the DNA-binding domain of a protein and the corresponding DNA, and the identification of just enough intermolecular NOEs to dock the protein onto the DNA in a unique orientation. The techniques for doing this were first established on the *Antennapedia* homeodomain–DNA complex, where the ability to ^{15}N label the protein and to perform ^{15}N-filtered NOESY experiments were critical to assigning NOEs between the protein and the DNA (Otting *et al*. 1990). The complete structure of this complex was reported a few years later (Billeter *et al*. 1993), at roughly the same time as structures of several other protein–DNA complexes (Chuprina *et al*. 1993; Omichinski *et al*. 1993; South & Summers, 1993). These structures gave insight into the specificity of DNA recognition by helix-turn-helix and zinc finger proteins. A wide variety of contacts were observed, including base-specific contacts, contacts to the phosphate backbone, and hydrogen bonds, which may be bridged by water molecules. Some of these contacts are dynamic rather than static, an attribute that could only be detected by NMR (Chuprina *et al*. 1993). These structures established NMR as a powerful means of identifying contacts between macromolecules in solution.

The earliest structures of protein–RNA complexes followed a few years later. In 1995 the first structure detailing protein–RNA contacts was reported, the solution structure of a 14-residue peptide from the bovine immunodeficiency virus Tat (*trans*-activator) protein in complex with its target RNA, TAR (Puglisi *et al*. 1995; Ye *et al*. 1995). This was followed shortly by the NMR structure of a peptide from the HIV-1 Rev (regulator of viral expression) protein bound to a stem-loop of the Rev response element (RRE) RNA (Battiste *et al*. 1996). Both of these complexes are vital for the regulation of viral replication and provide insights

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

into this process. At the same time, the structures contributed to the general understanding of protein–RNA recognition. In both structures arginine-rich peptides that are unstructured alone fold into conformations that nestle within the major groove of an RNA stem-loop structure. The conformational flexibility of the RNA is also important to the interactions, enabling a 'mutually induced' fit of protein and RNA. For example, the stem of TAR is an A-form helix containing two bases in bulges and a widened major groove. Upon binding to Tat peptide, one of the bulged bases, which had been stacked into the helix, becomes unstacked in order to contact the peptide, which in turn adopts a β-hairpin conformation that can fit deeply into the widened major groove of the TAR RNA. The interaction maximizes the interface between peptide and RNA both at the bases and along the backbone. Similarly extensive contacts are made in the Rev–RRE complex, but in this case the peptide adopts an α-helical conformation that fits deeply into a distinctively shaped major groove of an A-type RNA helix. These are two early examples of the diversity of structure in RNA–protein interactions. Since then, there have been several more solution structures of peptide–RNA complexes showing binding to different RNA conformations by equally small fragments of proteins. Structures of these complexes have been reviewed recently by Patel (1999).

Structures of larger complexes of protein modules and intact proteins with RNA have been slow in coming. The structure of the HIV-1 nucleocapsid protein bound to a step-loop in genomic **Ψ** RNA shows how a protein module can make extensive contacts with an RNA stem and an adjacent hairpin loop simultaneously (De Guzman *et al*. 1998). The N-terminus of the protein forms a 3_{10} helix that fits in the major groove of the RNA while the two zinc knuckles interact with exposed guanosines within the loop. The linker between the knuckles rigidifies upon binding to the RNA, but all other aspects of the induced fit involve changes in only the RNA structure. Not so for the recently reported structure of the complex of the ribosomal protein L30 with its helix–loop–helix target RNA (Mao *et al*. 1999). In this regulatory complex induced fit involves ordering of several loops at one end of the protein as well as internal loops in the RNA. Because mobile loops are so often critical for specific recognition and can present an obstacle to crystallographic studies, NMR may be particularly suitable for studies of proteins that are involved in such complexes.

2.3 RNA structures

The inherent flexibility of RNA makes it difficult to study, yet it is this malleability that enables it to perform such a wide range of functions within cells. Early NMR structures of RNA alone focused on defining the structures of the simple building blocks that make up more complex structures. First came the solution structure of a 12-residue hairpin by Tinoco and coworkers (Cheong *et al*. 1990; Varani *et al*. 1991), followed shortly by the structure of a more extended 25-base helical duplex containing non-Watson-Crick base pair (White *et al*. 1992). The structure of a short, parallel-stranded RNA tetraplex of G- and U-quartets added another type of structural motif to the mix (Cheong & Moore, 1992). Since then, methods for labeling RNA have been developed and it is now possible to work with larger and more complex RNAs.

One class of RNA, catalytic RNA, is critical in RNA processing and translation. The hairpin ribozyme is an example of a small catalytic RNA whose structure has recently been pieced together by NMR (Butcher *et al*. 1999; Cai & Tinoco, 1996). The RNA is 50 residues long and consists of two loops, one of which contains the bound substrate RNA. Solution

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

structures of these two loops have been solved separately and used to derive a model for the 22 kDa ribozyme. The larger of these is a 38-nucleotide piece containing a long stretch of non-Watson–Crick base pairs in its internal loop.

The structure of a 43-nucleotide RNA domain from a different type of RNA, part of the signal recognition particle (SRP) ribonucleoprotein, has recently been solved in its entirety (Schmitz *et al*. 1999a). This complex is responsible for co-translational protein targeting and translocation by directing ribosomes to the protein translocation apparatus of the endoplasmic reticulum. The domain solved is the *E. coli* homolog of one that is essential for all the known interactions of SRP. It contains two small internal loops, one of which is rigid and the other of which is structured yet flexible and thus may act as a hinge. The structure of the rigid loop had been solved previously (Schmitz *et al*. 1999b) and is identical to the structure within the larger fragment, one example of valuable information coming from a reductionistic approach to RNA structure. This structure and the hairpin ribozyme structure are both examples of non-A-form RNA.

Ribosomal RNAs are part of the most complex of RNA and protein machines, the ribosome. NMR spectroscopy and X-ray crystallography have been working hand in hand to reveal the complexities of the RNA structures within this enormous complex. At 120 nucleotides, the 5S RNA is the smallest of the rRNAs, and recently the solution and crystal structures of 'fragment 1', which comprises two thirds of the complete molecule, have been solved (Correll *et al*. 1997; Dallas & Moore, 1997). NMR has contributed structures of two other regions of the ribosome, namely the peptidyl transferase center of the 23S rRNA and the S8-binding site in the 16S rRNA (Kalurachchi *et al*. 1997). The latter structure was determined both in the presence and absence of bound S8 protein. Clearly studying rRNAs in complex with ribosomal proteins is a step in the right direction, since large conformational changes in RNA structure can occur upon binding to protein or adjacent RNA subunits. Indeed interactions with S8 indicate that this protein draws together two sections of the 16S rRNA. By breaking down this enormous complex into smaller pieces, we can hope to learn some rules for RNA–protein interactions that may be helpful in modeling other parts of the system.

2.4 Membrane-bound systems

Membrane-bound proteins make up a major fraction of cellular proteins and are responsible for such critical processes as ion transport and cellular responses to the environment. Yet there has been a lack of structural studies in this area due to the difficulty of preparing suitable samples and the need to mimic membrane conditions *in vitro*. Although complete structures of membrane proteins are difficult to obtain, NMR spectroscopy has been contributing to studies of membrane systems for over a decade. The gramicidin A transmembrane ion channel was an early subject of such studies (Arseniev *et al*. 1985). Using sodium dodecyl sulphate in sufficient concentration to form micelles, a membrane-like environment was created in solution. Two-dimensional proton NMR analysis of gramicidin A solubilized within these micelles revealed that the protein forms a pair of head-to-head right-handed, single-stranded β-helices (Arseniev *et al*. 1985). This contrasted with previous structural studies that had relied on ion-induced chemical shift perturbation of carbonyl carbons and had concluded that the gramicidin helices were left-handed (Urry *et al*. 1982, 1983). However, the solution structure has since been supported by further NMR studies in solution and by

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

the solid state NMR structure of gramicidin A as determined in lipid bilayers (Abdul-Manan & Hinton, 1994; Ketchem *et al*. 1997).

Current structural studies of membrane systems are still severely limited by molecular weight, but membrane proteins of more than 50 residues have been studied within homogeneous micelles. So far the structures of several bacteriophage coat proteins, the M2 pore-lining segments from two receptor channels, the acetylcholine receptor and the N-methyl D-aspartate receptor, and the transmembrane domain of glycophorin A have been solved by NMR using micelle samples. (Almeida & Opella, 1997; Henry & Sykes, 1992; MacKenzie *et al*. 1997; Opella *et al*. 1997; Van de Ven *et al*. 1993; Williams *et al*. 1996). The glycophorin A structure contains a 26-amino acid region that forms the transmembrane α-helix. Two such helices pack against one another into a symmetric dimer with a $-40°$ angle between them (MacKenzie *et al*. 1997). Interestingly, dimerization of the helices is mediated solely by van der Waals interactions, the only hydrogen bonds in the dimer lying within each helix. This example of how very specific interactions between transmembrane helices can be enforced has contributed to the understanding of how membrane protein structures are stabilized.

A different membrane model has been devised recently that may be an even better membrane model than micelles. Bicelles are bilayer discs composed of phospholipids and as such they have planar surfaces, very much like a membrane's. Because bicelles orient in the magnetic field of a spectrometer, bicelle samples can be used to measure residual dipolar couplings, which contain long-range angular information (Ottiger & Bax, 1998a, 1999; Tjandra & Bax, 1997) (see Section 3.2.1 below). The measurement of such constraints, which alone may be sufficient to determine the solution structure of a protein, has been the impetus for the development of these systems. Their usefulness in studies of membrane proteins is only beginning to be explored and is evidenced in a few solution structures of bicelle-bound peptides (Losonczi & Prestegard, 1998b; Vold *et al*. 1997). A distinct advantage of bicelle systems is that orientation confers a much greater dispersion in chemical shifts upon the sample than is observed in isotropic samples; however, bicelles are irregularly shaped and large, features that may limit their applicability to solution studies. The design of new lipids explicitly for reconstituting membranes in vitro is an area of active research that could result in bicelles having different properties (Losonczi & Prestegard, 1998a; Ottiger & Bax, 1998a).

Solid-state NMR spectroscopy can also be very useful for studying samples of peptides or proteins embedded in phospholipid bilayers. Sufficient resolution has been obtained to resolve the peaks in spectra of an ^{15}N-labeled 50-residue fd coat protein (Marassi *et al*. 1997). Because line widths in solid-state NMR spectra are independent of molecular weight, spectral complexity will be the major obstacle to studying larger systems rather than the line broadening experienced in solution. Thus solid-state NMR may be particularly suitable for studies of larger membrane-bound systems, although assigning such systems completely will be challenging. Progress in this area has been reviewed recently by Marassi & Opella (1998).

3. NMR spectroscopy today

As evidenced by the examples in the previous section, NMR spectroscopy has developed into a powerful tool for solving macromolecular structures only within the past decade. It is now possible, by combining state-of-the-art instrumentation and methodology with automated or semi-automated data analysis and structure calculations, to solve the structure of a moderate-

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

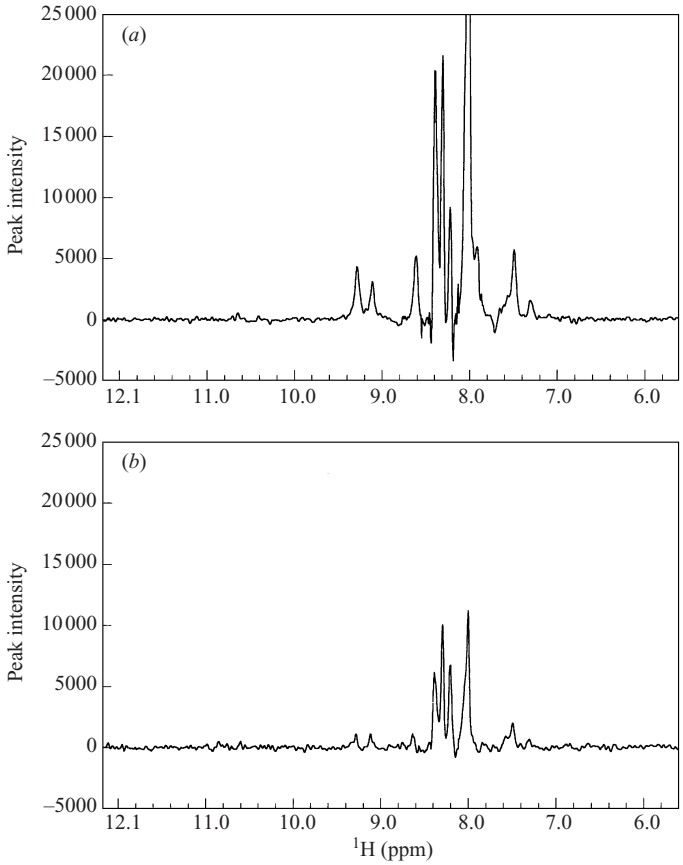


Fig. 3. The gain in signal-to-noise ratio from cryoprobe technology. (*a*) One-dimensional slice from the ^{1}H,^{15}N-HSQC of a 20 kDa protein acquired on a 500 MHz spectrometer equipped with a cryoprobe. (*b*) Analogous slice from the ^{1}H,^{15}N-HSQC spectrum of the same protein at 500 MHz using a state-of-the-art triple-resonance probe.

sized protein within a few months. The historic size limits for NMR analysis have been pushed back with the development of the cryoprobe and TROSY experiments (Pervushin *et al*. 1997), making a wider range of biologically important systems accessible to NMR spectroscopy than ever before. Nor should the unique capabilities of NMR be overlooked. It is the best technique for rapidly assessing whether a protein is folded or not and can be used to study the mobility of proteins, which may play a role in how they interact with ligands or other proteins *in vivo*. In this way, it is complementary to crystallography, which gives a static picture of protein structure and does not have a way of assessing the folding state of a protein.

3.1 State-of-the-art structure determination

Today's state-of-the-art spectrometers possess a range of tools that could not have been conceived a decade ago. Four channels allow detection of ^{1}H, ^{15}N, and ^{13}C nuclei with deuterium decoupling, while sensitivity has been greatly enhanced with the recent introduction of the cryoprobe. By cooling the probe coils to low temperatures (25 K), noise is significantly reduced leading to an increase in signal-to-noise ratio of up to a four-fold relative to standard state-of-the-art probes (Fig. 3). Thus, a triple-resonance experiment

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

normally requiring three to four days of instrument time can now be acquired in a single day, so one could expect to obtain all the data needed for complete assignment of a protein in less than two weeks. Previously, three weeks of instrument time were required just to acquire enough data for unambiguous backbone assignment. In cases where only dilute protein samples can be obtained, this increase in sensitivity will make the difference between being able to solve a structure efficiently and requiring inordinate amounts of instrument time to do so. For protein samples prone to degradation or aggregation this time differential can also make the difference between acquiring good data and data on a poor-quality sample.

Using a set of six triple-resonance experiments leads to complete and unambiguous assignment of most proteins. These experiments are HNCA and HN(CO)CA, HN(CA)CB and HN(COCA)CB, and HNCO and HN(CA)CO, which are performed on ^{2}H-, ^{15}N-, and ^{13}C-labelled samples for maximal signal. The Cβ chemical shifts are usually very well resolved and are sufficient to resolve most ambiguities in the Cα assignments. The C′ resonances are more poorly dispersed, but can also help resolve ambiguities. It is rare that any uncertainty is left in the assignments when Cα, Cβ, and C′ chemical shifts are all considered. The clarity of the data means that backbone assignments can now be obtained within a few days by manual analysis. More significantly, though, as ambiguity is minimized by high-resolution experiments, the potential for automated analysis of the data increases. At the present time, the user must check a program's peak-picking and assignments, but partial automation of the data analysis definitely shortens the amount of time spent in manual analysis. Automated assignment methods have been reviewed very recently by Mosely & Montelione (1999).

Once the backbone of a protein is assigned, one immediately has access to valuable structural information contained in the backbone chemical shifts. It has long been recognized that chemical shifts within regions of regular secondary structure have characteristic values. A 'chemical shift index' for Cα, Hα, Cβ, and C′ has been compiled that can be used to predict regions of secondary structure within a protein (Wishart *et al*. 1992; Wishart & Sykes, 1994). Thus as soon as the backbone has been assigned, the secondary structure can be determined. Moreover, with the newly developed method for direct observation of hydrogen bonds within a protein via an HNCO-type experiment (see Section 3.2.2), it is possible to determine the hydrogen-bonding pattern within a β-sheet or confirm the existence of a helix at a very early stage in the structure determination.

Dihedral angles can also be predicted on the basis of backbone chemical shifts. This is related to the prediction of secondary structures, and can provide additional constraints for regions of the protein lying outside the elements of secondary structure. This method has been dubbed TALOS (torsion angle likelihood obtained from shift and sequence similarity) (Cornilescu *et al*. 1999a). This program has a database of 20 proteins for which chemical shifts and high-resolution X-ray crystal structures are known. Using Cα, Cβ, C′, Hα, and N chemical shifts, the program searches for the closest matches to each triplet within a target protein based on chemical shift and sequence similarity. If the central residues in the best ten matches have similar backbone ϕ and ψ angles, those values can be used as dihedral angle restraints in structure calculations. Although about 3% of the predictions from TALOS may be wrong, mistakes can be identified by the inconsistency of a constraint with other types of data (e.g. NOEs).

With only backbone assignments in hand, one can also obtain residual dipolar coupling data from samples in dilute liquid crystals (see Section 2.1.3). The resulting information on

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

$N–H^N$, Cα–Hα, and Cα–C′ bond vector orientations could be used to orient elements of secondary structure relative to each other, establishing the tertiary structure of the target protein. Thus a preliminary structure of a protein could be determined without assigning any side chains or NOEs. This approach has recently been used to determine the global fold of calerythrin (Aitio *et al*. 1999).

Because Cβ and Cα chemical shifts have already been determined from the first set of experiments, side chain assignment can proceed efficiently using ^{13}C-edited experiments. The H(CC–CO)NH-TOCSY and (H)C(C–CO)NH-TOCSY experiments performed on partially deuterated protein provide a particularly facile means of assigning larger proteins, as demonstrated recently on the 160-residue histone acetyltransferase GCN5 (Lin *et al*. 1999). Since these experiments correlate side chain proton and carbon spins directly with the H^N and N chemical shifts, which have already been assigned, they provide immediate and unambiguous identification of spin systems with particular amino acids. Ambiguities about which proton goes with which carbon can be resolved judiciously with the ^{13}C-NOESY-HSQC, since each H–C pair must show at least some cross-peaks in this spectrum. In general, methyl groups are all observed in the HC(C–CO)NH-TOCSY experiments, but other resonances may be missing. When this occurs incomplete spin systems can be assigned using either the ^{13}C-NOESY-HSQC (Muhandiram *et al*. 1993) or HCCH-TOCSY (Kay *et al*. 1993) spectrum, which is relatively straightforward to interpret once some spins in each side chain have been assigned. Methyl resonances are particularly useful entries into these spectra, since entire spin systems are normally correlated to the methyl groups.

With assignments in hand, the gathering of NOEs and structure refinement can begin. The main sources of distance restraints are the ^{13}C-NOESY-HSQC and ^{15}N-NOESY-HSQC spectra (Fesik & Zuiderweg, 1988; Marion *et al*. 1989a; Vuister *et al*. 1992; Zuiderweg & Fesik, 1989). Both of these experiments are amenable to automated or semi-automated iterative analysis once the global fold of the protein has been determined and the side chain assignments have been completed. In one automated approach the program NOAH is used along with a set of preliminary structures (Mumenthaler *et al*. 1997; Xu *et al*. 1999b). Ambiguous NOESY cross-peaks are assigned to the proton pair predicted to cause fewest violations in the structure. Using the resulting restraints in another round of structure calculations gives a new set of structures in which incorrect assignments show up as consistent distance violations. The NOE list is refined and the process repeated, so that finally an expanded NOE list and family of structures are derived simultaneously. At the present time, some manual intervention is required for good results. Typically the preliminary structure for the iterative assignment of NOEs has been generated from a relatively small number of manually assigned long-range NOEs. When NOAH is run without some NOEs already assigned, the resulting NOE list is not as reliable.

Another approach to iterative assignment of NOEs is exemplified by ARIA, which is used in conjunction with X-PLOR (Nilges *et al*. 1997). In this strategy the chemical shifts of the NOESY cross-peaks are used to generate a set of ambiguous distance constraints that can be satisfied by proximity of any pair of protons at the given chemical shifts. Structures are calculated and ambiguities resolved by determining which pairs of protons are actually satisfying which constraints. Having a good preliminary structure is important for the success of the method, as is having complete side chain assignments. In order to apply this approach to large proteins, which are less likely to be completely assigned, it has been combined with chemical shift prediction methods (Hare & Wagner, 1999). This strategy resulted in

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

assignment of more than 40 side chain protons and 400 new unambiguous NOEs in a 28 kDa single-chain T-cell receptor (Hare & Wagner, 1999). With the improvements in experiments for side chain assignment, it can be anticipated that these automated NOE assignment methods will gain in accuracy and popularity. More detailed reviews of methods for structure refinement and automated analysis of NOESY spectra can be found elsewhere (Güntert, 1998; Mosely & Montelione, 1999).

3.2 New methods

3.2.1 Residual dipolar couplings

Recently, methods have been developed to obtain bond vector orientations from residual dipolar couplings, a measurement that provides a new type of structural constraint completely independent of interproton distances or dihedral angles. The uniqueness of these constraints is that they contain long-distance information relating the relative orientations of bond vectors in remote parts of a molecule. By dissolving a protein in a dilute mixture of phospholipid bicelles in water, a small degree of orientation is conferred upon the protein by the bicelles as they orient in the magnetic field (Tjandra & Bax, 1997). The NMR spectrum of such a sample will still have high resolution, but the residual dipolar couplings will no longer average to zero, as they do in an isotropic sample. Couplings for $N–H^N$, $C\alpha–H\alpha$, and $C\alpha–C'$ bond pairs have been measured and the resulting bond vector orientations have been used in structure calculations (Tjandra *et al*. 1997). $N–H^N$ dipolar couplings can be measured in a simple ^{15}N-HSQC or an IPAP ^{15}N-HSQC in which the multiplet components are separated (Ottiger *et al*. 1998), while the 3D (HA)CA(CO)NH and 2D H(N)CO allow one to detect $C\alpha–H\alpha$, and $C\alpha–C'$ couplings (Ottiger & Bax, 1998b). The stability of bicelle samples can be problematic, which has led to the development of alternate bicelle mixtures and the discovery that proteins can also be oriented by solutions of filamentous phage (Clore *et al*. 1998; Hansen *et al*. 1998).

The greatest value of residual dipolar coupling data may be to provide additional restraints in systems where the short-range NOE and dihedral angle restraints are insufficient to determine some aspect of the structure. Bond vector orientations have been used to determine the relative orientation of subdomains within the ribosomal protein S4 Δ41, resolving a discrepancy between the solution and crystal structures of that system (Markus *et al*. 1999). In another instance, a dearth of NOEs was causing difficulties in positioning two domains of barley lectin relative to each other and information from H–NH residual dipolar couplings was sufficient to define the domain orientation (Fischer *et al*. 1999). Such long-range constraints could also be very useful in determining solution structures of nucleic acids, which often suffer from a paucity of long-range NOEs.

3.2.2 Direct detection of hydrogen bonds

The decrease in line widths made possible by TROSY has been exploited in various experiments for detecting couplings through hydrogen bonds. These methods thus provide the first way to detect hydrogen bonds directly using NMR spectroscopy. The $^2J_{NN}$ couplings across Watson–Crick base pairs in RNA were the first such couplings detected (Dingley & Grzesiek, 1998). Using a quantitative J_{NN} HNN-COSY based on an HNHA experiment, combined with a TROSY-type strategy to increase the imino proton T_2, the signal to noise

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

was sufficiently high that couplings could be observed. While this method has the advantage that the couplings are relatively large (6–7 Hz) compared to other couplings through hydrogen bonds, it is limited to detection of only those hydrogen bonds in which nitrogen acts as the hydrogen bond acceptor. Therefore it cannot be applied to many non-Watson–Crick base pairs or to hydrogen bonding in proteins. Detection of $^{2}J_{NN}$ was confirmed using a [^{15}N,^{1}H]-TROSY on DNA (Pervushin *et al*. 1998), while a modification of the TROSY, dubbed the $^{h}J_{NN}$-correlation-[^{15}N,^{1}H]-TROSY, allowed for the first observation of the small (2–4 Hz) $^{h}J_{HN}$ couplings between the hydrogen in a hydrogen bond and the accepting nitrogen.

Observation of scalar couplings across hydrogen bonds in proteins has been reported only very recently (Cordier & Grzesiek, 1999; Cornilescu *et al*. 1999b). $^{3h}J_{NC'}$ interactions were detected using an HNCO modified to detect these very small couplings (0·25–0·9 Hz). The correlation between $^{3h}J_{NC'}$ and H^{N} chemical shift and the qualitative anti-correlation of $^{3h}J_{NC'}$ with hydrogen bond length, implies that a quantitative treatment of this coupling constant could provide a sensitive measure of hydrogen bond length. Indeed, such a correlation has been established (Cornilescu *et al*. 1999c). The relation should be particularly useful in studies of enzyme catalytic sites or other protein–ligand interactions in which hydrogen bond strength may play a key role.

Extension of these methods to studies of large proteins presents some challenges. With the necessity of deuterating the sample comes the problem of how to exchange deuterons in strong hydrogen bonds for protons. In a recent report of hydrogen bond detection in a 30 kDa protein, almost 40 % of the expected $^{3h}J_{NC'}$ correlations could not be established because tightly hydrogen-bonded amide protons could not be detected (Wang *et al*. 1999b). But this problem would not hamper data collection on hydrogen bonds to amines, since amino protons exchange readily with solvent. A TROSY modification of the HNCO allows such hydrogen bonds to be detected, providing a way to observe hydrogen bonds to side chains. The ability to identify such interactions would be particular useful in studies of protein–nucleic acid complexes, where side chains often make contacts with the DNA bases via hydrogen bonds.

3.2.3 Spin labeling

It has long been recognized that distance-dependent line broadening of resonances can be observed in NMR spectra of proteins containing paramagnetic electrons (e.g. metals or nitroxide spin labels (Krugh, 1976)). However, there are only a few reported attempts at translating this broadening into distance restraints for structure calculations (Gillespie & Shortle, 1997b; Girvin & Fillingame, 1995). Such distances would be unnecessary or redundant when NOEs are plentiful, but might be very valuable for systems in which only a limited number of NOEs can be assigned, as is often the case for large proteins. The usefulness of nitroxide spin labels in NMR was first demonstrated during studies of the unfolded state of staphococcal nuclease (Gillespie & Shortle, 1997a, b). The spin label reacts specifically with cysteine residues, so single cysteine mutants of staphococcal nuclease were engineered that would allow incorporation of spin labels throughout the protein. ^{15}N-HSQC spectra were acquired on the paramagnetic ('oxidized') protein and on the protein after reduction of the spin label. The difference in relaxation properties of the HN spins in the oxidized and reduced proteins should be directly attributable to the effects of the electron

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

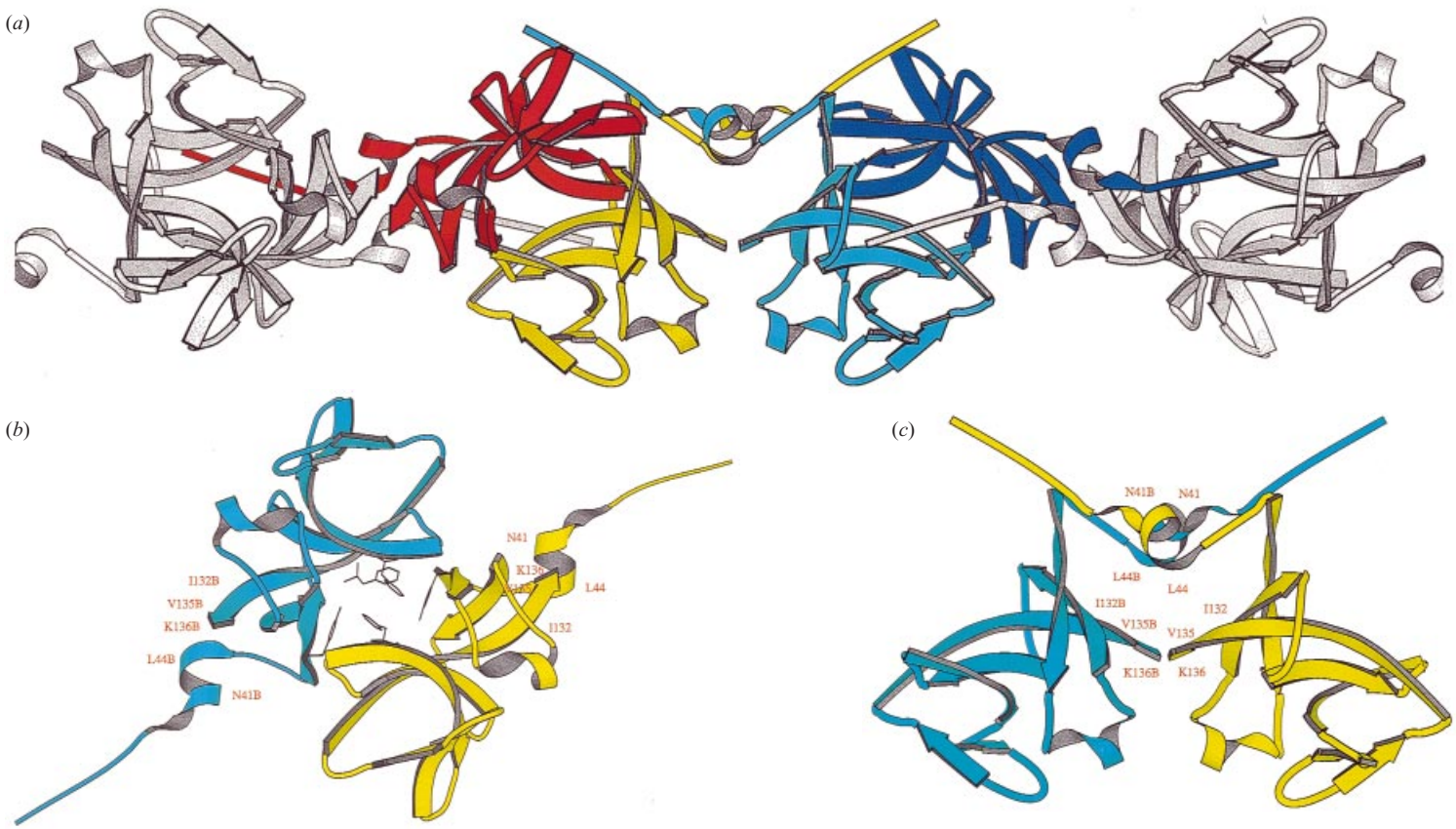


Fig. 4. (*a*) The filament of UmuD′ monomers observed in the X-ray crystallographic studies (Peat *et al*. 1996a). (*b*) The residues that generate inter-monomer NOEs in solution mapped onto the originally proposed molecular dimer of UmuD′. (*c*) The same residues in (*b*) mapped onto the originally designated 'filament dimer' of UmuD′. The filament dimer was concluded to be the dimeric form relevant in solution (Ferentz *et al*. 1997).



Fig. 5. (*a*) The solution structure of the class II single-chain T-cell receptor (Hare *et al*. 1999). (*b*) The crystal structure of the D10 T-cell receptor (blue) in complex with a foreign peptide (red) and the I-A self major histocompatibility complex class II molecule (gray) (Reinherz *et al*. 1999).

spin. Complete relaxation analysis of both spectra yielded information on distances between the protons and the spin label (Gillespie & Shortle, 1997b).

In an exciting recent development, the global fold of a perdeuterated protein has been determined largely on the basis of distance restraints derived from paramagnetic broadening of ^{15}N-HSQC resonances in site-specifically spin-labeled samples (Battiste & Wagner, 2000). In a simplified analysis of the data, relaxation rates were extracted indirectly from the HSQC data using the ratios of cross peak heights in the spectra of oxidized and reduced protein (I_{ox}/I_{red}). Cross peaks were divided into three categories, those that are undetectable in the oxidized HSQC spectrum ($< \sim 14$ Å from the spin label), those with I_{ox}/I_{red} less than 0·85 ($\sim$ 14–23 Å from the label), and those for which I_{ox}/I_{red} was $> 0{\cdot}90$ ($> \sim 23$ Å from the label). In this way five samples containing nitroxide spin labels attached at different residues yielded $\sim$ 500 semi-quantitative restraints for structure calculation. These restraints, along with H^N–H^N NOEs obtained from ^{15}N-edited NOESY spectra and loose backbone angle restraints for secondary structure elements (derived from the chemical shift index), correctly

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

determined the global fold of eIF4E with a backbone precision of 2·3 Å (Battiste & Wagner, 2000). Although eIF4E is only a 25 kDa protein, its structure is studied in CHAPS micelles having a total molecular weight of ~ 45–50 kDa. Thus it has relaxation properties similar to a protein of twice the molecular weight.

One concern about using data from spin-labeled samples is that the label will perturb the structure being studied. However, EPR studies have shown that the nitroxide spin label is readily introduced into all secondary structure elements (particularly helices) and in general does not significantly perturb protein structure (Mchaourab *et al*. 1996). Single-cysteine mutations of eIF4E were made in the loops of the protein or at other solvent-accessible positions using the known three-dimensional structure. With no prior knowledge of a structure, appropriate positions for spin labels could be identified based upon the secondary structure of the protein, as ascertained from the chemical shift index, and the hydrophilicity patterns of the sequence. The need for site-directed mutagenesis and the preparation of several distinct samples are drawbacks of the method. It is also not suitable for proteins containing many cysteine residues. But paramagnetic broadening restraints can provide data on systems too large to give many NOEs. In combination with restraints that rely only on the backbone assignments, such as H^N–H^N NOEs or orientational constraints for H^N–N bonds, paramagnetic broadening could provide extremely valuable information for determining global folding. The resulting preliminary structure and additional NOESY data could then be input into an iterative assignment routine for structure refinement.

3.2.4 Segmental labeling

A new approach being explored for studying large systems is to label only one part of a protein at a time. This strategy can dramatically reduce spectral complexity and allows one to focus on one region of interest within a large molecule or complex. Two different methods for linking together labeled and unlabeled segments have been reported so far, intein-based enzymatic ligation (Yamazaki *et al*. 1998) and chemical ligation (Xu *et al*. 1999a). The intein strategy is based on a naturally occurring system for protein splicing in which two amino acid sequences (the exteins) are linked together post-translationally in a reaction mediated by the intervening amino acid sequence (the intein). To label a segment of a protein using this reaction, a construct is created to express a fusion protein containing the N-terminus of the protein followed by the N-terminal half of the intein. Similarly the C-terminus of the intein is expressed as a fusion protein with the C-terminus of the protein. Since the two halves of the protein are expressed separately they can be labeled differently. The two hybrid proteins are then mixed and splicing occurs to release the segmentally labeled target protein. One drawback of this method is that splicing introduces extra amino acids at the junction of the two segments. It is therefore important to select splicing sites within loops of the protein, but such sites may not identifiable *a priori*. Another point of concern is that the segments of the protein must refold to form an effective intein for the splicing reaction to occur and that refolding and splicing conditions may vary dramatically from protein to protein. Even the two proteins segmentally labeled with this strategy to date, the carboxyl-terminal domain of the RNA polymerase α subunit and maltose binding protein, required different refolding and splicing conditions for optimal yield (Otomo *et al*. 1999).

Chemical ligation of two independently expressed protein fragments overcomes some of these difficulties, but splicing does require a cysteine residue C-terminal to the ligation site.

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

Therefore the C-terminal fragment must begin with a cysteine. In order to react with this moiety, the N-terminal fragment must contain a C-terminal α-thioester. This may be generated by expressing the N-terminus as a fusion protein containing a C-terminal intein and treating this hybrid with ethanethiol (Xu *et al*. 1999a). The resulting α-thioester reacts with the N-terminal cysteine of the C-terminal fragment upon mixing to yield full-length protein. This reaction is not quantitative, but has gone to 70% completion in the case of the ligation of an ethyl α-thioester derivative of an SH3 domain with an SH2 domain (Xu *et al*. 1999a). Such approaches could be extremely useful for studies of specific aspects of very large systems.

3.3 Protein complexes

The large size of many protein complexes and multimeric proteins is becoming less prohibitive with the recent technical advances discussed above, but complexes continue to present challenges for NMR studies. Sample preparation is more difficult than for a single protein and the subsequent data analysis is complicated by the contribution of signals from both proteins and the difficulty of distinguishing intra-protein NOEs from inter-protein NOEs. The latter problem is particularly pronounced in studies of multimeric proteins or heterodimers of two closely related proteins, and a wide variety of techniques have been developed to cope with this challenge. The general strategy is to make the two sides of the dimer or complex 'look' different by isotopic labeling and then to exploit this difference in experiments that select for NOEs between the two proteins.

By labeling a protein with ^{15}N and ^{13}C, and mixing it with unlabeled ligand or protein, a population of complexes is created in which one half of the complex is labeled. If the complex is a homodimeric protein, the resulting mixture will contain a 2:1:1 ratio of labeled–unlabeled heterodimer to labeled homodimer to unlabeled homodimer. Any of a number of experiments can then be used to select for NOEs between protons bound to ^{13}C or ^{15}N and protons bound to ^{12}C or ^{14}N. Such NOEs cannot be attributed to either homodimer, and therefore must be intermonomer NOEs in the labeled–unlabeled heterodimer. The first of this class of filtering experiments used an $X(\omega_1, \omega_2)$ double half filter. In this strategy four two-dimensional spectra are obtained that, by linear combination, allow one to derive a NOESY containing only NOEs between protons that are ^{13}C-labeled and those that are not. Originally used to identify NOEs in a cyclophilin–cyclosporin complex (Wider *et al*. 1990), double half filter experiments have also been applied to dimeric proteins (Folkers *et al*. 1993; Matsuo *et al*. 1995). The filtering concept has been extended to generate a three-dimensional ^{13}C-separated, ^{12}C-filtered NOESY experiment that reveals inter-monomer NOEs directly when one half of a complex is ^{13}C-labeled (Ikura & Bax, 1992). A more recent variation on these filters is a time-shared [^{15}N,^{13}C] double half-filtered 2D NOESY in which NOEs between all labeled and unlabeled protons are detected at once (Burgering *et al*. 1993, 1994). Unfortunately, all these experiments are quite insensitive, but any unambiguous NOEs detected allow the protein–protein or protein–ligand interface to be delineated allowing subsequent assignments to be made on the basis of the resulting preliminary structure.

Technically simpler methods for identifying inter-molecular NOEs include use of selectively deuterated samples and purely computational approaches. When samples of dimeric proteins are prepared in which only certain amino acids have been deuterated,

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

differences can be observed between the NOESY spectra of the dimers containing different deuteration patterns and mixed dimers containing 1:1 ratios of differently deuterated monomers. Differences in NOE intensities can be used to identify inter-monomer NOEs (Arrowsmith *et al*. 1990; Jia *et al*. 1994). In a purely computational approach that was used to solve a coiled-coil structure, NOEs involving residues in the coiled-coil were described as being ambiguous (either inter-monomer or intra-monomer) and the structure calculations were allowed to sort out how best to satisfy the constraints (Junius *et al*. 1996). With no *a priori* knowledge of the structure, it might be difficult to utilize such an approach efficiently. Although there have been cases in which identification of NOEs inconsistent with the rest of the structure was straightforward (Clore *et al*. 1989; Lodi *et al*. 1995), this is seldom trivial and experimental methods for explicitly identifying inter-monomer NOEs are generally the better choice.

A more recent method to identify inter-monomer NOEs in dimeric proteins is to acquire a ^{15}N-NOESY-HSQC on a mixture of 100% ^{2}H, ^{15}N-labeled protein and unlabeled protein (Walters *et al*. 1996). This spectrum selects for NOEs from amide protons attached to ^{15}N to side chain protons, which must be in the unlabeled monomer, and is far more sensitive than the half-filter experiments. Clearly, this experiment only detects NOEs from the backbone of one monomer to the side chains of the other. This limitation can be advantageous, though, because identification of the dimerization interface depends only upon the backbone assignments and can therefore be done very early on in NMR studies. Application of this approach to the UmuD′ dimer allowed for the rapid identification of the dimerization interface of that protein in solution (Ferentz *et al*. 1997). The result resolved an open question left by the crystal structure as to which of two possible interfaces observed in the crystal was the relevant interface in solution (Fig. 4) (Peat *et al*. 1996a, b).

Mapping the interface in a heterodimer complex can be very straightforward if the spectrum of each protein had been assigned prior to the study of the complex. When one component is labeled and the other is not, the latter can be titrated into the former and the ^{15}N-HSQC monitored. Surface residues of the labeled protein that become buried upon addition of the unlabeled component of the complex will experience a changed magnetic environment that will result in a change in chemical shift of the corresponding peak. Thus, as in the SAR by NMR approach discussed below, the binding surface can be mapped. If each component can be labeled separately, both sides of the interface can be delineated.

Another approach to studying complexes has been exemplified by work on the immune system. Single-chain antibodies were engineered long ago and have been used in many studies of antibody–antigen binding. An NMR spectroscopic comparison of the variable domain (Fv) of the McPC603 antibody to an Fv single-chain construct revealed that the linker did not significantly alter the structure of the Fv (Freund *et al*. 1994). The approach of joining together two weakly interacting modules with a flexible linker could be of general use for structural studies of complexes. Covalently linking the components reduces the entropic penalty for complex formation and ensures a one-to-one ratio of the molecules, solving stoichiometric problems in sample preparation. This approach may be particularly valuable for studying complexes with high dissociation constants or in cases where one component of the complex is insoluble on its own or expresses very poorly. One drawback is the resulting difficulty of identifying inter-domain NOEs, since it is no longer straightforward to label each half of the complex separately. A single-chain construct has recently been used to solve the structure of the variable domains from a major histocompatibility complex (MHC) class II T-

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

cell receptor and to give insight into the specificity of recognition by this receptor (Fig. 5(*a*)) (Hare *et al*. 1999). This solution structure was helpful in the crystal structure determination of the larger complex of this receptor with an MHC class II molecule and a foreign peptide (Fig. 5(*b*)) (Reinherz *et al*. 1999). Such combinations of NMR and crystallographic studies can be very powerful in approaching complexes or proteins that present difficulties for either method alone (see Section 4.6 below).

3.4 Mobility studies

Although static structures of proteins are very informative, they are often insufficient to account completely for the workings of a system. Mobility makes a critical contribution to the energetics of macromolecular interactions. Complex formation, whether between proteins, proteins and nucleic acids, or proteins and small molecule ligands, often involves rigidifying flexible regions of the interacting moieties. This entropic cost of binding helps to ensure the selectivity of the interactions. Thus identifying flexible regions of a molecule can give insight into its interactions and functioning. A variety of NMR methods now exist for monitoring motions in macromolecules on time scales from picoseconds to milliseconds (Kay, 1998). This broad range of time scales accessible to NMR allows one to observe modes of motion as diverse as bond vector movements on the picosecond to nanosecond time scale, and millisecond to microsecond motions characteristic of conformational exchange. The ability to measure $N–H^N$ and $C\alpha–C'$ motions and to measure both auto- and cross-correlation S^2 values means that different motional models can be distinguished, providing a detailed dynamic picture of proteins.

Relaxation properties of ^{15}N spins are the most commonly measured parameters for studies of protein dynamics on the picosecond to nanosecond time scale. These parameters allow one to monitor the mobility of $N–H^N$ bond vectors in the backbone (Kay *et al*. 1989) and side chains (Pascal *et al*. 1995) of a protein, which allows for determination of the overall anisotropic motion of the protein, the amplitude of internal motions, and the effective rate constant for internal dynamics. Techniques have also been developed to measure the relaxation of the alpha and carbonyl carbons, generating the corresponding data for $C\alpha–C'$ bond vectors (Dayie & Wagner, 1997; Fischer *et al*. 1997; Yamazaki *et al*. 1994c). This additional information allows one to distinguish types of motion within a protein and has been used to identify concerted motion of a helix around its axis in *E. coli* flavodoxin (Fischer *et al*. 1997). More recently, methods have been developed for measuring side chain dynamics. The interpretation of relaxation data on side chain carbon atoms has been vastly simplified by an innovative labeling technique in which every other carbon in the side chains carries a ^{13}C label, thus eliminating contributions from $^{13}C–^{13}C$ couplings (LeMaster & Kushlan, 1996). Side chain dynamics can also be monitored by measuring ^2H relaxation properties of methyl and methylene groups in proteins that have been fractionally deuterated and uniformly ^{13}C labeled (Muhandiram *et al*. 1995; Yang *et al*. 1998).

With all these types of data available, detailed dynamic pictures of proteins can now be obtained. Comparative studies of the mobility of proteins alone and in complexes is giving insight into the role dynamics play in complex formation and selectivity. The side chain mobilities of contact residues often decrease upon binding of a protein to its target, as would be expected within a tightly packed interface. Backbone dynamics are often affected as well,

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

with reduced mobility of flexible loops at the interface being commonly observed. Flexibility enables enzymes to accommodate a range of substrates and to release the product after catalysis, as in the case of β-lactamase, which binds substrates of many different shapes and sizes via a flexible β-hairpin loop (Scrofani *et al*. 1999). Similarly, the nucleotide excision repair protein XPA binds a myriad of lesions in single- or double-stranded DNA using flexible loops that become more rigid upon binding (Buchko *et al*. 1999).

However, binding is not always accompanied by decreased mobility of contact residues, nor can the dynamics at an interface be predicted based on the behavior of homologous systems. Even complexes that have very similar structural features may differ radically in their dynamic properties. For example, when the PLCC and NSyp SH2 domains bind to their target peptides, the methyl groups at the interface are much more mobile in PLCC than in Nsyp, a difference that exists despite their similar binding affinities and that could not have been anticipated based on the static structures of the two complexes (Kay *et al*. 1998). In cases where mobility at an interface is decreased upon binding, there are often compensatory increases in mobility at residues adjacent to the interface. This has been observed in the NSyp SH2 domain–phosphopeptide complex and in the binding of U1A protein to stem loop II of U1A snRNA (Kay *et al*. 1998; Mittermaier *et al*. 1999). Increasing entropy at a region of the protein not directly involved in binding may be one way in which proteins compensate for the entropic disadvantage of rigidifying side chains within the interface, thus fine tuning the stringency of their selectivity.

3.5 Determination of time-dependent structures

As described above, the vast majority of protein mobility studies by NMR have provided information on which parts of a protein are flexible and qualitative data on regions of a protein that change upon ligand binding. Relaxation data can also provide initial quantitative measurements of conformational entropies. The essence of protein mobility, however, would be to generate a complete description of time-dependent internal protein motions on the basis of experimental data. Such detailed information on motions is contained implicitly in two types of NMR data, NOEs and relaxation parameters. Thus, it should be possible to obtain quantitative information about the structures of conformational substates, and may even allow one to determine directions and amplitudes of motions.

There have been several recent approaches to gleaning such information from NMR data. James and coworkers have fit experimental NOE data on a DNA octamer with an ensemble of structures using the PARSE (Probability assessment via Relaxation rates of a Structural Ensemble) procedure. Despite substantial efforts, the authors have not yet been able to describe an ensemble of substates conclusively (Ulyanov *et al*. 1995). Meanwhile, Brüschweiler & Case (1994) have described motions around the average structure in a collective NMR relaxation model. Using normal mode analysis, they have defined collective coordinates and adjusted the amplitudes and directions of motions to fit the experimental dynamic parameters. Along a similar line, Nilges and Hilbers (Horstink *et al*. 1999) have been studying protein motions using molecular dynamics calculations. They have shown that order parameters calculated from their simulations were in good agreement with the experimental order parameters. The eigenvectors and eigenvalues of the covariance matrix (second moment matrix) yield directions (collective coordinates) and amplitudes of the motions, respectively. When applied to a single-stranded DNA-binding protein from phage Pf3, this approach

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

indicated that relatively few concerted motions could explain the protein's relaxation properties.

Recently, Kitao and Wagner have developed a slightly different approach for deriving populations of conformational substates along with directions and amplitudes of internal motions. The key to this strategy is to maintain consistency with structural constraints and dynamic parameters, such as order parameters, throughout the calculations (Kitao *et al*. 2000). The underlying assumption of the approach is that proteins exist in many diverse conformational substates. Internal motions are then considered to be fluctuations within and jumps between the minima. First, an ensemble of structures is calculated that is as diverse as possible within the bounds of the experimental NOE constraints. This is achieved by long molecular dynamics simulations that maximize the range of structures adopted by the protein. A representative set of conformations is picked along the time course of the MD calculations. These structures are then subjected to short energy minimizations and the resultant coordinates are taken as putative conformational substates. Short MD calculations that keep each structure within its local energy minimum are used to calculate the contribution of intra-substate motions to the order parameters. The averaging due to jumping among minima (JAM) is then analyzed by fitting populations of substates to the experimentally measured order parameters, using an equation for S^2 derived by Henry & Szabo (1985). The eigenvectors and eigenvalues of the second-moment matrix yield the directions and amplitudes of the motions along orthogonal collective coordinates. This JAM analysis was applied to the protein CD2. Very few conformational substates were found to be populated a significant amount of the time, and very few collective modes dominate the internal motions of human CD2. Remarkably, the directions of the motions resemble the conformational change occurring upon binding the counter-receptor CD58 (Kitao & Wagner, 2000; Kitao *et al*. 2000).

3.6 Drug discovery

Among structural techniques, NMR is uniquely capable of identifying weak interaction between macromolecules and ligands. This property has recently been exploited in a novel drug discovery technique called SAR (structure–activity relationships) by NMR (Shuker *et al*. 1996). In this strategy, binding of test compounds to a target protein can be assayed rapidly by observing chemical shift perturbations in the ^{15}N-HSQC spectrum of the protein. The beauty of the method is that very weak binding can be detected (K_d as high as 10 mM) so the 'hit' rate is relatively high. Lead compounds are then optimized for binding by screening analogs to identify more tightly binding molecules. If the HSQC spectrum has been assigned, binding sites are easily identified by noting the peaks that are perturbed by addition of lead compounds. Next, two ligands are identified that bind to adjacent sites in the protein and the structure of the ternary complex is determined by NMR or crystallography methods. With structural information in hand, a linker is designed to connect the two ligands. This results in a compound with a binding constant that is the product of the binding constants of the two separate ligands. Thus, a molecule with nanomolar binding can be derived from two ligands having micromolar binding constants. Using this approach, compounds have been designed that have nanomolar binding affinities for the FK506 binding protein (Shuker *et al*. 1996). The method has also resulted in nanomolar inhibitors of stromelysin (Hajduk *et al*. 1997; Olejniczak *et al*. 1997) and holds great promise for drug design.

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

Another recent tool for drug discovery that does not require that the target protein be labeled or assigned is the SHAPES strategy (Fejzo *et al*. 1999). This approach aims to generate lead compounds by screening binding of the 'SHAPES library' of 132 molecules. These compounds are soluble, easy to synthesize, and represent the full range of molecular frameworks found in known drugs. NMR screening is even simpler than in SAR by NMR. Instead of looking for changes in the HSQC spectrum of labeled target protein, line broadening and transferred NOEs are observed in the spectrum of the small molecule. When the target is large (> 60 kDa), changes in 1D tNOE NMR experiments are sufficient to discern binding, while smaller targets (10–60 kDa) require that additional 2D NOE spectra be acquired to detect micromolar to millimolar binding. One great advantage of this method over SAR by NMR is that there is no weight limit on the target protein and that the search requires only a single screen of a relatively small library. Therefore the quantities of protein required are very modest and can come from eukaryotic or prokaryotic sources. The results from this initial screen can be used to direct high-throughput screens. This typically results in a tenfold increase in hits relative to screening random compounds.

4. The future of NMR

Among the many useful applications of NMR, the most important for molecular biology are the abilities to solve solution structures of proteins and nucleic acids and to characterize molecular interactions. Developments over the past few years have increased the efficiency of the technique, leading to an optimistic outlook for the future of NMR in molecular biology despite its current limitations. From our perspective, the following aspects contribute to the bright future of NMR.

4.1 The ease of structure determination

With the technological refinements of the past few years, solving the structure of a small protein that gives good NMR spectra has become routine. Sophisticated, optimized pulse sequences are widely available through web-sites and are provided in packages from spectrometer manufacturers. The number of trained spectroscopists able to use these experiments efficiently is increasing. At the same time, protocols for structure calculations are becoming standardized in many laboratories. It has become common understanding that only completely unambiguous NOEs should be used to calculate the global fold of a protein, with ambiguous NOEs being added only after the initial fold has been determined with certainty. With these methods, structures of proteins up to 25 kDa can be solved within two months, provided the proteins are stable, soluble, do not aggregate at concentrations of at least 0·5 mM, and have good relaxation properties.

Improved instrumentation will further facilitate structure determination by NMR. The recently introduced cryoprobe boosts the sensitivity of protein spectra by at least a factor of three (Fig. 3). This can reduce the measuring time by almost tenfold for samples giving strong signals. The cryoprobe will also make it possible to record spectra with very high signal-to-noise in a reasonable amount of time and will increase the feasibility of solving structures of proteins with low solubility.

Automated assignment routines will certainly become functional soon. So far, programs

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

suffer from attempts to use spectra with low signal-to-noise ratios or limited resolution, or sets of spectra with small variations in peak position between experiments. As sensitivity increases with hardware innovations, such as the cryoprobe, complete sets of high-sensitivity experiments can be recorded sequentially. This should eliminate many of the problems that currently hamper automated assignment procedures.

4.2 The ease of making recombinant protein

Recent innovations in protein expression are expected to increase the efficiency with which protein structures can be determined. In present day NMR structure determination, the biggest problems are not the technical aspects of spectroscopy, spectral analysis, or structure calculations. The most time-consuming part of a project, and the most limiting factor, is the expression of soluble protein at high yield in a system that allows isotope labeling. Bacterial cells provide the most convenient and most commonly used expression system. Optimization of the yield may require testing of a variety of expression vectors and protein constructs. Commercial systems are now available that facilitate this process by enabling rapid transfer of genes into many different expression vectors (Liu *et al*. 1998; Walhout *et al*. 2000). They also allow for production of constructs having different lengths and a variety of purification tags.

Almost all labeled proteins are produced in *Escherichia coli*; however, the *E. coli* cytoplasm is maintained in a reduced state that disfavors the formation of stable disulfide bonds. Since many important proteins contain disulfides, this presents a substantial limitation for sample preparation. To overcome this problem, *E. coli* strains have recently been constructed that have a highly oxidizing cytoplasm yet grow normally (Bessette *et al*. 1999). Such strains will facilitate the production of labeled proteins containing multiple disulfides. Alternatively, the yeast *Picchia pastoris* can be used to express disulfide-containing proteins, but yeast requires a richer growth medium than bacteria and yields can be highly variable.

4.3 Post-translationally modified proteins

An area in which bacterial expression systems fall short is in the production and labeling of proteins that require post-translational modifications. Glycosylation, myristoylation, phosphorylation, and sulfonation do not take place in *E. coli*. Therefore, proteins with such modifications have been expressed primarily in mammalian cells, insect cells, or in yeast. Isotopic labeling in such expression systems is generally problematic. However, recent years have seen significant progress in producing large quantities of post-translationally modified proteins, and it is likely that more progress is imminent. The yeast *Picchia pastoris* has been used particularly successfully in expression of isotopically labeled proteins for NMR studies (Denton *et al*. 1998). Although the glycosylation pattern in *P. pastoris* may not be the same as in mammalian systems, glycans can still assist in producing folded and soluble samples. They are often needed to ensure that a protein folds properly and is soluble. Alternatively, we have recently demonstrated that one can eliminate the need for post-translational modification of glycoproteins by site-directed mutagenesis. The resulting mutants can be expressed in bacteria in high yield without loss of function, resulting in samples suitable for structure determination by NMR (Sun *et al*. 1999). Ikura and coworkers have shown that recoverin, a myristoylated protein can be expressed in *E. coli* when N-myristoyl CoA

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

transferase is co-expressed (Ames *et al*. 1994). Similar strategies can be expected to work for other post-translationally modified proteins.

4.4 Approaches to large and/or membrane-bound proteins

With the combination of recent innovations in protein expression, NMR hardware development and new pulse sequences, solution structures of large protein complexes may possibly be solved in the near future. NMR technology will soon be poised to tackle membrane protein structures. The high sensitivity of cryoprobes and the power of TROSY experiments will make it possible to solve structures of small membrane-bound proteins by NMR, provided the proteins can be expressed and labeled. Thus, the main obstacle is the difficulty of expressing labeled membrane proteins at high yield. One approach is to express the protein in *E. coli* and transfer it to micelles after purification. This method enabled Engelmann and coworkers to solve the structure of glycophorin bound to phospholipid micelles (MacKenzie *et al*. 1997). When this approach fails, as in the case of the G-protein coupled receptor rhodopsin (Eilers *et al*. 1999), an alternate strategy is to ^{15}N-label the protein by residue type using suspensions of HEK293 cells. As further labeling strategies are developed, the feasibility of solving NMR structures of membrane proteins will increase.

4.5 NMR in structural genomics

At the brink of a new millennium, the entire human genome is about to open up before us. The almost continuous sequence of human chromosome 22 has just been reported (Dunham *et al*. 1999), with the promise that, within a few years, the remaining chromosomal sequences will follow. This wealth of information will present new challenges for a range of fields, including structural biology. The impetus of structural genomics will be to express proteins rapidly in order to determine their structures. Even identifying the fold of a new protein can provide clues as to what it does and can be used to direct mutational studies and ligand-binding experiments. Thus, structural biology faces the challenge of developing methods for very rapid structure determination. Only in this way can we expect to be able to achieve the high-throughput that will be required for projects on a genome-wide scale.

NMR spectroscopy promises to become a tool of choice for structural genomics. Although large proteins and complexes still present challenges for solution structure determination, NMR spectroscopy is able to determine the fold of small proteins within a few weeks of obtaining a labeled sample. Therefore, when the rapidity of structure determination is the primary criterion, NMR should be the method of choice for small proteins. The only requirement is that the protein be sufficiently soluble and not aggregate. For small proteins fulfilling these criteria, the average time needed for structure determination using NMR spectroscopy may be even shorter than that needed for crystallographic structure determination, since finding conditions for crystallization may be quite time consuming.

4.6 Synergy of NMR and crystallography in protein structure determination

There is much to be gained by regarding protein structure determination as an effort of the structural biology community rather than of an individual NMR or crystallography group. NMR spectroscopy can play an important role in such an initiative by serving as a bridge

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

between sample production and final structure determination. The appearance of simple 1D or 2D NMR spectra indicate immediately whether a protein is folded or not. This information can help crystallographers decide whether it is reasonable to pursue crystallization or whether they must search further for conditions under which the protein will fold. Furthermore, NMR can detect large disordered segments within a protein and can guide efforts to truncate a protein to its structured region.

Synergy can go even further. For example, the structure of the important anti-apoptotic protein Bcl-xL was solved as a joint effort of crystallography and NMR. The protein contains a large unstructured loop, making determination of the crystal structure difficult. Complete structure determination by NMR was also difficult, but the secondary structure could be identified more easily using this technique. Thus, the structure could be solved most efficiently by a combination of the information from both methods (Muchmore *et al*. 1996). Another example is the combined effort to solve structures of glycosylated T-cell receptors. NMR studies have revealed the structure and functional significance of the N-linked high-mannose glycan in human CD2 (Wyss *et al*. 1995). This has led to the design of a mutant that is folded and functional even in the absence of the carbohydrate. Subsequently, a mutant form of the counter-receptor CD58 was designed that does not require its glycans either (Sun *et al*. 1999). This is remarkable since two thirds of the molecular weight of the wild-type protein is sugar. NMR spectra showed that the glycan-free mutants of CD2 and CD58 behave well and form a tight complex. The complex crystallized readily and a crystal structure of the complex could be solved rapidly (Wang *et al*. 1999a), revealing a novel hydrophilic interaction mode of immunoglobulin superfamily folds.

5. Conclusion

Although NMR spectroscopy is still a relatively new method for structure determination as compared to X-ray crystallography, it has come into its own over the course of the past decade. With the significant recent improvements in techniques for expressing labeled protein, the latest spectrometer hardware, and the newest methods for pulse programming and data analysis, it is now possible to determine structures of larger and more complex systems more rapidly than ever before. We look forward to a future in which NMR spectroscopy and X-ray crystallography work hand in hand to address structural questions on a genomic scale.

6. Acknowledgements

We thank Bryan Vought and Brian Hare for comments on the manuscript, Jim Sun for providing Fig. 3, and Brian Hare for Fig. 5.

7. References

Abdul-Manan, N. & Hinton, J. F. (1994). Conformation states of gramicidin A along the pathway to the formation of channels in model membranes determined by 2D NMR and circular dichroism spectroscopy. *Biochemistry* **33**, 6773–6783.

Aitio, H., Annila, A., Heikkinen, S., Thulin, E., Drakenberg, T. & Kilpelainen, I. (1999). NMR assignments, secondary structure, and global fold of calerythrin, an EF-hand calcium-binding protein from *saccharopolyspora erythraea*. *Protein Sci*. **8**, 2580–2588.

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

Almeida, F. & Opella, S. J. (1997). fd coat protein structure in membrane environments: structural dynamics of the loop between the hydrophobic transmembrane helix and the amphipathic in-plane helix. *J. molec. Biol.* **270**, 481–495.

Ames, J. B., Tanaka, T., Stryer, L. & Ikura, M. (1994). Secondary structure of myristoylated recoverin determined by three-dimensional heteronuclear NMR: implications for the calcium-myristoyl switch. *Biochemistry* **33**, 10743–10753.

Arnold, J. T., Dharmatti, S. S. & Packard, M. E. (1951). Chemical effects on nuclear induction signals from organic compounds. *J. chem. Phys.* **19**, 507.

Arrowsmith, C. H., Pachter, R., Altman, R. B., Iyer, S. B. & Jardetzky, O. (1990). Sequence-specific ^{1}H NMR assignments and secondary structure in solution of *Escherichia coli trp* repressor. *Biochemistry* **29**, 6332–6341.

Arseniev, A., Barsukov, I. L., Bystrov, V. F., Lomize, A. L. & Ovchinnikov, Y. (1985). ^{1}H-NMR study of gramicidin A transmembrane ion channel. Head-to-head right-handed, single-stranded helices. *FEBS Lett.* **186**, 168–174.

Battiste, J. L., Mao, H., Rao, N. S., Tan, R., Muhandiram, D. R., Kay, L. E., Frankel, A. D. & Williamson, J. R. (1996). Alpha helix–RNA major groove recognition in an HIV-1 Rev peptide–Rre RNA complex. *Science* **273**, 1547–1551.

Battiste, J. L. & Wagner, G. (2000). Utilization of site-directed spin labeling and high resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear Overhauser effect data. *Biochemistry* **39**, 5355–5365.

Bessette, P. H., Aslund, F., Beckwith, J. & Georgiou, G. (1999). Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm. *Proc. natn. Acad. Sci. USA* **96**, 13703–13708.

Billeter, M., Kline, A. D., Braun, W., Huber, R. & Wüthrich, K. (1989). Comparison of the high-resolution structures of the α-amylase inhibitor tendamistat determined by nuclear magnetic resonance in solution and by X-ray diffraction in single crystals. *J. molec. Biol.* **206**, 677–687.

Billeter, M., Qian, Y. Q., Otting, G., Müller, M., Gehring, W. J. & Wüthrich, K. (1993). Determination of the NMR solution structure of an Antennapedia homeodomain–DNA complex. *J. molec. Biol.* **234**, 446–462.

Bloch, F. (1946). Nuclear induction. *Phys. Rev.* **70**, 460.

Bolognesi, M., Gatti, G., Menagatti, E., Guarneri, M., Marquart, M., Papamokos, E. & Huber, R. (1982). Three-dimensional structure of the complex between pancreatic secretory trypsin inhibitor (Kazal type) and trypsinogen at 1·8 Å resolution. Structure solution, crystallographic refinement and preliminary structural interpretation. *J. molec. Biol.* **162**, 839–868.

Braun, W., Bösch, C., Brown, L. R., Go, N. & Wuthrich, K. (1981). Combined use of proton–proton Overhauser enhancements and a distance geometry algorithm for determination of polypeptide conformations. Application to micelle-bound glucagon. *Biochim. biophys. Acta* **667**, 377–396.

Braun, W., Epp, O., Wüthrich, K. & Huber, R. (1989). Solution of the phase problem in the X-ray diffraction method for proteins with the nuclear magnetic resonance solution structure as initial model. Patterson search and refinement for the alpha-amylase inhibitor tendamistat. *J. molec. Biol.* **206**, 669–676.

Braun, W. & Go, N. (1985). Calculation of protein conformations by proton–proton distance constraints. A new efficient algorithm. *J. molec. Biol.* **186**, 611–626.

Braun, W., Wider, G., Lee, K. H. & Wuthrich, K. (1983). Conformation of glucagon in a lipid–water interphase by 1H nuclear magnetic resonance. *J. molec. Biol.* **169**, 921–948.

Brüschweiler, R. & Case, D. A. (1994). Collective NMR relaxation model applied to protein dynamics. *Phys. Rev. Lett.* **72**, 940–943.

Buchko, G. W., Daughdrill, G. W., de Lorimier, R., Rao, S., Isern, N. G., Lingbeck, J. M., Taylor, J.-S., Wold, M. S., Gochin, M., Spicer, L. D., Lowry, D. F. & Kennedy, M. A. (1999). Interactions of human nucleotide excision repair protein XPA with DNA and RPA 70ΔC327: chemical shift mapping and ^{15}N NMR relaxation studies. *Biochemistry* **38**, 15116–15128.

Burgering, M., Boelens, R. & Kaptein, R. (1993). Observation of intersubunit NOEs in a dimeric P22 Mnt repressor mutant by a time-shared [^{15}N,^{13}C] double half-filter technique. *J. biomol. NMR* **3**, 709–714.

Burgering, M. J. M., Boelens, R., Gilbert, D. E., Breg, J. N., Knight, K. L., Sauer, R. T. & Kaptein, R. (1994). Solution structure of dimeric Mnt repressor (1–76). *Biochemistry* **33**, 15036–15045.

Butcher, S. E., Allain, F. H.-T. & Feigon, J. (1999). Solution structure of the loop B domain from the hairpin ribozyme. *Nature struct. Biol.* **6**, 212–216.

Caffrey, M., Cai, M., Kaufman, J., Stahl, S. J., Wingfield, P. T., Covell, D. G., Gronenborn, A. M. & Clore, G. M. (1998). Three-dimensional solution structure of the 44 kDa ectodomain of SIV gp41. *EMBO J.* **17**, 4572–4584.

Cai, Z. & Tinoco, I. J. (1996). Solution structure of loop A from the hairpin ribozyme from tobacco ringspot virus satellite. *Biochemistry* **35**, 6026–6036.

Cheong, C. & Moore, P. B. (1992). Solution structure of an unusually stable RNA tetraplex containing G- and U-quartet structures. *Biochemistry* **31**, 8406–8414.

Cheong, C., Varani, G. & Tinoco, I. (1990). Solution structure of an unusually stable RNA hairpin, 5′GGAC(UUCG)GUCC. *Nature* **346**, 680–682.

Chuprina, V. P., Rullman, J. A. C., Lamerichs,

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

R. M. J. N., Van Boom, J. H., Boelens, R. & Kaptein, R. (1993). Structure of the complex of lac repressor headpiece and an 11 base-pair half-operator determined by nuclear magnetic resonance spectroscopy and restrained molecular dynamics. *J. molec. Biol.* **234**, 446–462.

Clore, G. M., Appella, E., Yamada, M., Matsushima, K. & Gronenborn, A. M. (1989). Determination of the secondary structure of interleukin-8 by nuclear magnetic resonance spectroscopy. *J. biol. Chem.* **264**, 18907–18911.

Clore, G. M. & Gronenborn, A. M. (1998). NMR structure determination of proteins and protein complexes larger than 20 kDa. *Curr. Opin. Chem. Biol.* **2**, 564–570.

Clore, G. M., Starich, M. R. & Gronenborn, A. M. (1998). Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. *J. Am. chem. Soc.* **120**, 10571–10572.

Clore, G. M., Wingfield, P. T. & Gronenborn, A. M. (1991). High-resolution three-dimensional structure of interleukin 1β in solution by three- and four-dimensional nuclear magnetic resonance spectroscopy. *Biochemistry* **30**, 2315–2323.

Clubb, R. T., Thanabal, V. & Wagner, G. (1992). A constant-time three-dimensional triple-resonance pulse scheme to correlate intraresidue proton ($^{1}H^{N}$), nitrogen-15 and carbon-13 ($^{13}C'$) chemical shifts in nitrogen-15-carbon-13-labeled proteins. *J. magn. Reson.* **97**, 213–217.

Cordier, G. & Grzesiek, S. (1999). Direct observation of hydrogen bonds in proteins by interresidue $^{3h}J_{NC'}$ scalar couplings. *J. Am. chem. Soc.* **121**, 1601–1602.

Cornilescu, G., Delaglio, F. & Bax, A. (1999a). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. biomol. NMR* **13**, 289–302.

Cornilescu, G., Hu, J.-S., Bax, A. (1999b). Identification of the hydrogen bonding network in a protein by scalar coupling. *J. Am. chem. Soc.* **121**, 2949–2950.

Cornilescu, G., Ramirez, B. E., Frank, M. K., Clore, G. M., Gronenborn, A. M. & Bax, A. (1999c). Correlation between $^{3h}J_{NC'}$ and hydrogen bond length in proteins. *J. Am. chem. Soc.* **121**, 6275–6279.

Correll, C. C., Freeborn, B., Moore, P. B. & Steitz, T. A. (1997). Metals, motifs and recognition in the crystal structure of a 5S rRNA domain. *Cell* **91**, 705–712.

Crippen, G. M. & Havel, T. F. (1978). Stable calculation of coordinates from distance information. *Acta crystallogr.* A**34**, 282–284.

Dallas, A. & Moore, P. (1997). The loop E loop D region of Escherichia coli 5S rRNA: the solution structure reveals an unusual loop that may be important for binding ribosomal proteins. *Structure* **5**, 1639–1653.

Dayie, K. T. & Wagner, G. (1997). Carbonyl carbon probe of local mobility in ^{13}C,^{15}N-enriched proteins using high-resolution nuclear magnetic resonance. *J. Am. chem. Soc.* **119**, 7797–7806.

De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N. & Summers, M. F. (1998). Structure of the HIV-1 nucleocapsid protein bound to the SL3 Ψ-RNA recognition element. *Science* **279**, 384–388.

Denton, H., Smith, M., Husi, H., Uhrin, D., Barlow, P. N., Batt, C. A. & Sawyer, L. (1998). Isotopically labeled bovine beta-lactoglobulin for NMR studies expressed in Pichia pastoris. *Protein Expr. Purif.* **14**, 97–103.

Dingley, A. J. & Grzesiek, S. (1998). Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide $^{2}J_{NN}$ couplings. *J. Am. chem. Soc.* **120**, 8293–8297.

Dötsch, V. & Wagner, G. (1998). New approaches to structure determination by NMR. *Curr. Opin. struct. Biol.* **8**, 619–623.

Dubs, A., Wagner, G. & Wüthrich, K. (1979). Individual assignments of amide proton resonances in the proton NMR spectrum of the basic pancreatic trypsin inhibitor. *Biochim. biophys. Acta* **577**, 177–194.

Dunham, I., Shimizu, N., Roe, B. A., Chissoe, S. *et al.* (1999). The DNA sequence of human chromosome 22. *Nature* **402**, 489–495.

Eilers, M., Reeves, P. J., Ying, W., Khorana, H. G. & Smith, S. O. (1999). Magic angle spinning NMR of the protonated retinylidene Schiff base nitrogen in rhodopsin: expression of 15N-lysine- and 13C-glycine-labeled opsin in a stable cell line. *Proc. natn. Acad. Sci. USA* **96**, 487–492.

Fejzo, J., Lepre, C. A., Peng, J. W., Bemis, G. W., Ajay, Murcko, M. A. & Moore, J. M. (1999). The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chemistry & Biology* **6**, 755–769.

Ferentz, A., Opperman, T., Walker, G. & Wagner, G. (1997). Dimerization of the UmuD′ protein in solution and its implications for regulation of SOS mutagenesis. *Nature struct. Biol.* **4**, 979–983.

Fesik, S. W. & Zuiderweg, E. R. P. (1988). Heteronuclear three dimensional NMR spectroscopy. A strategy for the simplification of homonuclear two dimensional NMR spectra. *J. magn. Reson.* **78**, 588–593.

Fischer, M. W., Losonczi, J. A., Weaver, J. L. & Prestegard, J. H. (1999). Domain orientation and dynamics in multidomain proteins from residual dipolar couplings. *Biochemistry* **38**, 9013–9022.

Fischer, M. W. F., Zeng, L., Pang, Y., Hu, W., Majumdar, A. & Zuiderweg, E. R. P. (1997).

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

Experimental characterization of models for backbone picosecond dynamics in proteins. Quantification of NMR auto- and cross-correlation relaxation mechanisms involving different nuclei of the peptide plane. *J. Am. chem. Soc.* **119**, 12629–12642.

Folkers, P. J. M., Folmer, R. H. A., Konings, R. N. H. & Hilbers, C. W. (1993). Overcoming the ambiguity problem encountered in the analysis of nuclear overhauser magnetic resonance spectra of symmetric dimer proteins. *J. Am. chem. Soc.* **115**, 3798–3799.

Freund, C., Ross, A., Pluckthun, A. & Holak, T. A. (1994). Structural and dynamics properties of the Fv fragment and the single-chain Fv fragment of an antibody in solution investigated by heteronuclear three-dimensional NMR spectroscopy. *Biochemistry* **33**, 3296–3303.

Garrett, D. S., Seok, Y.-J., Liao, D.-I., Peterkofsky, A., Gronenborn, A. M. & Clore, G. M. (1997). Solution structure of the 30 kDa N-terminal domain of enzyme I of the *Escherichia coli* phosphoenolpyruvate:sugar phosphotransferase system by multidimensional NMR. *Biochemistry* **36**, 2517–2530.

Garrett, D. S., Seok, Y. J., Peterkofsky, A., Gronenborn, A. M. & Clore, G. M. (1999). Solution structure of the 40,000 Mr phosphoryl transfer complex between the N-terminal domain of enzyme I and HPr. *Nat. struct. Biol.* **6**, 166–173.

Gillespie, J. R. & Shortle, D. (1997a). Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. *J. molec. Biol.* **268**, 158–169.

Gillespie, J. R. & Shortle, D. (1997b). Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures. *J. molec. Biol.* **268**, 170–184.

Girvin, M. E. & Fillingame, R. H. (1995). Determination of local protein structure by spin label difference 2D NMR: the region neighboring Asp61 of subsunit c of the F_1F_0 ATP synthase. *Biochemistry* **34**, 1635–1645.

Gordon, S. L. & Wüthrich, K. (1978). Transient proton–proton Overhauser effects in horse ferrocytochrome c. *J. Am. chem. Soc.* **100**, 7094–7096.

Grzesiek, S., Anglister, J. & Bax, A. (1993). Correlation of backbone amide and aliphatic sidechain resonances in 13C/15N-enriched proteins by isotropic mixing of 13C magnetization. *J. magn. Reson.* B**101**, 114–119.

Grzesiek, S. & Bax, A. (1992a). An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J. magn. Reson.* **99**, 201–207.

Grzesiek, S. & Bax, A. (1992b). Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *J. magn. Reson.* **96**, 432–440.

Grzesiek, S. & Bax, A. (1993). Amino acid type determination in the sequential assignment procedure of uniformly $^{13}C/^{15}N$-enriched proteins. *J. biomol. NMR* **3**, 185–204.

Güntert, P. (1998). Structure calculations of biological macromolecules from NMR data. *Q. Rev. Biophys.* **31**, 145–237.

Gutowsky, H. S.., McCall, E. W. & Slichter, C. P. (1951). Coupling among nuclear magnetic dipoles in molecules. *Phys. Rev.* **84**, 589–596.

Hahn, E. L. & Maxwell, D. E. (1951). Chemical shift and field independent frequency modulation of the spin echo envelope. *Phys. Rev.* **84**, 1246–1247.

Hajduk, P. J., Sheppard, G., Nettesheim, D. G., Olejniczak, E. T., Shuker, S. B., Meadows, R. P., Steinman, D. H., Carrera, Jr., G. M., Marcotte, P. A., Severin, J., Walter, K., Smith, J., Gubbins, E., Simmer, R., Holtzman, T. F., Morgan, D. W., Davidsen, S. K., Summers, J. B. & Fesik, S. W. (1997). Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR. *J. Am. chem. Soc.* **119**, 5818–5827.

Hansen, M. R., Mueller, L. & Pardi, A. (1998). Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nature struct. Biol.* **5**, 1065–1074.

Hare, B. J. & Wagner, G. (1999). Application of automated NOE assignment to three-dimensional structure refinement of a 28 kDa single-chain T cell receptor. *J. biomol. NMR* **15**, 103–113.

Hare, B. J., Wyss, D. F., Osburne, M. S., Kern, P. S., Reinherz, E. L. & Wagner, G. (1999). Structure, specificity and CDR mobility of a class II restricted single-chain T-cell receptor. *Nature struct. Biol.* **6**, 574–581.

Havel, T. F., Kuntz, I. D. & Crippen, G. M. (1983). Theory and practice of distance geometry. *Bull. math. Biol.* **45**, 665–720.

Havel, T. F. & Wüthrich, K. (1984). A distance geometry program for determining the structure of small proteins and other macromolecules from nuclear magnetic resonance measurements of intramolecular ^1H–^1H proximities in solution. *Bull. math. Biol.* **46**, 673–698.

Henry, E. R. & Szabo, A. (1985). Influence of vibrational motion on solid state line shapes and NMR relaxation. *J. chem. Phys.* **82**, 4753–4761.

Henry, G. D. & Sykes, B. D. (1992). Assignments of ^1H and ^{15}N NMR resonances in detergent solubilized M13 coat protein: a model for the coat protein dimer. *Biochemistry* **31**, 5284–5297.

Horstink, L. M., Abseher, R., Nilges, M. & Hilbers, C. W. (1999). Functionally important correlated motions in the single-stranded DNA-binding protein

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

encoded by filamentous phage Pf3. *J. molec. Biol.* **287**, 569–577.

Ikura, M. & Bax, A. (1992). Isotope-filtered 2D NMR of a protein-peptide complex: study of a skeletal muscle myosin light chain kinase fragment bound to calmodulin. *J. Am. chem. Soc.* **114**, 2433–2440.

Ikura, M., Kay, L. E. & Bax, A. (1990a). A novel approach for sequential assignment of ^{1}H, ^{13}C and ^{15}N spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* **29**, 4659–4667.

Ikura, M., Marion, D., Kay, L. E., Shih, H., Krinks, M., Klee, C. B. & Bax, A. (1990b). Heteronuclear 3D NMR and isotopic labeling of calmodulin. Towards the complete assignment of the 1H NMR spectrum. *Biochem. Pharmacol.* **40**, 153–160.

Jasanoff, A., Wagner, G. & Wiley, D. C. (1998). Structure of a trimeric domain of the MHC class II-associated chaperonin and targeting protein Ii. *Embo J.* **17**, 6812–6818.

Jia, X., Reisman, J. M., Hsu, V. L., Geiduschek, E. P., Parello, J. & Kearns, D. R. (1994). Proton and nitrogen NMR sequence-specific assignments and secondary structure determination of the Bacillus subtilis SPO1-encoded transcription factor 1. *Biochemistry* **33**, 8842–8852.

Junius, F. K., O'Donoghue, S., Nilges, M., Weiss, A. S. & King, G. F. (1996). High resolution NMR solution structure of the leucine zipper domain of the c-Jun homodimer. *J. biol. Chem.* **271**, 13663–13667.

Kalurachchi, K., Uma, K., Zimmermann, R. A. & Nikonowicz, E. P. (1997). Structural features of the binding site for ribosomal protein S8 in *Escherichia coli* 16S rRNA defined using NMR spectroscopy. *Structure* **5**, 1639–1653.

Kaptein, R., Zuiderweg, E. R., Scheek, R. M., Boelens, R. & van Gunsteren, W. F. (1985). A protein structure from nuclear magnetic resonance data. lac repressor headpiece. *J. molec. Biol.* **182**, 179–182.

Kay, L. E. (1998). Protein dynamics from NMR. *Nature Struct. Biol.* **5**, 513–517.

Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. (1990). Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J. magn. Reson.* **89**, 496–514.

Kay, L. E., Muhandiram, D. R., Wolf, G., Shoelson, S. E. & Forman-Kay, J. D. (1998). Correlation between binding and dynamics at SH2 domain interfaces. *Nature struct. Biol.* **5**, 156–162.

Kay, L. E., Torchia, D. & Bax, A. (1989). Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**, 8972–8979.

Kay, L. E., Xu, G.-Y., Singer, A. U., Muhandiram, D. R. & Forman-Kay, J. D. (1993). A gradient-enhanced HCCH-TOCSY experiment for recording side-chain ^{1}H and ^{13}C correlations in H_2O samples of proteins. *J. magn. Reson.* B**101**, 333–337.

Kay, L. E., Xu, G. Y. & Yamazaki, T. (1994). Enhanced-sensitivity triple-resonance spectroscopy with minimal H_2O saturation. *J. magn. Reson.* A**109**, 129–133.

Kendrew, J. C. *et al.* (1960). Structure of myoglobin. *Nature* **185**, 422–427.

Ketchem, R. R., Roux, B. & Cross, T. A. (1997). High-resolution polypeptide structure in a lamellar phase lipid environment from solid state NMR derived orientational constraints. *Structure* **5**, 1655–1669.

Kitao, A. & Wagner, G. (2000). Structure refinement of human CD2 versus geometric and dynamic constraints. *J. molec. Biol.* (submitted).

Kitao, A., Wagner, G. & Sci, P. N. A. (2000). A space-time structure determination of human CD2 reveals the CD58-binding mode. *Proc. natn. Acad. Sci. USA* (in press).

Kline, A. D., Braun, W. & Wüthrich, K. (1986). Studies by 1H nuclear magnetic resonance and distance geometry of the solution conformation of the alpha-amylase inhibitor tendamistat. *J. molec. Biol.* **189**: 377–382.

Kline, A. D., Braun, W. & Wüthrich, K. (1988). Determination of the complete three-dimensional structure of the α-amylase inhibitor tendamistat in aqueous solution by nuclear magnetic resonance and distance geometry. *J. molec. Biol.* **204**, 675–724.

Knight, W. D. (1949). Nuclear magnetic resonance shift in metals. *Phys. Rev.* **76**, 1259–1260.

Krugh, T. R. (1976). In *Spin Labeling Theory and Applications* (ed. L. J. Berliner), pp. 339–372. New York: Academic Press.

LeMaster, D. M. & Kushlan, D. M. (1996). Dynamical mapping of *E. coli* thioredoxin via ^{13}C NMR relaxation analysis. *J. Am. chem. Soc.* **118**, 9255–9264.

LeMaster, D. M. & Richards, F. M. (1985). 1H–15N heteronuclear NMR studies of Escherichia coli thioredoxin in samples isotopically labeled by residue type. *Biochemistry* **24**, 7263–7268.

Lin, Y., Fletcher, C. M., Zhou, J., Allis, C. D. & Wagner, G. (1999). Solution structure of the catalytic domain of GCN5 histone acetyltransferase bound to coenzyme A. *Nature* **400**, 86–89.

Lin, Y. & Wagner, G. (1999). Efficient side-chain and backbone assignment in large proteins: application to tGCN5. *J. biomol. NMR* **15**, 227–239.

Liu, Q., Li, M. Z., Leibham, D., Cortez, D. & Elledge, S. J. (1998). The univector plasmid-fusion system, a method for rapid construction of recombinant DNA without restriction enzymes. *Curr. Biol.* **8**, 1300–1309.

Lodi, P. J., Ernst, J. A., Kuszewski, J., Hickman, A. B., Engelman, A., Craigie, R., Clore, G. M. &

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

Gronenborn, A. M. (1995). Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry* **34**, 9826–9833.

Logan, T. M., Olejniczak, E. T., Xu, R. X. & Fesik, S. W. (1993). A general method for assigning NMR spectra of denatured proteins using 3D HC(CO)NH-TOCSY triple resonance experiments. *J. biomol. NMR* **3**, 225–231.

Loria, J. P., Rance, M. & Palmer, A. G. (1999). Transverse-relaxation-optimized (TROSY) gradient-enhanced triple-resonance NMR spectroscopy. *J. magn. Reson.* **141**, 180–184.

Losonczi, J. A. & Prestegard, J. H. (1998a). Improved dilute bicelle solutions for high-resolution NMR of biological macromolecules. *J. biomol. NMR* **12**, 447–451.

Losonczi, J. A. & Prestegard, J. H. (1998b). Nuclear magnetic resonance characterization of the myristoylated N-terminal fragment of ADP-ribosylation factor 1 in a magnetically oriented membrane array. *Biochemistry* **37**, 706–716.

MacKenzie, K. R., Prestegard, J. H. & Engelman, D. M. (1997). A transmembrane helix dimer: structure and implications. *Science* **276**, 131–133.

Mao, H., White, S. A. & Williamson, J. R. (1999). A novel loop–loop recognition motif in the yeast ribosomal protein L30 autoregulatory RNA complex. *Nature struct. Biol.* **6**, 1139–1147.

Marassi, F. M. & Opella, S. J. (1998). NMR structural studies of membrane proteins. *Curr. Opin. struct. Biol.* **8**, 640–648.

Marassi, F. M., Ramamoorthy, A. & Opella, S. J. (1997). Complete resolution of the solid-state NMR spectrum of a uniformly ^{15}N-labeled membrane protein in phospholipid bilayers. *Proc. natn. Acad. Sci. USA* **94**, 8551–8556.

Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M. & Clore, G. M. (1989a). Overcoming the overlap problem in the assignment of ^{1}H NMR spectra of larger proteins by use of three-dimensional heteronuclear ^{1}H–^{15}N Hartmann–Hahn–multiple quantum coherence and nuclear Overhauser–multiple quantum coherence spectroscopy: application to interleukin 1β. *Biochemistry* **28**, 6150–6156.

Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A. & Bax, A. (1989b). Three-dimensional heteronuclear NMR of ^{15}N-labeled proteins. *J. Am. chem. Soc.* **111**, 1515–1517.

Markus, M. A., Gerstner, R. B., Draper, D. E. & Torchia, D. A. (1999). Refining the overall structure and subdomain orientation of ribosomal protein S4 Δ41 with dipolar couplings measured by NMR in uniaxial liquid crystalline phases. *J. molec. Biol.* **292**, 375–387.

Matsuo, H., Li, H., McGuire, A. M., Fletcher, C. M., Gingras, A. C., Sonenberg, N. & Wagner, G. (1997). Structure of translation factor eIF4E bound to m7GDP and interaction with 4E-binding protein. *Nat. struct. Biol.* **4**, 717–724.

Matsuo, H., Li, H. & Wagner, G. (1996). A sensitive HN(CA)CO experiment for deuterated proteins. *J. magn. Reson.* B**110**, 112–115.

Matsuo, H., Shirakawa, M. & Kyogoku, Y. (1995). Three-dimensional dimer structure of the λ-Cro repressor in solution as determined by heteronuclear multidimensional NMR. *J. molec. Biol.* **254**, 668–680.

Mchaourab, H. S., Lietzow, M. A., Hideg, K. & Hubbell, W. L. (1996). Motion of spin-labeled side chains in T4 lysozyme. Correlation with protein structure and dynamics. *Biochemistry* **35**, 7692–7704.

Meissner, A. & Sørensen, O. W. (1999). Optimization of three-dimensional TROSY-type HCCH NMR correlation of aromatic (1)H-(13)C groups in proteins. *J. magn. Reson.* **139**, 447–450.

Mittermaier, A., Varani, L., Muhandiram, D. R., Kay, L. E. & Varani, G. (1999). Changes in side-chain and backbone dynamics identify determinants of specificity in RNA recognition by human U1A protein. *J. molec. Biol.* **294**, 967–979.

Montelione, G. T. & Wagner, G. (1990). Triple resonance experiments for establishing conformation-independent sequential NMR assignments in isotope-enriched polypeptides. *J. magn. Reson.* **87**, 183–188.

Moseley, H. N. B. & Montelione, G. T. (1999). Automated analysis of NMR assignments and structures for proteins. *Current Opin. struct. Biol.* **9**, 635–642.

Muchmore, S. W., Sattler, M., Liang, H., Meadows, R. P., Harlan, J. E., Yoon, H. S., Nettesheim, D., Chang, B. S., Thompson, C. B., Wong, S. L., Ng, S. L. & Fesik, S. W. (1996). X-ray and NMR structure of human Bcl-xL, an inhibitor of programmed cell death. *Nature* **381**, 335–341.

Muhandiram, D. R., Farrow, N. A., Xu, G.-Y., Smallcombe, S. H. & Kay, L. E. (1993). A gradient ^{13}C NOESY-HSQC experiment for recording NOESY spectra of ^{13}C-labeled proteins dissolved in H_2O. *J. magn. Res.* B**102**, 317–321.

Muhandiram, D. R. & Kay, L. E. (1994). Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. *J. magn. Reson.* B**103**, 203–216.

Muhandiram, D. R., Yamazaki, T., Sykes, B. D. & Kay, L. E. (1995). Measurement of ^{2}H T_1 and $T_{1\rho}$ relaxation times in uniformly ^{13}C-labeled and fractionally ^{2}H-labeled proteins in solution. *J. Am. chem. Soc.* **117**, 11536–11544.

Mumenthaler, C., Güntert, P., Braun, W. & Wüthrich, K. (1997). Automated combined assignment of NOESY spectra and three-dimensional protein structure determination. *J. biomol. NMR* **10**, 351–362.

Nilges, M., Macias, M. J., O'Donoghue, S. I. &

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

Oschkinat, H. (1997). Automated NOESY interpretation with ambiguous distance restraints: the refined NMR solution structure of the pleckstrin homology domain from β-spectrin. *J. molec. Biol.* **269**, 408–422.

Olejniczak, E. T., Hajduk, P. J., Marcotte, P. A., Nettesheim, D. G., Meadows, R. P., Edalji, R., Holtzman, T. F. & Fesik, S. W. (1997). Stromelysin inhibitors designed from weakly bound fragments: effects of linking and cooperativity. *J. Am. chem. Soc.* **119**, 5828–5832.

Omichinski, J. G., Clore, G. M., Schaad, O., Felsenfeld, G., Trainor, C., Appella, E., Stahl, S. J. & Gronenborn, A. M. (1993). Solution structure of the specific DNA complex of the zinc containing DNA binding domain of the erythroid transcription factor Gata-1 by multidimensional NMR. *Science* **261**, 438–446.

Opella, S. J., Gesell, J., Valente, A. P., Marassi, F. M., Oblatt-Montal, M., Sun, W., Ferrer-Montiel, A. & Montal, M. (1997). Structural studies of the pore-lining segments of neurotransmitter-gated ion channels. *Chemtracts Biochem. molec. Biol.* **10**, 153–174.

Otomo, T., Teruya, K., Uegaki, K., Yamazaki, T. & Kyogoku, Y. (1999). Improved segmental labeling of proteins and application to a larger protein. *J. biomol. NMR* **14**, 105–114.

Ottiger, M. & Bax, A. (1998a). Characterization of magnetically oriented phospholipid micelles for measurement of dipolar couplings in macromolecules. *J. biomol. NMR* **12**, 361–372.

Ottiger, M. & Bax, A. (1998b). Determination of relative N–H^N, N–C′, Cα–C′, and Cα–Hα effective bond lengths in a protein by NMR in a dilute liquid crystalline phase. *J. Am. chem. Soc.* **120**, 12334–12341.

Ottiger, M. & Bax, A. (1999). Bicelle-based liquid crystals for NMR-measurement of dipolar couplings at acidic and basic pH values. *J. biomol. NMR* **13**, 187–191.

Ottiger, M., Delaglio, F. & Bax, A. (1998). Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. magn. Reson.* **131**, 373–378.

Otting, G., Qian, Y. Q., Billeter, M., Müller, M., Affolter, M., Gehring, W. J. & Wüthrich, K. (1990). Protein-DNA contacts in the structure of a homeodomain–DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J* **9**, 3085–3092.

Otting, G., Senn, H., Wagner, G. & Wüthrich, K. (1986). Editing of 2D 1H NMR spectra using X half-filters: combined use with residue selective ^{15}N-labeling of proteins. *J. magn. Reson.* **70**, 500–505.

Papamokos, E., Weber, E., Bode, W., Huber, R., Empie, M. W., Kato, I. & Laskowski Jr., M. (1982). Crystallographic refinement of Japanese quail ovomucoid, a Kazal-type inhibitor, and model building studies of complexes with serine proteases. *J. molec. Biol.* **158**, 515–537.

Pascal, S. M., Yamazaki, T., Singer, A. U., Kay, L. E. & Forman-Kay, J. D. (1995). Structural and dynamics characterization of the phosphotyrosine binding region of a Src homology 2 domain–phosphopeptide complex by NMR relaxation, proton exchange, and chemical shift approaches. *Biochemistry* **34**, 11353–11362.

Patel, D. J. (1999). Adaptive recognition in RNA complexes with peptides and protein modules. *Curr. Opin. struct. Biol.* **9**, 74–87.

Peat, T. S., Frank, E. G., McDonald, J. P., Levine, A. S., Woodgate, R. & Hendrickson, W. A. (1996a). Structure of the UmuD′ protein and its regulation in response to DNA damage. *Nature* **380**, 727–730.

Peat, T. S., Frank, E. G., Mcdonald, J. P., Levine, A. S., Woodgate, R. & Hendrickson, W. A. (1996b). The UmuD′ protein filament and its role in damage induced mutagenesis. *Structure* **4**, 1401–1412.

Perutz, M. F. (1960). Structure of hemoglobin. A three-dimensional Fourier synthesis at 5·5 Å resolution, obtained by X-ray analysis. *Nature* **185**, 416–422.

Pervushin, K., Ono, A., Fernández, C., Szyperski, T., Kainosho, M. & Wüthrich, K. (1998a). NMR scalar couplings across Watson–Crick base pair hydrogen bonds in DNA observed by transverse relaxation-optimized spectroscopy. *Proc. natn. Acad. Sci. USA* **95**, 14147–14151.

Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. (1998b). Transverse relaxation-optimized spectroscopy (TROSY) for NMR studies of aromatic spin systems in 13C-labeled proteins. *J. Am. chem. Soc.* **120**, 6394–6400.

Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. (1997). Attenuating T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. natn. Acad. Sci. USA* **94**, 12366–12371.

Pflugrath, J., Wiegand, E., Huber, R. & Vertesy, L. (1986). Crystal structure determination, refinement and the molecular model of the α-amylase inhibitor Hoe-467A. *J. molec. Biol.* **189**, 383–386.

Puglisi, J. D., Chen, L., Blanchard, S. & Frankel, A. D. (1995). Solution structure of a bovine immunodeficiency virus Tat-Tar peptide–RNA complex. *Science* **270**, 1200–1203.

Purcell, E. M., Torrey, H. C. & Pound, R. V. (1946). Resonance absorption by nuclear magnetic moments in solid. *Phys. Rev.* **69**, 37–38.

Reinherz, E. L., Tan, K., Tang, L., Kern, P., Liu, J.-h., Xiong, Y., Hussey, R. E., Smolyar, A., Hare,

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

B., Zhang, R., Joachimiak, A., Chang, H.-C., Wagner, G. & Wang, J.-h. (1999). The crystal structure of a T cell receptor in complex with peptide and MHC class II. *Science* **286**, 1913–1921.

Riek, R., Wider, G., Pervushin, K. & Wüthrich, K. (1999). Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules. *Proc. natn. Acad. Sci. USA* **96**, 4918–4923.

Salzmann, M., Pervushin, K., Wider, G., Senn, H. & Wüthrich, K. (1999a). [13C]-constant-time [15N,1H]-TROSY-HNCA for sequential assignments of large proteins. *J. biomol. NMR* **14**, 85–88.

Salzmann, M., Pervushin, K., Wider, G., Senn, H. & Wüthrich, K. (1998). TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc. natn. Acad. Sci. USA* **95**, 13585–13590.

Salzmann, M., Wider, G., Pervushin, K. & Wüthrich, K. (1999b). Improved sensitivity and coherence selection for [15N,1H]-TROSY elements in triple resonance experiments. *J. biomol. NMR* **15**, 181–184.

Saunders, M., Wishnia, A. & Kirkwood, J. G. (1957). The nuclear magnetic resonance spectrum of ribonuclease. *J. Am. chem. Soc.* **79**, 3289.

Schmitz, U., Behrens, S., Freymann, D. M., Keenan, R. J., Lukavsky, P., Walter, P. & James, T. L. (1999a). Structure of the phylogenetically most conserved domain of SRP RNA. *RNA*, **5**, 1419–1429.

Schmitz, U., James, T. L., Lukavsky, P. & Walter, P. (1999b). Structure of the most conserved internal loop in SRP RNA. *Nature struct. Biol.* **6**, 634–638.

Scrofani, S. D. B., Chung, J., Huntley, J. J. A., Benkovic, S. J., Wright, P. E. & Dyson, H. J. (1999). NMR characterization of the metallo-β-lactamase from *Bacteroides fragilis* and its interaction with a tight-binding inhibitor: role of an active-site loop. *Biochemistry* **38**, 14507–14514.

Shan, X., Gardner, K. H., Muhandiram, D. R., Rao, N. S., Arrowsmith, C. H. & Kay, L. E. (1996). Assignment of ^{15}N, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, and H^{N} resonances in an ^{15}N, ^{13}C, ^{2}H labeled 64 kDa Trp repressor–operator complex using triple-resonance NMR spectroscopy and ^{2}H-decoupling. *J. Am. chem. Soc.* **118**, 6560–6579.

Shuker, S. B., Hajduk, P. J., Meadows, R. P. & Fesik, S. W. (1996). Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **274**, 1531–1534.

South, T. L. & Summers, M. F. (1993). Zinc- and sequence-dependent binding to nucleic acids by the N-terminal zinc finger domain of the HIV-1 nucleocapsid protein: NMR structure of the complex with the Psi-site analog, dACGCC. *Protein Sci.* **2**, 3–19.

Sun, Z. Y., Dötsch, V., Kim, M., Li, J., Reinherz, E. L. & Wagner, G. (1999). Functional glycan-free adhesion domain of human cell surface receptor CD58: design, production and NMR studies. *Embo J.* **18**, 2941–2949.

Takeda, M. & Jardetzky, O. (1957). Proton magnetic resonance of simple amino acids and dipeptides in aqueous solution. *J. chem. Phys.* **26**, 1346–1347.

Tjandra, N. & Bax, A. (1997). Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* **278**, 1111–1114.

Tjandra, N., Omichinski, J., Gronenborn, A., Clore, G. & Bax, A. (1997). Use of dipolar ^{1}H–^{15}N and ^{1}H–^{13}C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nature struct. Biol.* **4**, 732–738.

Ulyanov, N. B., Schmitz, U., Kumar, A. & James, T. L. (1995). Probability assessment of conformational ensembles: sugar repuckering in a DNA duplex in solution. *Biophys. J.* **68**, 13–24.

Urry, D. W., Trapane, T. L. & Prasad, K. U. (1983). Is the gramicidin A transmembrane channel single-stranded or double-stranded helix? A simple unequivocal determination. *Science* **221**, 1064–1067.

Urry, D. W., Walker, J. T. & Trapane, T. L. (1982). Ion interactions in (1-13C)D-Val 8 and D-Leu14 analogs of gramicidin A, the helix sense of the channel and location of ion binding sites. *J. membr. Biol.* **69**, 225–231.

Van de Ven, F. J. M., Os, J. W. M., Aelen, J. M. A., Wymenja, S. S., Remerowski, M. L., Konings, R. N. H. & Hilbers, C. W. (1993). Assignment of ^{1}H, ^{15}N and backbone ^{13}C resonances in detergent-solubilized M13 coat protein via multinuclear multidimensional NMR: a model for the coat protein monomer. *Biochemistry* **32**, 8322–8328.

Varani, G., Cheong, C. & Tinoco, I. (1991). Structure of an unusually stable RNA hairpin. *Biochemistry* **30**, 3280–3289.

Vold, R. R., Prosser, R. S. & Deese, A. J. (1997). Isotropic solution of phospholipic bicelles: a new membrane mimetic for high-resolution NMR studies of polypeptides. *J. biomol. NMR* **9**, 329–335.

Vuister, G. W., Boelens, R., Kaptein, R., Burgering, M. & van Zijl, P. C. M. (1992). Gradient-enhanced 3D NOESY-HMQC spectroscopy. *J. biomol. NMR* **2**, 301–305.

Wagner, G. (1997). An account of NMR in structural biology. *Nature struct. Biol.* **4**, 841–844.

Wagner, G. & Wüthrich, K. (1979). Truncated driven nuclear overhauser effect (TOE). A new technique for studies of selective ^{1}H–^{1}H Overhauser effects in the presence of spin diffusion. *J. magn. Reson.* **33**, 675–680.

Wagner, G. & Wüthrich, K. (1982). Sequential resonance assignments in protein ^{1}H nuclear magnetic resonance spectra. Basic pancreatic trypsin inhibitor. *J. molec. Biol.* **155**, 347–366.

Walhout, A. J., Sordella, R., Lu, X., Hartley, J. L.,

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press

Temple, G. F., Brasch, M. A., Thierry-Mieg, N. & Vidal, M. (2000). Protein interaction mapping in C. elegans using proteins involved in vulval development [see comments]. *Science* **287**, 116–122.

Walters, K. J., Matsuo, H. & Wagner, G. (1996). Use of deuteration to distinguish intermonomer NOEs in homodimeric proteins with C_2 symmetry, submitted for publication. *J. Am. chem. Soc.* **119**, 5958–5959.

Wang, J. H., Smolyar, A., Tan, K., Liu, J. H., Kim, M., Sun, Z. Y., Wagner, G. & Reinherz, E. L. (1999a). Structure of a heterophilic adhesion complex between the human CD2 and CD58 (LFA-3) counter-receptors. *Cell* **97**, 791–803.

Wang, Y.-X., Jacob, J., Cordier, F., Wingfield, P., Stahl, S. J., Lee-Huang, S., Torchia, D., Grzesiek, S. & Bax, A. (1999b). Measurement of $^{3h}J_{NC'}$ connectivities across hydrogen bonds in a 30 kDa protein. *J. biomol. NMR* **14**, 181–184.

White, S. A., Nilges, M., Huang, A., Brunger, A. T. & Moore, P. B. (1992). NMR Analysis of Helix I from the 5S RNA of Escherichia Coli. *Biochemistry* **31**, 1610–1621.

Wider, G., Weber, C., Traber, R., Widmer, H. & Wüthrich, K. (1990). Use of a double-half filter in two-dimensional ^{1}H nuclear magnetic resonance studies of receptor-bound cyclosporin. *J. Am. chem. Soc.* **112**, 9015–9016.

Wider, G. & Wüthrich, K. (1999). NMR spectroscopy of large molecules and multimolecular assembles in solution. *Curr. Opin. struct. Biol.* **9**, 594–601.

Williams, K. A., Farrow, N. A., Deber, C. M. & Kay, L. E. (1996). Structure and dynamics of bacteriophage Ike major coat protein in MPG micelles by solution NMR. *Biochemistry* **35**, 5145–5157.

Williamson, M. P., Havel, T. F. & Wüthrich, K. (1985). Solution conformation and proteinase inhibitor IIA from bull seminal plasma by proton NMR and distance geometry. *J. molec. Biol.* **182**, 295–315.

Wishart, D., Sykes, B. & Richards, F. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure. *Biochemistry* **31**, 1647–1651.

Wishart, D. S. & Sykes, B. D. (1994). The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. biomol. NMR* **4**, 171–180.

Wittekind, M. & Mueller, L. (1993). HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. *J. magn. Reson.* B**101**, 201–205.

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

Wyss, D. F., Choi, J. S., Li, J., Knoppers, M. H., Willis, K. J., Arulanandam, A. R., Smolyar, A., Reinherz, E. L. & Wagner, G. (1995). Conformation and function of the N-linked glycan in the adhesion domain of human CD2. *Science* **269**, 1273–1278.

Xu, R., Ayers, B., Cowburn, D. & Muir, T. W. (1999a). Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies. *Proc. natn. Acad. Sci. USA* **96**, 388–393.

Xu, Y., Wu, J., Gorenstein, D. & Braun, W. (1999b). Automated 2D NOESY assignment and structure calculation of crambin (S22/I25) with the self-correcting distance geometry based NOAH/DIAMOD programs. *J. magn. Reson.* **136**, 76–85.

Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R. & Kay, L. E. (1994a). A suite of triple resonance NMR experiments for the backbone assignment of ^{15}N, ^{13}C, ^{2}H labeled proteins with high sensitivity. *J. Am. chem. Soc.* **116**, 11655–11666.

Yamazaki, T., Lee, W., Revingtom, M., Mattiello, D. L., Dahlquist, F. W., Arrowsmith, C. H. & Kay, L. E. (1994b). An HNCA pulse scheme for the backbone assignment of ^{15}N,^{13}C,^{2}H-labeled proteins: application to a 37-KDa Trp repressor-DNA complex. *J. Am. chem. Soc.* **116**, 6464–6465.

Yamazaki, T., Muhandiram, R. & Kay, L. E. (1994c). NMR experiments for the measurement of carbon relaxation properties in highly enriched, uniformly ^{13}C,^{15}N-labeled proteins: application to $^{13}C^{\alpha}$ carbons. *J. Am. chem. Soc.* **116**, 8266–8278.

Yamazaki, T., Otomo, T., Oda, N., Kyogoku, Y., Uegaki, K., Ito, N., Ishino, Y. & Nakamura, H. (1998). Segmental isotope labeling for protein NMR using peptide splicing. *J. Am. chem. Soc.* **120**, 5591–5592.

Yang, D., Mittermaier, A., Mok, Y. K. & Kay, L. E. (1998). A study of protein side-chain dynamics from new ^{2}H auto-correlation and ^{13}C cross-correlation NMR experiments: application to the N-terminal SH3 domain from drk. *J. molec. Biol.* **276**, 939–954.

Ye, X., Kumar, R. A. & Patel, D. J. (1995). Molecular recognition in the bovine immunodeficiency virus Tat peptide–Tar RNA complex. *Chem. Biol.* **2**, 827–840.

Zuiderweg, E. R. P. & Fesik, S. W. (1989). Heteronuclear three-dimensional NMR spectroscopy of the inflammatory protein C5a. *Biochemistry* **28**, 2387–2391.

https://doi.org/10.1017/S0033583500003589 Published online by Cambridge University Press