Annu. Rev. Biochem. 2004. 73:107–146
doi: 10.1146/annurev.biochem.73.011303.074004

First published online as a Review in Advance on April 2, 2004

NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY OF HIGH-MOLECULAR-WEIGHT PROTEINS

Vitali Tugarinov,[1,2,3] Peter M. Hwang,[2] and Lewis E. Kay[1,2,3]

Departments of Medical Genetics,[1] Biochemistry,[2] and Chemistry,[3] University of Toronto, Ontario, Canada M5S 1A8; email: vitali@pound.med.utoronto.ca, peter@pound.med.utoronto.ca, kay@pound.med.utoronto.ca

Key Words protein NMR, resonance assignments, methyl TROSY, deuteration

■ **Abstract** Recent developments in NMR spectroscopy, which include new experiments that increase the lifetimes of NMR signals or that precisely define the orientation of internuclear bond vectors with respect to a common molecular frame, have significantly increased the size of proteins for which quantitative structural and dynamic information can be obtained. These experiments have, in turn, benefited from new labeling strategies that continue to drive the field. The utility of the new methodology is illustrated by considering applications to malate synthase G, a 723 residue enzyme, which is the largest single polypeptide chain for which chemical shift assignments have been obtained to date. New experiments developed specifically to address the complexity and low sensitivity of spectra recorded on this protein are presented. A discussion of the chemical information that is readily available from studies of systems in the 100 kDa mol wt range is included. Prospects for membrane protein structure determination are discussed briefly in the context of an application to an *Escherichia coli* enzyme, PagP, localized to the outer membrane of gram-negative bacteria.

CONTENTS

INTRODUCTION

Overall Goals

In the past decade, solution NMR spectroscopy has continued to evolve as a powerful tool for the study of biomolecular structure and dynamics. The development of multidimensional, multinuclear NMR spectroscopy (1) in parallel with new labeling methodologies (2) has significantly increased the size of molecules that are now amenable to NMR studies. The introduction of media for obtaining weak alignment of solute molecules (3–7), along with new experiments for measuring the resultant residual couplings (8), and chemical shift changes that accompany alignment (9) have facilitated the measurement of powerful orientational restraints. This, in turn, has led to substantial improvements in the accuracy of NMR-derived structures, especially in cases where only low densities of other restraints are available (10). The development of transverse relaxation-optimized spectroscopy (TROSY) experiments (11, 12) in which the lifetimes of NMR signals are substantially increased has significantly impacted the size limitations that have plagued NMR studies of macromolecules in the past. These advances in concert hold the promise for many interesting applications involving a wide range of biochemical systems.

The new methodologies have been described in detail in a series of reviews, and the interested reader is referred to the literature (1, 12–20). In this review, we choose to not be comprehensive and to not focus on what has been reviewed previously. Instead, we concentrate primarily on one protein that our laboratory has studied over the past several years, malate synthase G, MSG, and illustrate concepts and approaches using this system. MSG is a monomeric enzyme, composed of 723 residues (82 kDa) (21), and at present is close to a factor of two larger than other proteins that have been assigned by multidimensional NMR methods. Experiments that work well for proteins of 40 kDa fail, in some cases, for this large system, and in this review we describe new experimental approaches that circumvent such problems. As described below, MSG is a four-domain enzyme that binds a number of ligands. We present results from studies that measure the relative orientation of domains as a function of ligand,

the thermodynamics and kinetics of ligand binding, as well as hydrogen exchange and site-specific dynamics. An application to the study of structure and dynamics of a membrane enzyme, PagP, (22) is also discussed; it illustrates how the new methodologies have also impacted the study of this important group of molecules. These examples provide the reader with a feel for the information that is available from NMR studies of large proteins and the strong complimentarity between results from NMR and X-ray crystallography. Building on the labeling strategies described herein, a methyl-TROSY experiment is presented that compliments TROSY spectroscopy of backbone amide correlations and promises to extend NMR studies further to even higher mol wt systems.

A Brief Introduction to MSG

MSG catalyzes the Claisen condensation of glyoxylate with an acetyl group of acetyl-coenzyme A (acetyl-CoA), producing malate, an intermediate in the citric-acid-cycle (Figure 1*a*). The abstraction of a proton from the methyl group of the acetyl-CoA thioester presents both a kinetic and a thermodynamic challenge for the weak bases that are typically available in proteins, and it has led to an interest in structural studies of this enzyme. The crystal structure of MSG complexed with magnesium and glyoxylate has been solved at a resolution of 2.0 Å (21). As illustrated in Figure 1*b*, MSG is composed of four main domains: (*a*) a centrally located core domain of the molecule based on a highly stable $\beta 8/\alpha 8$ barrel fold, (*b*) an N-terminal α-helical clasp linked to the first strand of the barrel by a long extended loop, (*c*) an α/β domain appended to the molecular core, and (*d*) the C-terminal end of the enzyme consisting of a five-helix plug connected to the barrel by an extended loop. The active site of the enzyme is located in a cleft at the interface between the C-terminal plug and loops at the C-terminal ends of several of the β-strands of the core barrel (21). In order to complement the existing crystallographic structural studies on the glyoxylate-bound state of the enzyme, our laboratory has initiated studies on the apo form of the protein, as described in detail below.

BACKBONE AND SIDE CHAIN ASSIGNMENTS IN LARGE PROTEINS

Backbone Resonance Assignments

A first step in any detailed study of protein structure and/or dynamics involves backbone resonance assignments. That is, the frequency of absorption of each NMR active spin along the backbone must be determined to be used subsequently as an "atomic signature," which identifies each site in the protein. A number of important developments in the past decade have led to a significant increase in the size of macromolecules for which backbone resonances can be obtained. A major advance in this regard has been the use of highly deuterated

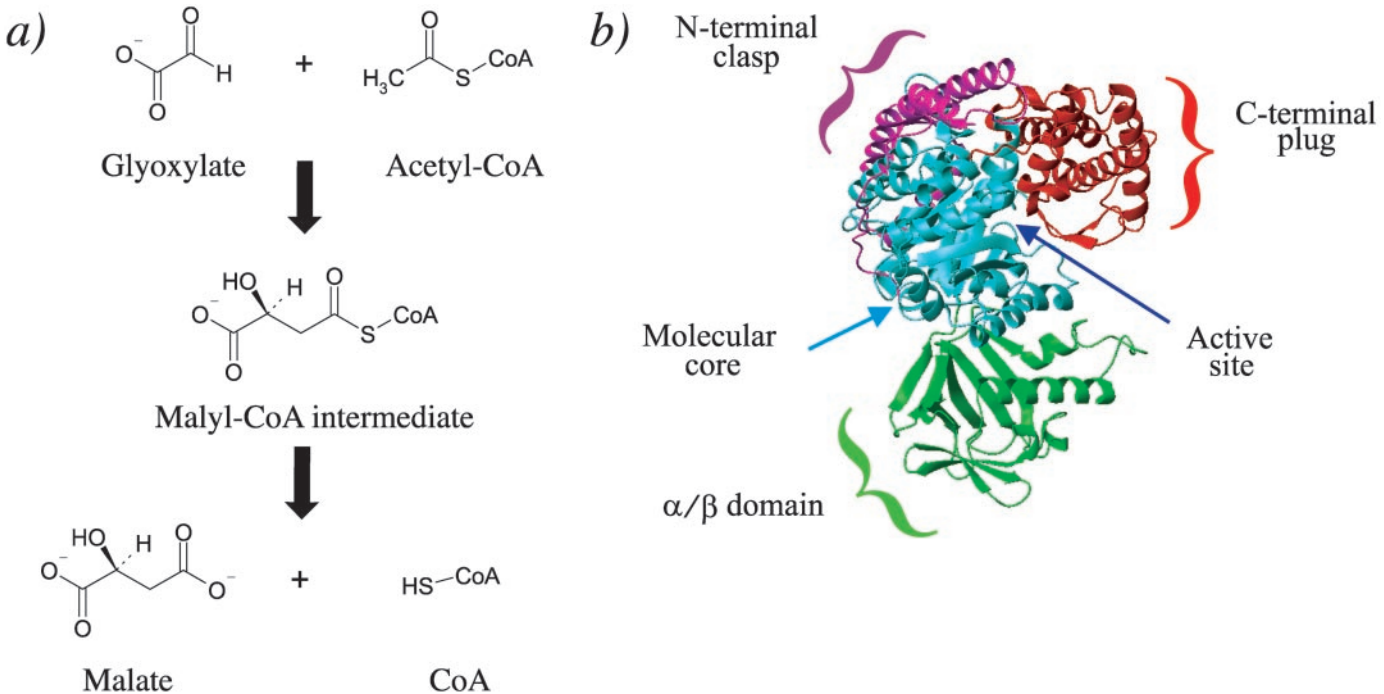


Figure 1 (*a*) Reaction catalyzed by MSG. (*b*) Ribbon diagram of MSG illustrating the four domains of the molecule. MSG coordinates were obtained from PDB accession code (1d8c) (21). Reproduced with permission from Tugarinov & Kay (48).

protein samples, leading to attenuation of the relaxation rates of NMR active nuclei and concomitant sensitivity and resolution gains in spectra (15, 16, 23, 24). Another important contribution has been the development of ^{1}HN-^{15}N TROSY (11) in which the components of ^{1}HN and ^{15}N signals that relax slowly in macromolecules at high magnetic fields are selected. TROSY experiments are also best conducted with samples that are highly deuterated. Not surprisingly, sample preparation is one of the key (and most time consuming) steps in any NMR study of a large protein. Many of the issues have been reviewed previously (15); aspects of sample preparation that required new approaches in the case of MSG are discussed below.

IN VITRO REFOLDING OF LARGE PROTEINS OVEREXPRESSED IN D_2O-BASED MEDIA

Overexpression of proteins in pure D_2O using [^{2}H,^{13}C]-D-glucose as the main carbon source leads to the production of perdeuterated proteins, with deuterium incorporation also at exchangeable backbone and side chain positions. Although subsequent protein purification and sample preparation steps are normally conducted in $^{1}H_2O$ over the course of several days, $^{2}H \rightarrow {}^{1}H$ exchange may be incomplete for amides in very stable structural motifs. Because state-of-the-art NMR methodologies for protein backbone assignments rely on a high occupancy of protons at the amide sites, incomplete $^{2}H \rightarrow {}^{1}H$ exchange can lead to severe losses (or total absence) of NMR signals (25, 26). This problem necessitates the development of efficient protocols for unfolding/refolding of proteins in $^{1}H_2O$ in order to achieve complete exchange in a reasonable time frame.

In order to demonstrate the importance of unfolding/refolding in the case of MSG, a fully protonated sample of the protein was dissolved in D_2O at 25°C and ^{1}HN-^{15}N TROSY-HSQC spectra were recorded over a period of 2 months. Approximately 120 backbone amides remained protonated after 7 days, and roughly 70 were not completely exchanged after a period of 6 weeks. Although near complete backbone assignments were reported for a number of large systems without the need for refolding (27–29), these studies involved proteins composed of several noncovalently linked identical units. The molecular core of each of these relatively small units may be less stable and/or more accessible to solvent than that of a large monomeric molecule.

In vitro refolding of large multidomain proteins is problematic due to complex and poorly understood thermodynamics and kinetics of the folding process (30–32). Several methods are commonly in use; these include (*a*) unfolding the protein with denaturant (6 M GuHCl or urea) followed by slow refolding using dialysis, (*b*) refolding His-tagged proteins on a metal-affinity column using a gradient with decreasing denaturant concentrations, or (*c*) partial protein denaturation (~2 M GuHCl) followed by fast dilution refolding. The latter approach was used previously for NMR studies of perdeuterated MBP in our laboratory (25). However, in the case of MSG, these methods failed to provide sufficiently high yields of refolded protein (26). In contrast, complete denaturation of MSG with 6 M GuHCl and subsequent fast dilution refolding into a denaturant-free

buffer gave refolding yields of 60 ± 15% (26). Rapid dilution from a fully unfolded state was necessary to avoid the aggregation and ensuing precipitation that occur if a partially denatured state is allowed to accumulate. For multidomain proteins with less favorable refolding properties, it may be necessary to resort to refolding chromatography with immobilized minichaperones (33) or to a fast dilution method with the addition of nondetergent sulfobetaines (chemical chaperones) (34). In some cases the addition of binding substrates (small molecules or metal ions) and/or protein stabilizing agents (e.g., sucrose and trehalose) to the refolding medium can help to stabilize the folded state, thereby minimizing aggregation of partially folded intermediates (35). Once refolding is complete, the structural integrity of the protein can be established by comparing peak positions in 2D ^{1}HN-^{15}N correlation spectra with the subset of peaks appearing in spectra of protein that has not been refolded.

BACKBONE ASSIGNMENTS OF LARGE PROTEINS USING 4D TROSY The complete backbone assignments of a number of systems of high-mol wt but in which the spectral complexity is moderate have been reported in the past several years. Included in this list is the 110 kDa homo-octameric protein, 7,8-dihydroneopterin aldolase (27), as well as a number of 8-stranded β-barrel membrane proteins, OmpX (36, 37), OmpA (38), and PagP (22) dissolved in lipid detergents, with effective mol wt in the 50–60 kDa range. In the case of MSG, both the high-mol wt and the large number of residues (723, correlation time of 37 ns at 37°C) necessitated the use of four-dimensional (4D) TROSY (26). In the design of these experiments, special care has been taken to optimize sensitivity. When one considers that a 1 mM solution of a 100 kDa monomeric protein consists of ~10% by weight protein, it is clear that for many systems it will be necessary to record spectra at concentrations significantly below 1 mM. Concentrations ranging from 0.5–0.9 mM were commonly used for MSG. Low concentrations, fast relaxing signals, and peak overlap contribute to the challenges associated with studies of large monomeric proteins.

A suite of 4D assignment experiments has been developed over the past several years to address the issues mentioned above. Initially a pair of 4D TROSY experiments were described; these included the 4D TROSY-HNCACO (providing correlations of the form [$\omega_{C\alpha}(i),\omega_{CO}(i),\omega_{N}(i),\omega_{HN}(i)$] and in some cases [$\omega_{C\alpha}$(i-1),$\omega_{CO}$(i-1),$\omega_{N}(i),\omega_{HN}(i)$]) and the 4D TROSY-HNCOCA (providing correlations of the form [$\omega_{C\alpha}$(i-1),ω_{CO}(i-1),$\omega_{N}(i),\omega_{HN}(i)$]) (39). Backbone chemical shifts are obtained by matching either $^{13}C^{\alpha}$ -^{13}CO or ^{1}HN-^{15}N chemical shift pairs for different residues from these 4D data sets, as described in some detail below. Approximately 95% of the expected intra- and inter-residue correlations were obtained in HNCACO and HNCOCA data sets, respectively, of maltose binding protein at 5°C, where the overall tumbling time of the protein is 46 ns (39). Later, these experiments were supplemented by the introduction of the 4D TROSY-HNCO$_{i-1}$CA$_{i}$ (providing correlations [$\omega_{C\alpha}(i),\omega_{CO}$(i-1),$\omega_{N}(i),\omega_{HN}(i)$] and in some cases [$\omega_{C\alpha}$(i-1),$\omega_{CO}$(i-1),$\omega_{N}(i),\omega_{HN}(i)$]) (40) and by

the use of the 4D $^{15}N,^{15}N$-edited NOESY (41, 42), so that ambiguities arising in the course of the assignments could be resolved (see below). Backbone assignments of a dimeric form of the human tumor suppressor protein p53 (67 kDa homodimer, 279 residues) have been reported (28), utilizing this array of four 4D data sets and a pair of 3D TROSY data sets that correlate backbone amide and side chain $^{13}C^{\beta}$ chemical shifts (39, 43). These early studies demonstrated the utility of 4D-TROSY for the investigation of proteins with moderate numbers of residues. However, the full potential of this methodology is realized only in applications to systems of increasing complexity, such as MSG. Below we explain in some detail the process of resonance assignments in MSG using 4D TROSY NMR experiments (26).

Figure 2 illustrates the procedure for backbone resonance assignments for the segment of polypeptide chain in MSG extending from the amide of Ala 33 to the amide of Phe 35 (Figure 2*a*). Prior to the start of the backbone assignment process, it is useful to simplify the ^{1}HN-^{15}N correlation map by recording a modified 2D HN(CACB) experiment in which only amide correlations of Ala and those residues immediately following Ala are selected (26). The distinct Ala $^{13}C^{\beta}$ chemical shift makes this selection particularly straightforward. Alanine is one of the most abundant amino acids in protein sequences (the most abundant in MSG, 10.1%) and serves as a good starting point for resonance assignments. Figure 2*b* shows a region of such a spectrum in which correlations from only Ala residues (or, rarely, residues following Ala) are observed.

Figure 2*c–i* illustrates how the family of 4D data sets are used for sequential assignment. Starting from Ala 33 (shown in red in Figure 2*b*), a $^{13}C^{\alpha}-^{13}CO$ slice of the 4D HNCACO at the amide ^{1}HN and ^{15}N chemical shifts of Ala 33 (Figure 2*c*) gives the assignments of the $^{13}C^{\alpha}$ (55.0 ppm) and ^{13}CO(181.4 ppm) shifts of this residue because the ($^{1}HN,^{15}N$) correlation of Ala 33 is well resolved in the ^{1}HN-^{15}N HSQC spectrum. A weaker crosspeak in this plane (Figure 2*c*) appears at $^{13}C^{\alpha}/^{13}CO$ chemical shifts of Ala 32. Next, the ($^{1}HN,^{15}N$) slice of the 4D HNCOCA at $^{13}C^{\alpha}$(55.0 ppm)/^{13}CO(181.4 ppm) corresponding to the shifts of Ala 33 (Figure 2*d*) allows the immediate assignment of the $^{1}HN/^{15}N$ chemical shifts of the subsequent residue—Ala 34 (shown in blue)—because the corresponding crosspeak is unique in this slice of the 4D HNCOCA. Figure 2*e–g* shows slices from 4D data sets at the ^{1}HN(8.07 ppm)/^{15}N(121.8 ppm) shifts of Ala 34. In this case, the 4D HNCACO spectrum alone is not sufficient for the assignment of $^{13}C^{\alpha}/^{13}CO$ shifts of Ala 34 as there are five residues in MSG with approximately the same pairs of backbone amide shifts (five peaks in Figure 2*g*). To choose the right correlation out of the five possibilities, 4D HNCOCA (Figure 2*e*) and 4D $HNCO_{i-1}CA_i$ (Figure 2*f*) data sets are used. The $^{13}C^{\alpha}/^{13}CO$ chemical shifts of Ala 33 are known, so the $^{13}C^{\alpha}$ shift of Ala 34 can be obtained by identifying a peak in the 4D $HNCO_{i-1}CA_i$ spectrum at the same ^{13}CO position as that of Ala 33 and a $^{13}C^{\alpha}$ position different from that of Ala 33 (such a crosspeak is shown in blue in Figure 2*f*). The assignment of the ^{13}CO of Ala 34 is then immediately obtained from the HNCACO slice (Figure 2*g*), thereby resolving the

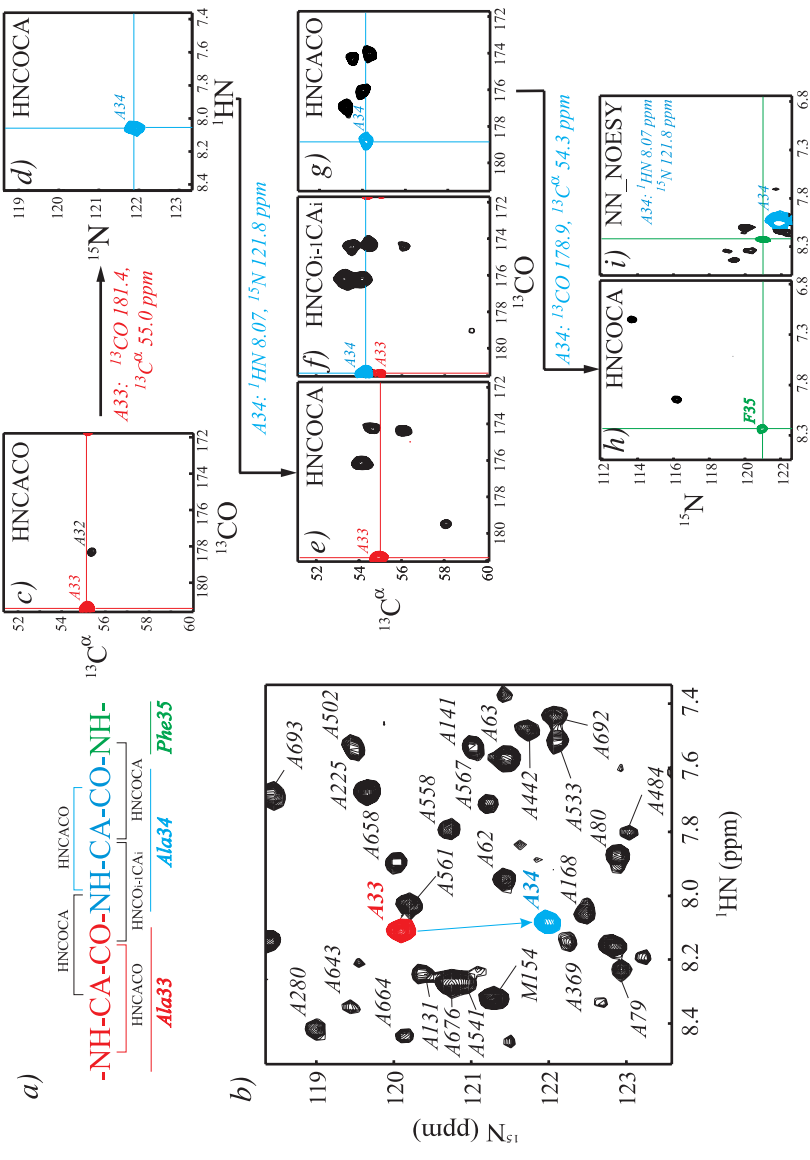


Figure 2 Sequential backbone assignments of the Ala 33-Phe 35 segment of MSG using 4D TROSY. Adapted from Tugarinov et al. (26).

fivefold degeneracy described above. In the absence of chemical shift degeneracies, the $^{13}C^{\alpha}$(54.3 ppm)/^{13}CO(178.9 ppm) chemical shifts of Ala 34 are sufficient to assign the amide shifts of Phe 35 from the 4D HNCOCA, Figure 2*h*. However, because of the threefold degeneracy of this pair of carbon shifts (three crosspeaks in Figure 2*h*), the correct assignment can only be made with additional information. In this case, the ambiguity can be resolved using ^{1}HN-^{1}HN NOEs from a 4D ^{15}N,^{15}N-edited NOESY data set (Figure 2*i*). The correct ^{1}HN,^{15}N assignment for Phe 35 is shown in green in Figure 2*h*, which coincides with the position of the NOE crosspeak in Figure 2*i* marked with green crosshairs. Generally, a high degree of degeneracy among $^{13}C^{\alpha}$/^{13}CO pairs of chemical shifts is observed, and the 4D NOESY data set is very helpful in these cases. In fact, a recent study indicates that NOE information may be essential for assignment of certain residues in MSG (44). Typically, in β-sheets and loops, sequential ^{1}HN-^{1}HN NOE crosspeaks cannot be detected, and in cases of chemical shift degeneracy, assignments are made based on the connectivities from 3D TROSY experiments that correlate amide shifts with inter- or intraresidue $^{13}C^{\beta}$ shifts. The residues of MSG were connected sequentially, as described above, starting from Ala or Gly until another residue of an unambiguous type (one of Ala, Gly, Ser, or Thr) was reached. The connected stretches, e.g., Ala(Gly)-$(X)_n$-Ala, were positioned in the MSG sequence taking into account residue-type information of all the intervening residues X based on $^{13}C^{\beta}$ chemical shifts.

In favorable circumstances, a single sample of the refolded [U-^{15}N,^{13}C,^{2}H]-labeled protein is sufficient to record all NMR experiments mentioned in this section. The redundancy of information available from the combination of 4D and 3D data sets ensures a high degree of reliability of the resulting assignments.

COVALENT MODIFICATIONS AND SECONDARY STRUCTURE FROM HETERONUCLEAR NMR DATA Deviations of $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and ^{13}CO chemical shifts from mean random coil values corrected for deuterium isotope shifts can be used to predict the secondary structure of a protein on the basis of the chemical shift index of Wishart & Sykes (45) or using the database approach of Bax and coworkers (46). In the case of MSG, the NMR-based secondary structure prediction was in good agreement with that derived from the X-ray structure (26). In addition to secondary structure, which emerges from the initial assignments, it is also possible in some cases to observe rare covalent modifications of the polypeptide chain. Figure 3*a* illustrates an isoaspartyl linkage resulting from the deamidation of the side chain of an Asn residue via a cyclic imide intermediate (30). In Figure 3*b*, strips from 3D TROSY-HNCACB (correlations primarily of the form $[\omega_{C\alpha}(i)/\omega_{C\beta}(i),\omega_N(i),\omega_{HN}(i)]$) and 3D TROSY-HN(CO)CACB ($[\omega_{C\alpha}(i\text{-}1)/\omega_{C\beta}(i\text{-}1),\omega_N(i),\omega_{HN}(i)]$) data sets have been used to distinguish between the expected -Asn(305)-Gly(306)- connectivity and the observed -IsoAsp(305)-Gly(306)- connection. Crosspeaks involving $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ nuclei are of opposite signs in these spectra, illustrated in the HNCACB strip plot at the amide shifts of Asn 305, with the crosspeak of the carbon directly attached to the carbonyl carbon

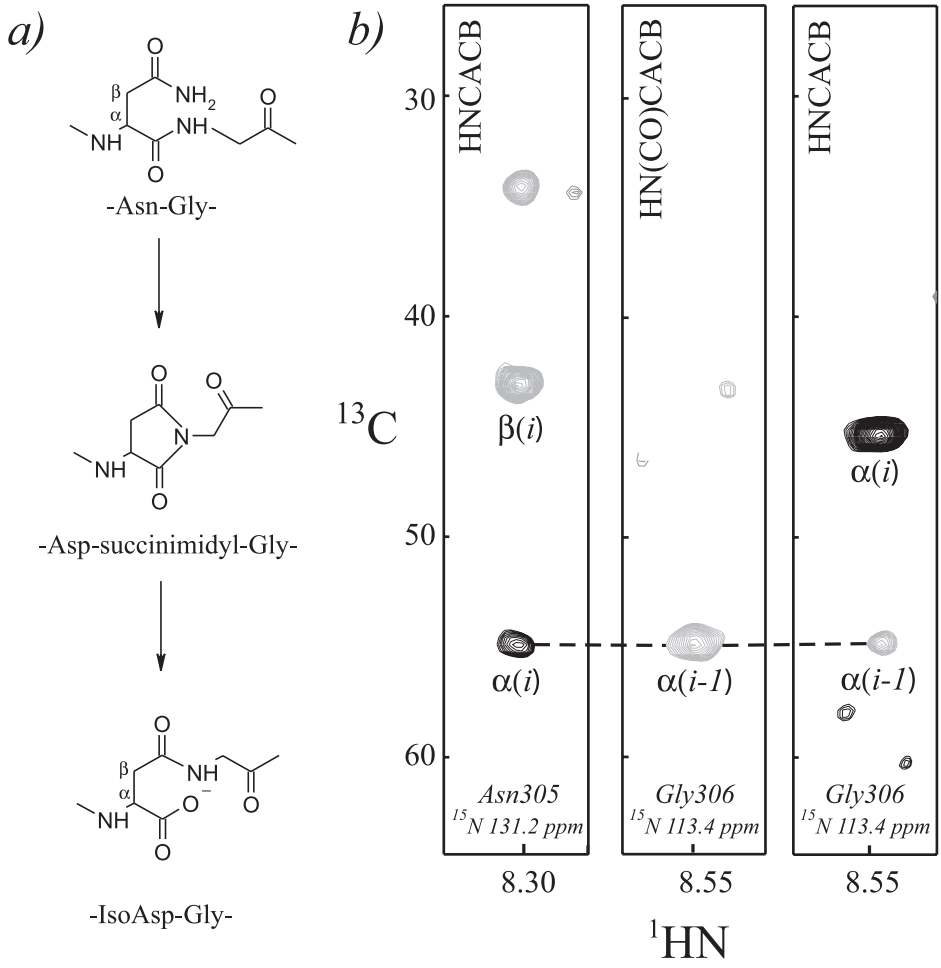


Figure 3 (*a*) The deamidation pathway of the Asn 305 side chain in MSG resulting in the isoaspartyl linkage between Asn 305 and Gly 306. (*b*) Slices from 3D HNCACB and HN(CO)CACB spectra extracted at ^{1}HN and ^{15}N chemical shifts of Asn 305 and Gly 306. Negative peaks are shown with gray contours. Adapted from Tugarinov et al. (26).

($^{13}C^{\alpha}$, for a typical peptide bond) plotted positive (in all of the spectra in the Figure). The compelling evidence that the regular -Asn-Gly- linkage does not exist comes from the negative sign of the crosspeak in the HN(CO)CACB spectrum at the ^{1}HN,^{15}N shifts of Gly306 and the $^{13}C^{\alpha}$ chemical shift of Asn 305. Conversely, the data of Figure 3*b* are consistent with the -IsoAsp-Gly- structure. Iso-linkages occur predominantly in Asn-Gly pairs (30). Earlier, Chazin et al. (47) used ^{1}H NMR spectroscopy to confirm one such linkage in calbindin D_{9k}. Of special interest is the fact that the isoaspartyl linkage in MSG was identified

exactly in the middle of a flexible stretch (residues 300–310) for which coordinates could not be determined by X-ray crystallography (21). Two β-strands are predicted in this region from the chemical shift index, with the Asn 305-Gly 306 pair possibly forming a flexible β-hairpin connection between them. The unusual flexibility of this segment is supported by ^{15}N relaxation data (26, 48) and is likely to be a prerequisite for the formation of the cyclic imide intermediate.

Assignments of Side Chains in Large Proteins

The success of NMR methodology for assignment of backbone spins in proteins in the 100 kDa range has encouraged the development of new methods for assignment of side chains (49–51). A main strategy in our laboratory has been to focus on assignments of protonated methyl groups, specifically Ile(δ1), Leu, and Val methyls, in otherwise perdeuterated proteins. This labeling pattern preserves many of the important features of perdeuteration with respect to sensitivity gains in spectra, while maintaining a critical number of protons at important side chain positions for further studies of protein structure (52–54) and dynamics (55, 56). ^{1}H-^{13}C correlation spectra of methyls are, in general, of high quality even in applications involving high-mol-wt systems (57), because many of the transitions that contribute to methyl crosspeaks are long lived (58) and because the intensity of each correlation is derived from three degenerate methyl protons. The strategies for biosynthetic incorporation of protons into Ile(δ1), Leu, and Val methyl groups of perdeuterated proteins are well established (52, 53). A concise overview of these biosynthetic methods is given below.

THE IMPORTANCE OF α-KETOACIDS AS BIOSYNTHETIC PRECURSORS FOR METHYL-CONTAINING RESIDUES IN PROTEINS It has been long known that certain α-ketoacids can serve as biosynthetic precursors of a number of methyl-containing amino acids (Ala, Ile, Leu, and Val) in proteins overexpressed in minimal media (59). The list of α-ketoacids that are routinely used in our laboratory for a variety of applications is shown in Figure 4. Addition of α-ketoisovalerate (compound I) to [D_2O, ^{15}N, U-^{13}C,^{2}H-glucose]-based minimal media for protein expression in *Escherichia coli* leads to the production of uniformly ^{15}N,^{13}C, highly deuterated, Leu, Val methyl-protonated proteins (53), whereas the addition of α-ketobutyrate (compound II) generates proteins with methyl protonation restricted to the Ile δ1 position (52). Typically, both α-ketobutyrate and α-ketoisovalerate are added to growth media to generate highly deuterated proteins with protonation at methyl positions of Ile (δ1), Leu, and Val.

In addition to selective methyl protonation in deuterated, uniformly ^{15}N,^{13}C-labeled proteins, different labeling patterns involving side chains of Ile, Leu, and Val are possible by using some of the other precursors shown in Figure 4. In an effort to produce proteins with ^{13}C only at methyl sites in a cost-effective manner for high throughput screening of ligands, the Abbott group (60) has developed synthetic methods for production of ^{13}C methyl-labeled α-ketobutyric and

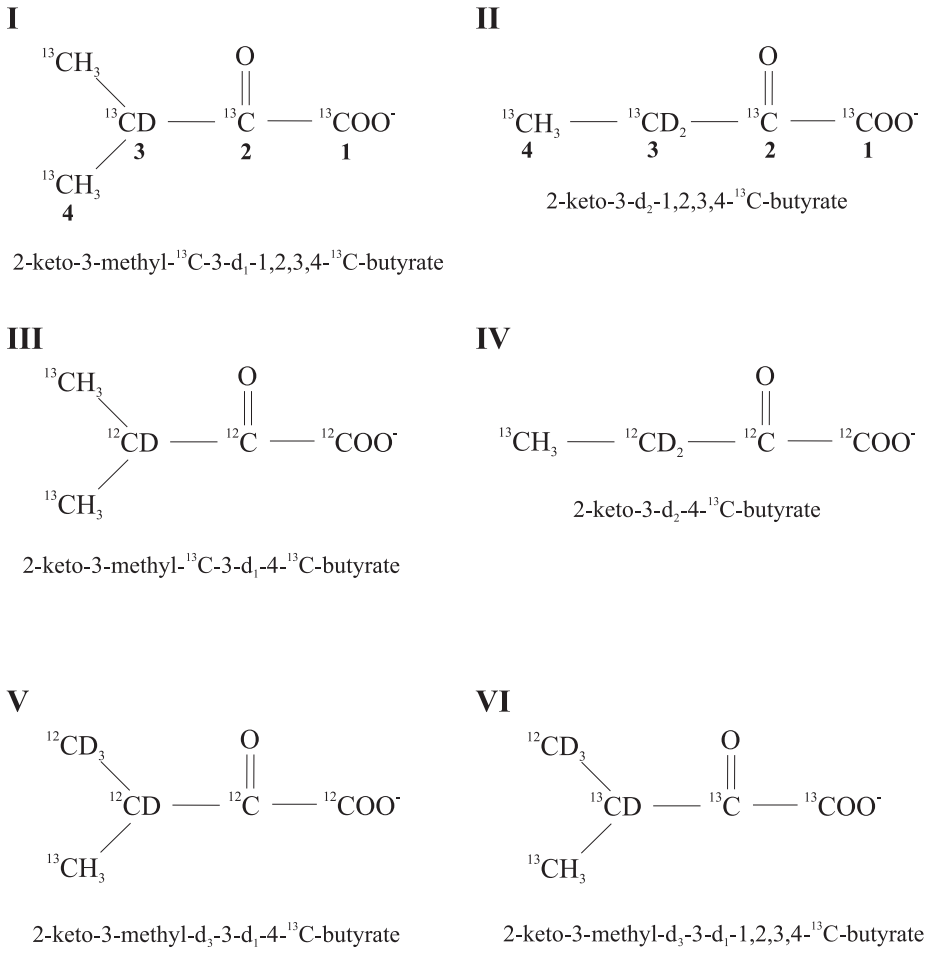


Figure 4 Chemical formulae of the set of isotopically labeled α-ketoacids used as biosynthetic precursors for methyl-containing residues in proteins. Sodium salts of α-ketobutyric and α-ketoisovaleric acids protonated at the methyl and the 3-positions are commercially available and can be quantitatively exchanged to 3-^{2}H at high pH in D_2O (53). These precursors are added in the amounts of ~50 mg (butyric acid) and ~80 mg (valeric acid) per liter of [D_2O; U-^{13}C,^{2}H-glucose]-based growth media ~1 h prior to induction of protein overexpression to produce Ile (δ1 only), Leu, and Val methyl protonated-[^{13}C,^{2}H] labeled proteins (53).

α-ketoisovaleric acids, compounds III and IV in Figure 4. An alternative and apparently even more cost-effective synthetic strategy for production of α-ketoacids with this labeling pattern using Grignard chemistry has been proposed recently by Wagner and coworkers (61). Our laboratory has demonstrated that for

assignments of methyls in MSG a labeling strategy that makes use of compound VI in Figure 4 is particularly advantageous. Below we describe in detail optimal labeling strategies and recently developed NMR experiments for assignments of methyl groups in MSG, because standard approaches that have worked so well for smaller systems failed in this case.

ASSIGNMENTS OF Ile, Leu, AND Val METHYL GROUPS USING COSY-BASED HN-DETECTED NMR EXPERIMENTS The commonly used NMR approach for assignment of Ile, Leu, and Val methyl ^{13}C and ^{1}H resonances involves a transfer of magnetization originating on methyl groups to the backbone $^{13}C^{\alpha}$ carbon using homonuclear ^{13}C-^{13}C Hartmann-Hahn mixing schemes (62). Subsequently the magnetization is relayed through carbonyl spins to the amide nitrogens and protons for detection (63–67). Because the backbone spins are already assigned, correlation of side chain and backbone chemical shifts immediately leads to the assignment of the side chain ^{13}C and ^{1}H spins. These experiments have been used for assignment of methyl groups in a highly deuterated maltose binding protein (MBP, 42 kDa) sample at 600 MHz and 37°C (25) and for assignment of methyls in a large number of smaller systems. Recently, ^{1}HN-^{15}N TROSY versions of these experiments were used by Wüthrich and coworkers (68) to assign methyl groups of the membrane protein OmpX in micelles (effective mol wt ~60 kDa). For larger systems and at higher magnetic field strengths, experiments that avoid the transfer through fast-relaxing carbonyl nuclei are attractive alternatives. However, the branch points in each of the side chain carbon scaffolds of Ile, Leu, and Val degrade the efficiency of $^{13}C^{m} \rightarrow {}^{13}C^{\alpha}$ magnetization transfer. For example, consider Ile. Signal originating on $^{13}C^{m}$ is transferred up the side chain. At the $^{13}C^{\beta}$ carbon, magnetization can go in one of two ways, $^{13}C^{\beta} \rightarrow {}^{13}C^{\alpha}$ or $^{13}C^{\beta} \rightarrow {}^{13}C^{\gamma 2}$. The former transfer is productive in the sense that only the signal transferred to $^{13}C^{\alpha}$ will ultimately be detected, whereas the later transfer siphons off the signal. This bifurcation leads to a reduction in sensitivity of almost a factor of 2.5, which is unacceptable for applications to high-mol-wt proteins where signal to noise is limited in the first place. With this in mind, our laboratory has designed COSY-based experiments that transfer magnetization down the side chain for assignment of Ile methyl groups in large proteins (49). Each transfer step is sequential in this approach, affording control over the flow of magnetization, unlike the case in TOCSY-transfer. This is possible for Ile because the chemical shifts of all ^{13}C spins at the branch point are well separated so that magnetization transfer from $^{13}C^{\beta}$ to $^{13}C^{\gamma 2}$ (unproductive) can be prevented by application of $^{13}C^{\gamma 2}$-selective 180° pulses (49). The efficacy of this approach was demonstrated on a sample of MBP at 5°C (correlation time of 46 ns, equivalent to ~100 kDa protein at 37°C) for which both the resolution and sensitivity were considerably improved over the more traditional TOCSY-based experiment (49).

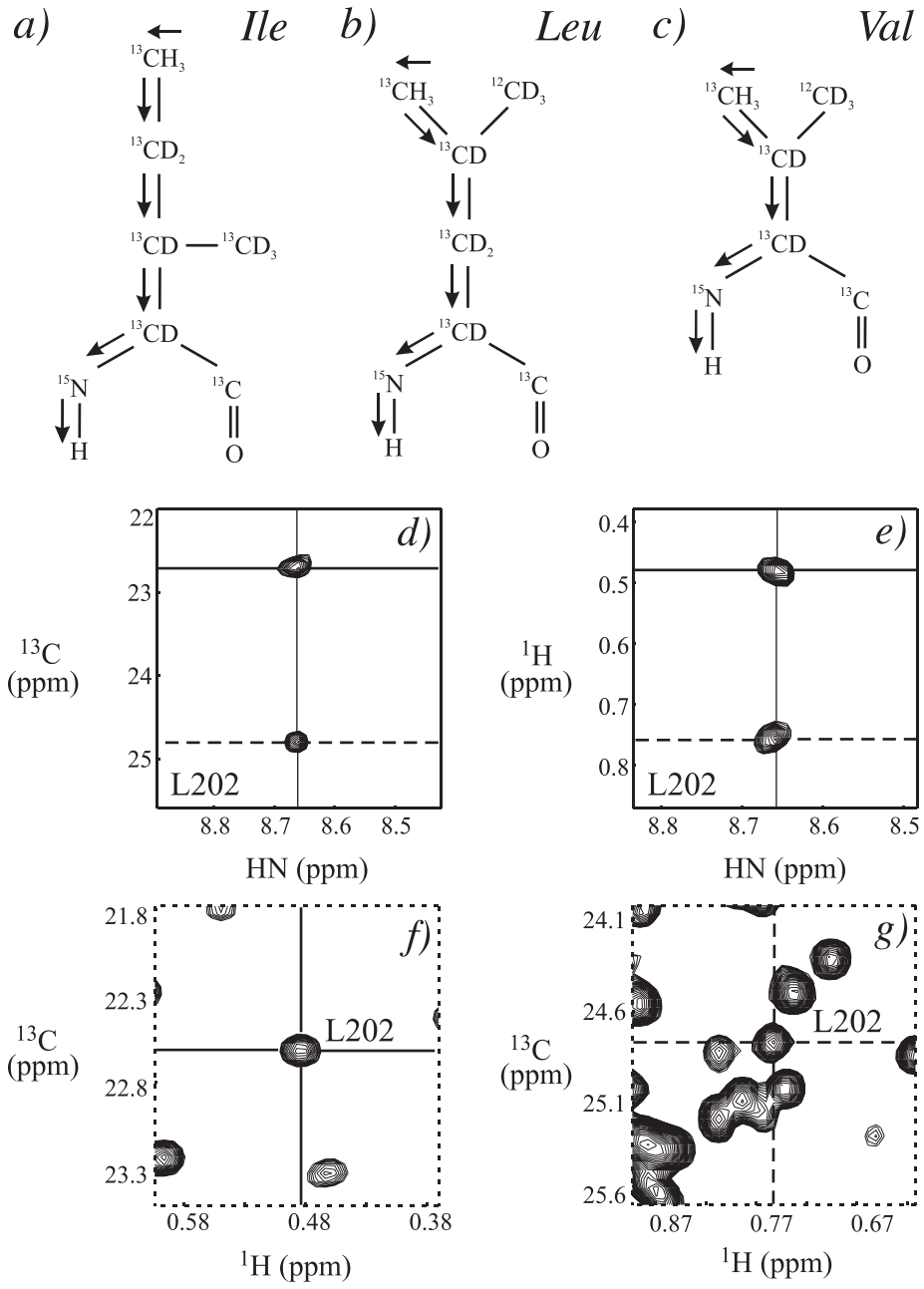
Unfortunately, the approach described for Ile is not possible for Val and Leu side chains in which pro-R and pro-S methyls (and Leu $^{13}C^{\gamma}$ spins) all resonate at similar frequencies. We have, therefore, modified the labeling protocol so that

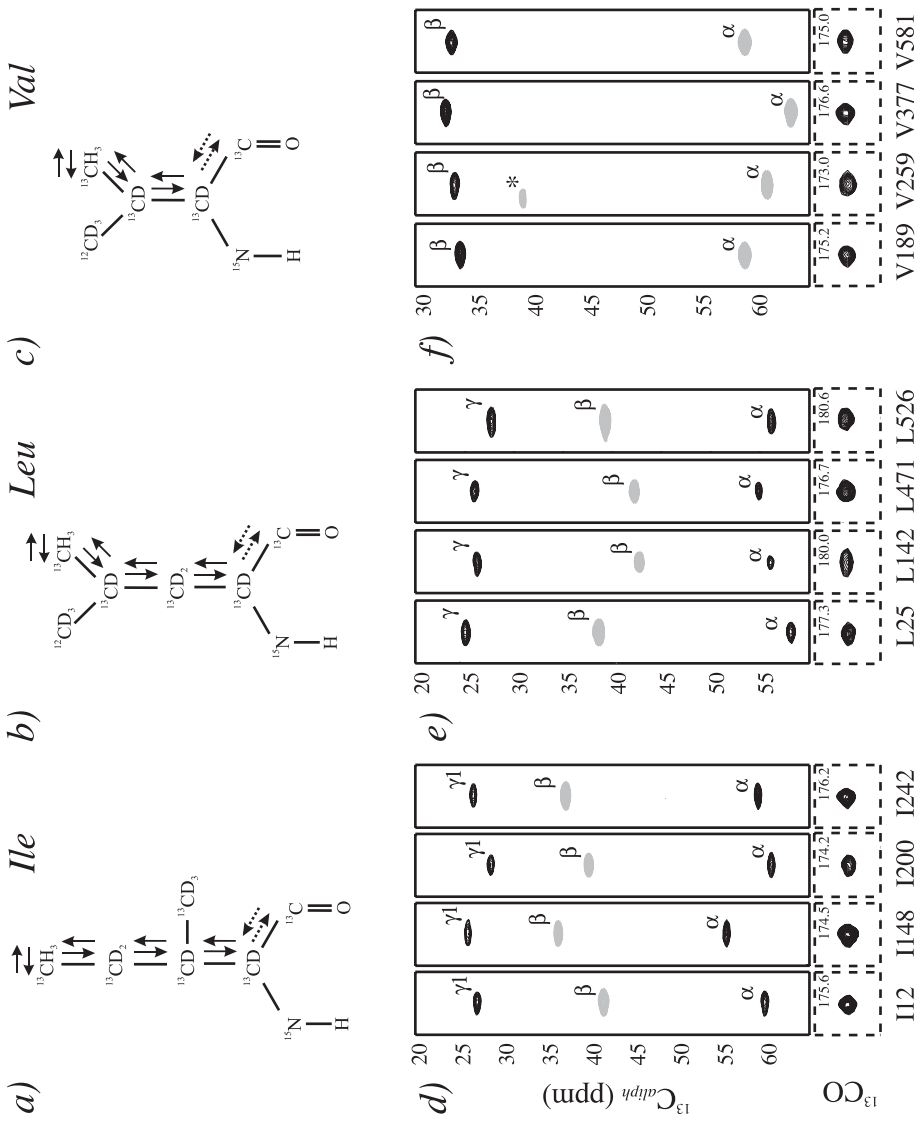
proteins are produced with Val and Leu residues ^{13}C-enriched at only a single methyl site. This labeling strategy involves the use of α-ketoisovalerate ^{13}C-labeled (nonstereospecifically) at one methyl, while the other is of the $^{12}CD_3$ variety (compound VI in Figure 4) to generate a linear ^{13}C spin system for these residues, effectively avoiding the problem associated with the branch point, described above. The gain in sensitivity that results outweighs the losses associated with the effective twofold dilution of the methyl groups (50). A methyl protonated {I(δ1 only), L($^{13}CH_3$,$^{12}CD_3$), V($^{13}CH_3$,$^{12}CD_3$)} U-[^{15}N,^{13}C,^{2}H] sample of MSG was produced using U-[^{13}C]-4-[^{1}H] α-ketobutyrate (compound II) and [$^{13}CH_3$/$^{12}CD_3$]-α-ketoisovalerate (compound VI) (50).

Figure 5*a–c* illustrates the magnetization flow in the array of COSY-based HN-detected NMR experiments that have been recorded on MSG at 37°C (276 protonated methyl groups). Three-dimensional data sets containing correlations of the form [$\omega_{Cm}(i),\omega_N(i),\omega_{HN}(i)$] and [$\omega_{Hm}(i),\omega_N(i),\omega_{HN}(i)$] are generated (50). Figure 5*d,e* shows planes from these 3D data sets, extracted at the amide ^{15}N shift of Leu 202 and used to assign the ^{13}C and ^{1}H methyl chemical shifts of this residue. In order to ascertain which ^{1}H shift is to be paired with which ^{13}C, a high-resolution ^{1}H-^{13}C 2D correlation map is employed, Figure 5*f,g*, showing clearly that only a single ^{1}H/^{13}C combination is possible. Using this set of experiments 90%, 55%, and 88% of the Ile, Leu, and Val methyls could be assigned in MSG (50).

METHYL-DETECTED OUT-AND-BACK EXPERIMENTS FOR ASSIGNMENT OF Ile, Leu, AND Val METHYLS IN LARGE PROTEINS A drawback of the HN-detected experiments described above is that the net magnetization that can be transferred from the methyl to the amide site is effectively that associated with a single proton (rather than the reservoir from all 3 of the methyl protons) (50). As an alternative, "out-and-back" experiments can be employed in which signal both originates and is detected on the methyl groups, circumventing losses associated with net transfer to amides (50). The magnetization transfer steps in this class of

Figure 5 (*a-c*) Schematic diagrams of the magnetization flow in the HN-detected experiments for assignments of Ile(δ1), Leu and Val methyls (note that the methyls in Leu and Val are of the $^{13}CH_3$/$^{12}CD_3$ variety). (*d,e*) $^{13}C^m$-HN (correlations of the form [$\omega_{Cm}(i),\omega_N(i),\omega_{HN}(i)$]) and $^{1}H^m$-HN (correlations of the form [$\omega_{Hm}(i),\omega_N(i),\omega_{HN}(i)$]) slices at the ^{15}N chemical shift of Leu 202 from HN-COSY experiments recorded on U-[^{2}H,^{15}N,^{13}C], [$^{13}CH_3$]-Ile(δ1) [$^{13}CH_3$/$^{12}CD_3$]-Leu,Val MSG, 800 MHz, 37°C. (*f,g*) regions of a 2D ^{1}H-^{13}C constant time (CT) HMQC spectrum showing the corresponding crosspeaks of Leu 202. $^{13}C^m$ and $^{1}H^m$ chemical shifts from 3D spectra in panels (*d,e*) indicated with solid and dashed lines correspond to peak positions in the CT-HMQC shown with solid and dashed lines, respectively, illustrating how proton-carbon connectivities can be established from a 2D ^{1}H-^{13}C correlation map. Adapted from Tugarinov & Kay (50).


a)
Ile
b)
Leu
c)
Val
d)
e)
f)
g)
L202
HN (ppm)
^{13}C (ppm)
^{1}H (ppm)


a)
Ile
b)
Leu
c)
Val
d)
e)
f)
$^{13}C_{aliph}$ (ppm)
^{13}CO
I12 I148 I200 I242
L25 L142 L471 L526
V189 V259 V377 V581

experiment are illustrated schematically in Figure 6*a–c*. A single 3D data set with correlations of the form [$\omega_{C\gamma,C\beta,C\alpha}(i)$, $\omega_{Cm}(i)$, $\omega_{Hm}(i)$] for Ile and Leu and [$\omega_{C\beta,C\alpha}(i)$, $\omega_{Cm}(i)$, $\omega_{Hm}(i)$] for Val is obtained. Additional experiments can be performed to relay signal to the ^{13}CO spin and back, providing correlations of the form [$\omega_{CO}(i)$, $\omega_{Cm}(i)$, $\omega_{Hm}(i)$]. As with the HN-detected experiments, these schemes benefit from Leu and Val-[$^{13}CH_3/^{12}CD_3$] isotopic labeling and from using selective pulses to prevent the transfer of magnetization from $^{13}C^{\beta}$ to $^{13}C^{\gamma 2}$ in Ile (50). On average, sensitivity gains ranging from factors of 5 to 10 relative to the HN-detected data sets were observed for MSG (50).

Figure 6*d–f* shows $^{1}H_m$-$^{13}C_{aliph}$ and $^{1}H_m$-^{13}CO strips from the data sets described above for selected residues of MSG. The sequence-specific assignments of Ile, Leu, and Val methyls can be obtained by matching three ^{13}C frequencies ($^{13}C^{\alpha}$, $^{13}C^{\beta}$, and ^{13}CO) to those available from the compiled list of backbone and $^{13}C^{\beta}$ assignments. Note that, unlike HN-detected data sets, the methyl-detected experiments inherently contain information on ^{1}H-^{13}C connectivities within a given methyl group, simplifying the assignment process.

The high sensitivity of the methyl-detected experiments significantly reduces the requirements for both NMR instrument time and long-term protein stability. Less than a week of machine time was required to record all of these experiments on a 0.9 mM sample of MSG, whereas close to three weeks of NMR time was employed for the HN-detected experiments. On the basis of data from the methyl-detected data sets exclusively, 95% of Ile δ1, 64% of Leu, and 93% of Val methyls could be assigned. Degeneracy in Leu $^{13}C^{\alpha}/^{13}C^{\beta}$ chemical shift pairs, lower sensitivity of the CO experiment, and strong scalar coupling between $^{13}C^{\delta}$ and $^{13}C^{\gamma}$ nuclei in 15 Leu spin-systems precluded full assignment of Leu methyls. Combining the connectivities obtained in both methyl- and HN-data sets gave assignments for 95%, 88%, and 99% of Ile, Leu, and Val residues with a high degree of confidence. Finally, a 3D [^{13}C-F_1, ^{13}C-F_2]-edited NOESY spectrum of a {L($^{13}CH_3$,$^{13}CH_3$), V($^{13}CH_3$,$^{13}CH_3$)} U-[^{15}N,^{2}H] MSG sample (mixing time of 30 ms) was used to obtain intramethyl NOE correlations. In addition to confirming our assignments, this data set was used to assign Leu methyls in cases for which assignment of only one of the two methyls was available via the through-bond experiments described above, owing to strong coupling effects for example. In this way, the number of Leu methyls that was assigned was increased to 91%.

←

Figure 6 (*a-c*) Schematic diagrams of the magnetization transfer steps employed in the methyl-detected experiments recorded on U-[^{2}H,^{15}N,^{13}C], [$^{13}CH_3$]-Ile(δ1) [$^{13}CH_3/^{12}CD_3$]-Leu,Val MSG. (*d-f*) Selected $^{1}H^{m}$-$^{13}C^{aliph}$ strips, with correlations at [$\omega_{C\gamma,C\beta,C\alpha}(i)$, $\omega_{Cm}(i)$, $\omega_{Hm}(i)$] (upper row, solid frame) and $^{1}H^{m}$-^{13}CO strips, with correlations at [$\omega_{CO}(i)$, $\omega_{Cm}(i)$, $\omega_{Hm}(i)$] (dashed-lined frame) for Ile (*d*), Leu (*e*) and Val (*f*). Negative peaks ($^{13}C^{\beta}$ for Ile and Leu; $^{13}C^{\alpha}$ for Val) are shown in gray. The peak labeled * derives from another spin system and is more intense on another slice. Adapted from Tugarinov & Kay (50).

STRUCTURAL AND DYNAMIC INFORMATION FROM NMR STUDIES OF LARGE PROTEINS

The chemical shifts assigned from the experiments described in the previous sections form the basis for NMR studies of biomolecular structure and dynamics. MSG is a particularly rich system for such studies. A high-resolution X-ray structure of the glyoxylate bound state has been published (21), and thus a structural framework is available, on the one hand, to interpret dynamics and binding data that can be readily obtained from NMR. On the other hand, structures of other liganded states or of the apo state of the enzyme have not been determined, and NMR studies can provide insight into how the domains of the protein change in response to ligand binding. In this context, NMR can be used in concert with the X-ray structure to rapidly generate high-quality models of the protein in its different states. Below, we illustrate a number of applications involving MSG and establish that the same sort of quantitative studies that have to date been reserved for applications involving small proteins can also be performed on larger systems, albeit with some modifications to the existing methodology. Finally, a structural and dynamic study of the membrane enzyme PagP is presented, illustrating that in favorable cases the methodology can also be applied successfully to small membrane proteins.

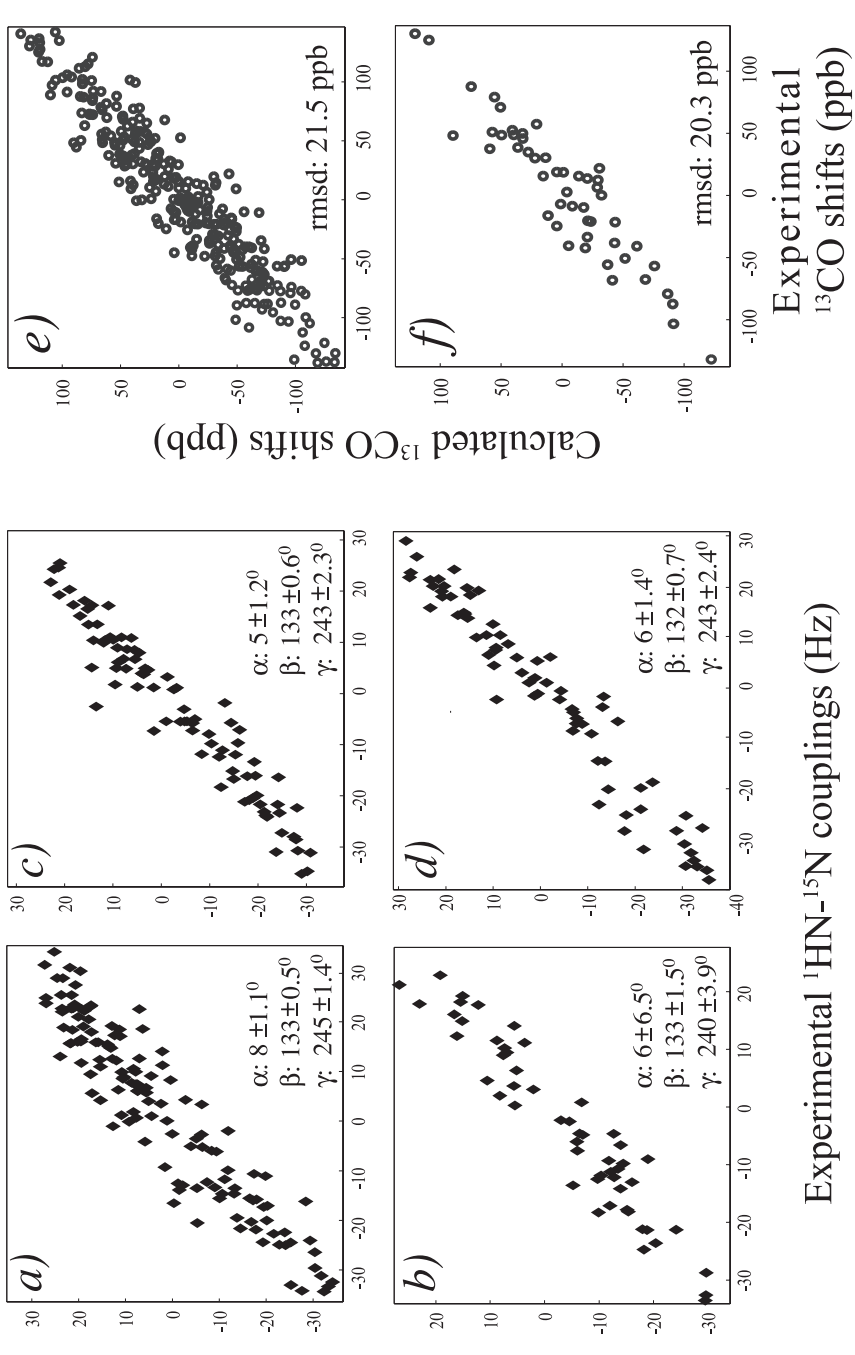
Domain Orientation in MSG from Residual Dipolar Couplings and Chemical Shift Changes Upon Alignment

Many spin interactions that contain useful structural (70, 71) and potentially dynamic information (72) average to zero for molecules that tumble in isotropic solution. The introduction of a small amount of molecular alignment (typically on the order of 0.1%) restores these interactions, at least partially, so that they can be easily measured (3). In this case, the dipolar coupling between pairs of proximal NMR active spins is nonzero, and the magnitude of this coupling is related to the orientation of the vector connecting the spins in a molecular alignment frame (3, 71). This dipolar coupling leads to peak splittings in spectra in a manner analogous to scalar couplings. A large number of experiments, including schemes optimized for applications involving high-mol-wt proteins, such as MSG (73–75), have recently been developed to facilitate the measurement of such couplings. Chemical shift changes upon alignment also provide structural information that is complementary to that derived from dipolar couplings (76–79). Because in both cases information is obtained that relates orientations of bond vectors or axis systems (chemical shift tensors) to a single molecular frame, the couplings or shift changes measured provide particularly powerful structural restraints.

The ^{1}HN-^{15}N dipolar coupling ($^{1}D_{H\text{-}N}$) is by far the largest of those accessible in uniformly deuterated, ^{15}N,^{13}C-labeled proteins, and a total of 415 such couplings ranging from -40 to $+35$ Hz have been measured for the apo form of MSG (48). These couplings were supplemented by 320 ^{13}CO chemical shift changes that occur upon alignment. A number of important technical issues emerge in these measurements concerning the application to large proteins, and the interested reader is referred to the literature (48, 73, 74).

In the case of MSG, our goal was to use these restraints to understand how the domains in the enzyme change orientation in response to ligand binding. In particular, it was believed that the position of the C-terminal domain of the protein might differ in apo (open) and substrate-bound (closed) states of the molecule (21). Earlier circular dichroism studies of both yeast and maize malate synthases indicated that significant conformational changes likely occur in the enzyme upon ligand binding (80, 81). Domain reorientations that occur upon ligand binding in pyruvate kinase (82), a close structural analogue of MSG, have been observed from X-ray studies (83). In addition, dimeric citrate synthase (84) that possesses a totally different fold but catalyzes a similar reaction, undergoes a large conformational change in which its smaller domain rotates from the main body of the dimer by $\sim$18° (21, 84).

It is straightforward to use the dipolar couplings and chemical shifts recorded on the unligated protein to reorient the domains from the X-ray ligated state to the solution apo conformation. Using the X-ray structures of the domains on an individual basis, molecular alignment frames were determined for each domain by minimizing the difference between measured couplings and those calculated on the basis of the structure of the domain. Figure 7 compares calculated couplings obtained using the glyoxylate-bound X-ray structure versus experimental couplings from the apo solution state for each of the four domains in the protein (see Figure 1). The excellent correlation in all cases indicates that the intradomain structures of the glyoxylate-bound and apo forms of the protein are the same. Also included in the Figure are the Euler angles, $\{\alpha, \beta,$ and $\gamma\}$ that effectively describe the transformation from the domain orientation described by the X-ray structure to the solution orientation. Of note, Euler angles from the four domains are equivalent to within experimental error. Therefore, contrary to predictions based on earlier studies with *E. coli* MSG and related proteins, the domain orientations in the apo form are not significantly different from those in the glyoxylate-bound structure determined by X-ray crystallography (21). The correlation between experimentally measured carbonyl chemical shift changes and those calculated from X-ray coordinates of the MSG-glyoxylate complex for the whole molecule (Figure 7*e*) and for the C-terminal plug (Figure 7*f*) confirm the conclusions of the dipolar coupling analysis. Thus, the gross features of the enzyme do not change upon ligand binding; much more subtle structural changes are involved. In an attempt to characterize these changes as well as the kinetics and thermodynamics of ligand binding, we


a)
α: 8 ±1.1°
β: 133±0.5°
γ: 245 ±1.4°
b)
α: 6±6.5°
β: 133 ±1.5°
γ: 240 ±3.9°
c)
α: 5 ±1.2°
β: 133 ±0.6°
γ: 243 ±2.3°
d)
α: 6 ±1.4°
β: 132 ±0.7°
γ: 243 ±2.4°
e)
rmsd: 21.5 ppb
f)
rmsd: 20.3 ppb
Calculated ^{1}HN-^{15}N couplings (Hz)
Experimental ^{1}HN-^{15}N couplings (Hz)
Calculated ^{13}CO shifts (ppb)
Experimental ^{13}CO shifts (ppb)

have measured a variety of NMR parameters that are sensitive to motions spanning a wide range of timescales.

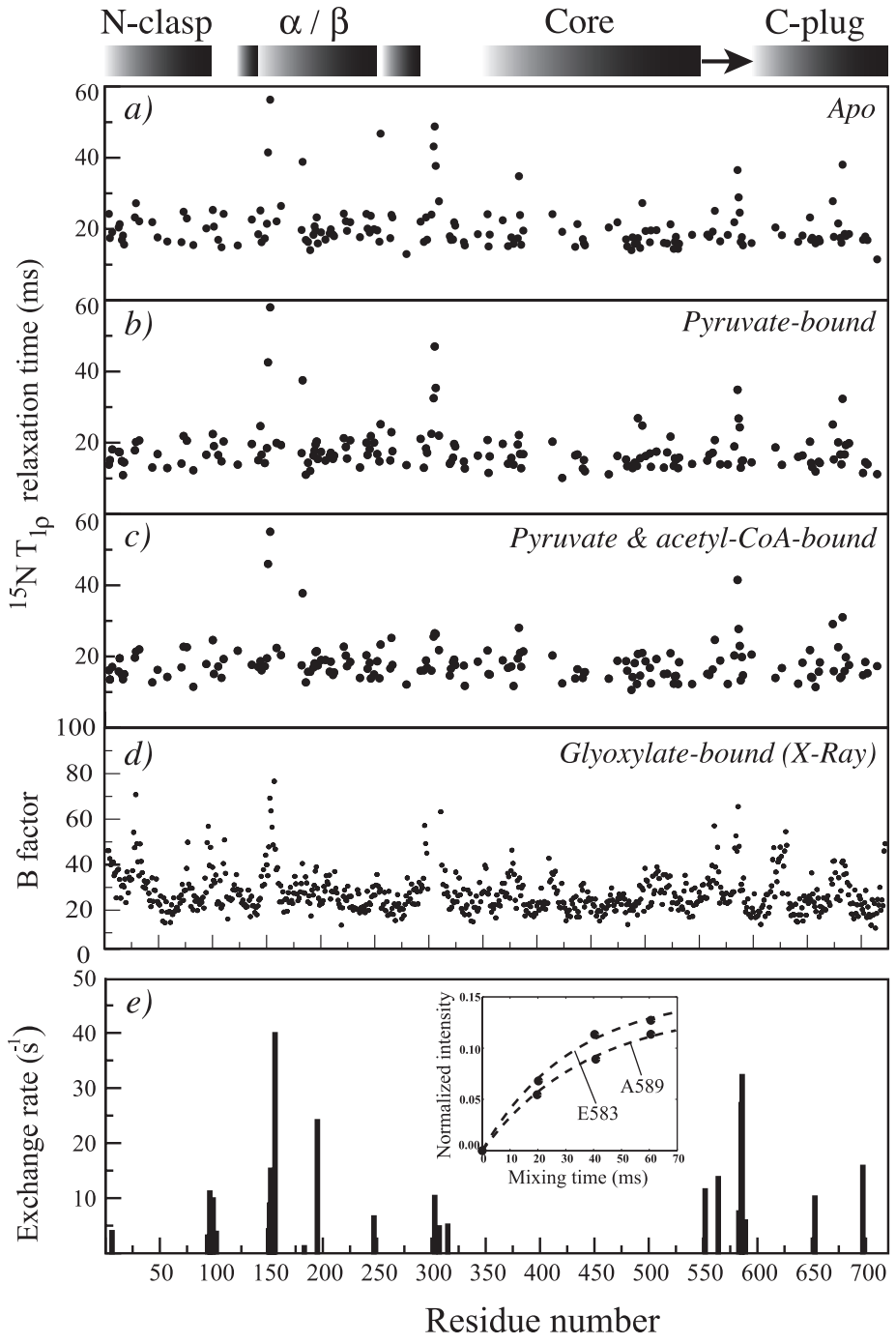
Quantitative NMR Studies of Ligand Binding to MSG

NMR is a particularly powerful probe of ligand-induced structural and dynamical changes in macromolecules, and not surprisingly, large numbers of binding studies have been reported over a period of many years. NMR is also a powerful tool for determining the kinetic parameters of ligand binding not easily accessible by other methods (85). Below we summarize the sort of information that can be obtained from studies of systems as large as MSG.

^{15}N RELAXATION AND HYDROGEN EXCHANGE IN Apo AND LIGAND BOUND FORMS OF MSG

^{15}N spin relaxation measurements of apo MSG establish that the protein tumbles isotropically in solution, with correlation times for each of the domains ranging from 35 to 37 ns (37°C) (26). The X-ray structure shows that all the domains in the glyoxylate-bound form make extensive surface contacts with the core of the molecule, with contact surface areas of ~2300, ~1330, and ~2000 Å^2 between the core and the N-terminal, α/β, and the C-terminal domains, respectively. These close contacts contribute to the overall globular shape of the enzyme and ensure that the molecule tumbles as a single entity, despite the fact that the domains are connected to the core of the protein through flexible linkers. ^{15}N $T_{1\rho}$ values (related inversely to line width) have been measured for well-separated peaks of the protein in the apo, pyruvate-bound (pyruvate serves as a glyoxylate-mimicking inhibitor of the enzymatic reaction), and pyruvate/acetyl-CoA-bound states, and these values are plotted as a function of residue number in Figure 8*a–c*. The relaxation times are homogeneous throughout the sequence for all forms of the protein, with the exception of several residues in loops and linkers (48). Interestingly, the regions that are flexible on the ps-ns timescale in solution are also disordered in the crystal state (Figure 8*d*). In order to establish whether slower processes might be affected by binding, hydrogen exchange with solvent was measured using 3D TROSY-HNCO type (86) exchange experiments (87). The exchange rates of 28 amides could be quantified accurately (rates greater than ~3 s^{-1}) and are shown in Figure 8*e* for the apo state of MSG. In general, regions with high exchange rates correlate well with those having elevated ^{15}N transverse relaxation times. Similar exchange profiles

Figure 7 Experimental versus calculated $^1D_{HN}$ values for the (*a*) core domain, (*b*) N-terminal α-clasp, (*c*) α/β domain, and (*d*) C-terminal plug of MSG. Experimental versus calculated carbonyl shifts upon alignment of MSG for (*e*) the whole molecule and (*f*) residues of the C-terminal plug. The alignment tensor obtained from the full set of $^1D_{HN}$ data was used for the calculation of carbonyl shifts. MSG was aligned using Pf1 phage, ~12 mg/ml (4). Reproduced with permission from Tugarinov & Kay (48).


N-clasp
α / β
Core
C-plug
a)
Apo
b)
Pyruvate-bound
c)
Pyruvate & acetyl-CoA-bound
d)
Glyoxylate-bound (X-Ray)
e)
^{15}N $T_{1\rho}$ relaxation time (ms)
B factor
Exchange rate (s^{-1})
Normalized intensity
A589
E583
Mixing time (ms)
Residue number

are obtained for the various bound states of the protein. The absence of significant changes in hydrogen exchange and ^{15}N relaxation profiles suggests that the changes in dynamics that do accompany binding, like the changes in structure, are minor. Notably, there is not rigidification of the linker between the C-terminal domain and the molecular core, in contrast to predictions in the literature.

KINETICS AND THERMODYNAMICS OF LIGAND BINDING TO MSG Figure 9*a* illustrates a well-resolved section from the ^{1}HN-^{15}N TROSY-HSQC spectrum of apo MSG, with trajectories of peak displacements resulting from the titration of apo MSG with pyruvate, indicated with arrows. The equilibrium dissociation constants (K_d) of various MSG complexes can be determined from the shift changes that accompany ligand binding (48, 88). The best-fit curves in both ^{15}N and ^{1}HN dimensions for Val 620 titrated with pyruvate and glyoxylate are shown in Figure 9*b*,*c*. Estimated K_d values averaged over 20 well-resolved peaks with significant shifts in either ^{1}HN or ^{15}N dimensions are 1.02 ± 0.15 mM and 600 ± 70 μM for pyruvate and glyoxylate, respectively. As expected, glyoxylate, the physiological substrate of MSG, binds stronger than pyruvate.

The binding of glyoxylate or pyruvate occurs in the intermediate exchange regime on the ^{15}N and ^{1}HN chemical shift timescales for many resonances in MSG, as evidenced by considerable line-broadening at intermediate ligand concentrations (Figure 9*a*). It is therefore straightforward to obtain quantitative estimates of the kinetic parameters of ligand binding by fitting peak line shapes as a function of added ligand. The exchange kinetics of binding were determined for both pyruvate and glyoxylate assuming that the binding process obeys a simple second-order reaction,

$$P + L \underset{k_{off}}{\overset{k_{on}}{\longleftrightarrow}} PL,$$

in which P, L, and PL denote free protein, free ligand, and complex, respectively. Figure 9*d*,*e* shows two examples of line-shape simulations from the titration data for glyoxylate (^{1}H dimension of Ile 482) and pyruvate (^{15}N dimension of Val 620). The on rates for pyruvate derived from line-shape analysis vary between

Figure 8 (*a*-*c*) ^{15}N $T_{1\rho}$ relaxation times versus residue number for a subset of 165 residues of MSG as a function of different binding states of the molecule; (*d*) crystallographic B factors versus residue number of the glyoxylate-bound enzyme; (*e*) solvent exchange rates for the amides of apo MSG versus residue number. The inset shows the build-up of peak intensities of Glu 583 and Ala 589 as a function of mixing time. Adapted from Tugarinov & Kay (48).

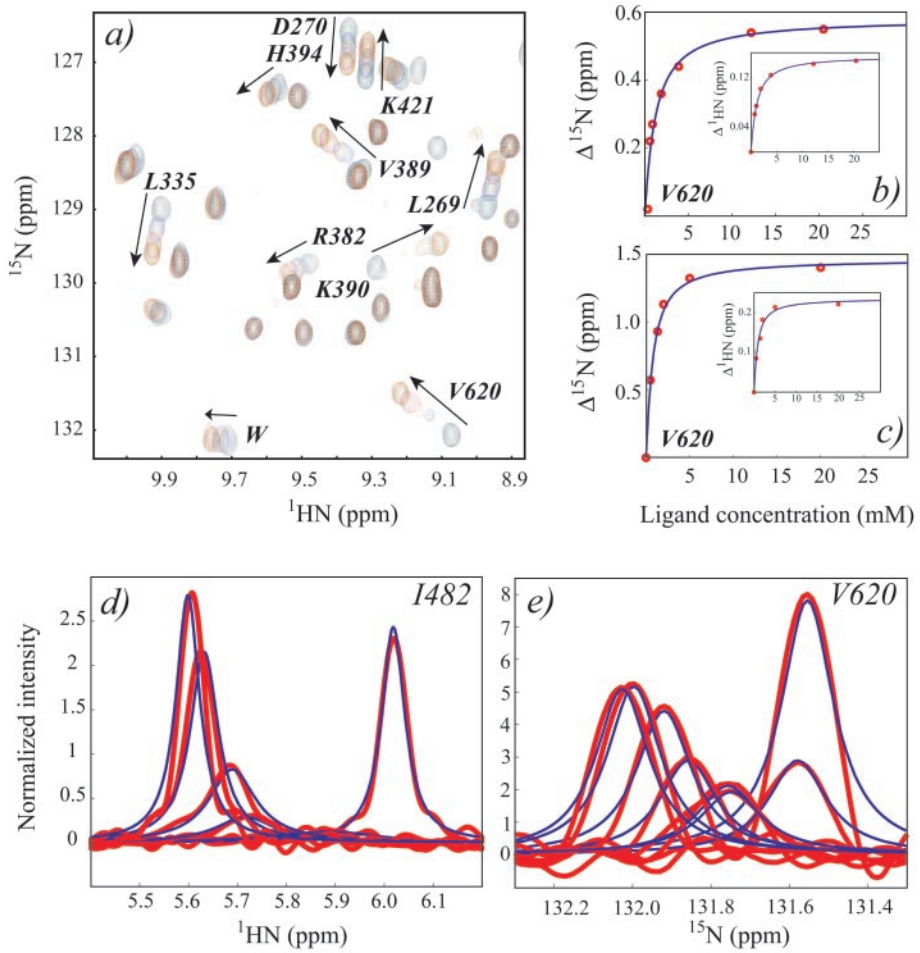


Figure 9 (*a*) Titration of [^{15}N,^{2}H]-labeled MSG with pyruvate. A selected region of the ^{1}HN-^{15}N TROSY-HSQC is shown. (*b*,*c*) K_d determination using chemical shift changes of Val 620 upon titration with (*b*) pyruvate and (*c*) glyoxylate. (*d*,*e*) Examples of line-shape simulations for (*d*) Ile 482 (^{1}HN dimension) in titration with glyoxylate and (*e*) Val 620 (^{15}N dimension) in titration with pyruvate. Experimental line shapes are shown in red, and the results of the simulations are shown in blue. Modified from Tugarinov & Kay (48).

$0.8\text{x}10^6$ and $1.5\text{x}10^6$ $M^{-1}s^{-1}$ and are slightly higher for glyoxylate, $1.5\text{x}10^6$-$3.8\text{x}10^6$ $M^{-1}s^{-1}$. These on rates are about two orders of magnitude lower than biomolecular diffusion-controlled rates ($\sim 10^9$ $M^{-1}s^{-1}$) and are similar to values reported for substrates of many enzymes (10^6-10^8 $M^{-1}s^{-1}$) (30, 32). The entry of ligands into the binding crevice of apo MSG might require subtle structural rearrangements consistent with a large number of small chemical shift changes

that were quantified upon ligand binding (48), and this may account for the decreased on rates.

FUTURE PROSPECTS—DE NOVO STRUCTURAL STUDIES OF MSG At this point we are well poised to attempt de novo structural studies of MSG. To this end NOESY data sets quantifying methyl-methyl, methyl-HN, and HN-HN distances are being recorded. These distance restraints will be used in concert with the ^{1}HN-^{15}N dipolar couplings and ^{13}CO shift changes upon alignment, which have been measured previously, to attempt to generate global folds of the enzyme. Figure 10 illustrates the quality of structures obtained using a similar approach for β-cyclodextrin-loaded maltose binding protein. In Figure 10*a*, the 10 lowest energy structures calculated on the basis of distance restraints of the type indicated above in combination with hydrogen bond and dihedral angle data are shown (54), and Figure 10*b* shows the structures obtained when dipolar coupling data are included (77). Of note, the relative orientation of the two domains in this protein differs by 10° between solution and crystal conformations (77, 89), but in the apo- and maltotriose-loaded forms of the protein, the interdomain structures obtained from NMR and X-ray analysis are identical (90). This emphasizes the importance of a method alternative to X-ray crystallography for structural studies of multidomain proteins (89, 91).

Structure and Dynamics of Membrane Proteins by Solution NMR Spectroscopy—A Case Study of PagP

PagP is an outer membrane enzyme found in gram-negative bacteria that catalyzes the transfer of a palmitate chain from the *sn*-1 position of a phospholipid molecule in the inner leaflet of the outer membrane to lipid A, in the outer leaflet, forming hepta-acylated lipid A (92), Figure 11*a*. This modification confers resistance to certain antimicrobial peptides that are part of the host innate immune system (93). In an effort to understand the catalytic mechanism of the enzyme as well as the potential role of dynamics in its function, NMR studies on this protein have been initiated.

EXPRESSION AND RECONSTITUTION OF PagP Membrane proteins are notoriously difficult to overexpress, presenting a significant obstacle to structural studies. In some favorable cases, the membrane protein of interest can be made to accumulate in a misfolded form. For example, gram-negative bacterial outer membrane proteins require a signal sequence to direct them to the outer membrane. When the signal sequence is removed, the protein forms dense inactive aggregates within the cytoplasm (94). Although nonfunctional expression is generally avoided due to problems associated with reconstitution, this is not a deterrent to solution NMR studies because unfolding and refolding are often required steps for working with large perdeuterated systems (see In Vitro Refolding of Proteins, above). In fact, this approach is the one that we used to obtain sufficient quantities of protein for the analyses described below.

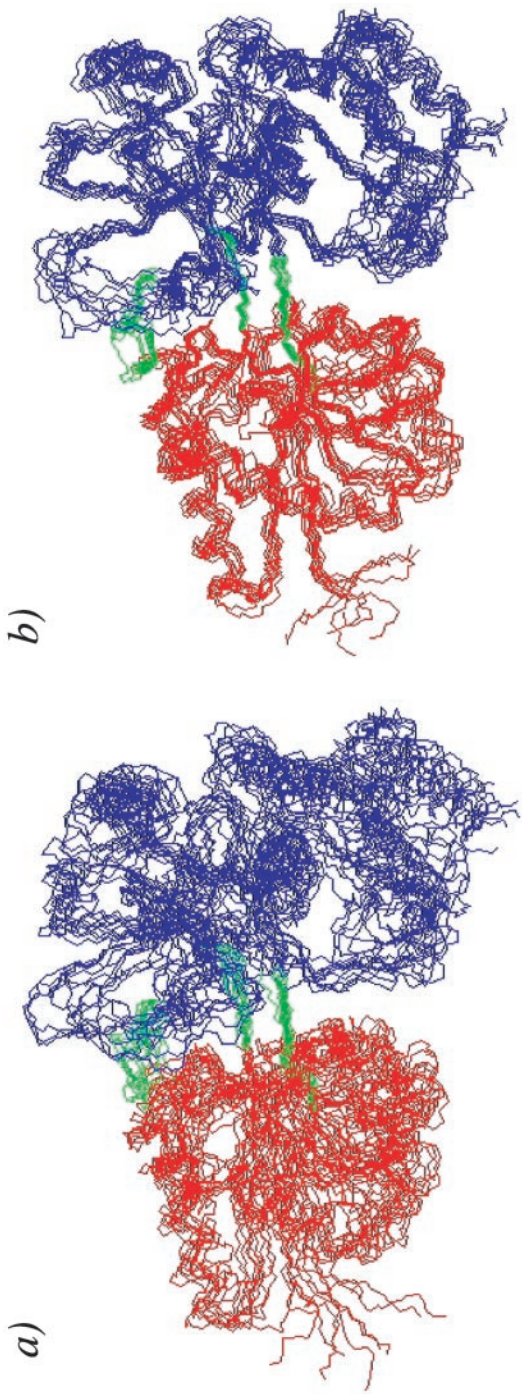


Figure 10 The 10 lowest-energy NMR structures of maltose binding protein (MBP), complexed to β-cyclodextrin (not shown) calculated from (*a*) NOE (826 HN-HN, 769 HN-CH_3, 348 CH_3-CH_3), hydrogen bond (48) and dihedral angle (464) restraints, and (*b*) NOE, hydrogen bond, dihedral angle, carbonyl chemical shift changes upon alignment (278) and residual dipolar coupling (279 ^{1}H-^{15}N, 275 $^{13}C^{\alpha}$-^{13}CO, 261 ^{15}N-^{13}CO) restraints. Domains are in red and blue, and the linkers between them are in green. Modified from (77).

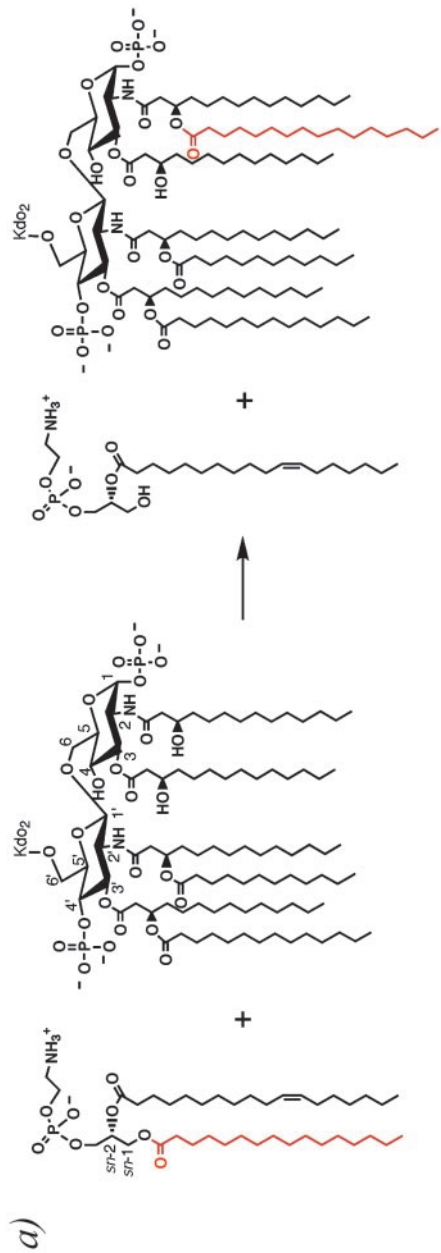


Figure 11 (*a*) PagP-catalyzed palmitoylation of lipid A involving acyl transfer from PtdEtn (phosphatidylethanolamine) to lipid A. (*b*) Ribbon diagram of PagP based on the structure determined in DPC. The protein is embedded in a membrane with phospholipids on the inner leaflet and lipopolysaccharide (lipid A) on the outside. The orientation of the protein relative to the bilayer is not available from the NMR experiments. Residues colored in red have backbone ^{15}N T_1s that are less than 80% of the value predicted for a 20 nanosecond overall correlation time [see Hwang et al. (22) for details]. Yellow residues have NMR signals too faint to be detected in ^{1}HN-^{15}N HSQC spectra, and green residues have weak signal due to conformational exchange. All other residues are shown in blue. The key catalytic residues are indicated. Adapted from Hwang et al. (22).

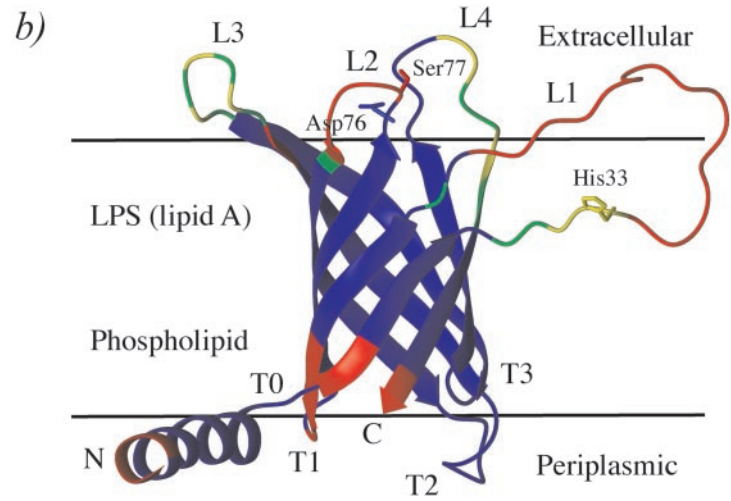


Figure 11 Continued

A number of β-barrel membrane proteins have been successfully overexpressed and reconstituted (94), making this class of membrane protein particularly attractive for structural work. Denatured PagP in 6 M guanidine hydrochloride could be refolded completely by rapid dilution into dodecylphosphocholine containing buffer, as monitored by CD spectroscopy. Samples of PagP in *n*-octyl-β-D-glucoside were also prepared and studied, as described in Hwang et al. (22). Recently developed lipopeptide detergents, amphipathic helical peptides fused to a long-chain fatty acid, are particularly promising for the study of membrane proteins (95). Excellent spectra were obtained of PagP in one such detergent, LPD-14.

STRUCTURE DETERMINATION OF PagP Once a well-folded and stable sample has been produced, there is still no guarantee that it will be amenable to NMR analysis. Fortunately, an ^{1}HN-^{15}N HSQC spectrum run on an ^{15}N-labeled sample is sufficient for determining whether it is worthwhile to continue with structural studies at this stage.

There are currently three solution NMR structures of multi-span integral membrane proteins, including OmpX (37), OmpA (38), and PagP (22). For all three, nonfunctional expression in highly deuterated, ^{15}N, ^{13}C media followed by reconstitution produced homogeneous samples of perdeuterated ^{15}N,^{13}C-labeled protein, protonated at all exchangeable sites. In the case of PagP, backbone chemical shift assignments were carried out using conventional TROSY-based 3D NMR experiments (86, 97). The $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and ^{13}CO chemical shifts obtained during the course of the assignment are very reliable indicators of secondary structure (45), and empirically based programs, such as TALOS, can

be used to extract backbone dihedral angle restraints for most residues found within secondary structure elements (46).

Structural calculations for PagP were carried out using HN-HN NOEs, dihedral angle restraints, and hydrogen bond restraints, the latter established on the basis of hydrogen exchange experiments (22). For many protein topologies this information is insufficient to accurately define global folds. However, in the case of β-barrels, cross-strand NOEs are extremely effective at constraining the structure. Figure 11*b* shows a ribbon diagram of PagP, with the residues important for catalytic activity highlighted. The active site is thus located at the extracellular membrane interface in an optimal position for interacting with the polar groups of lipid A. This also implies that phospholipids, which are located exclusively in the inner leaflet, must somehow migrate to the outer leaflet for the acylation of lipid A to occur, Figure 11*a*.

Although the barrel of PagP is well defined in the solution NMR structure, the N-terminal α-helix could not be properly positioned using only HN-HN NOEs and dihedral angle restraints. In this regard an approach based on selective methyl labeling (see above) might be useful, as described recently in the context of structural studies of OmpX (68). Dipolar couplings (see above) would be of utility in reorienting the helical "domain" relative to the barrel. However, because the linker region T0 is ill-defined, restraints measured in a pair of alignment media would be necessary to uniquely define the helix orientation (99).

DYNAMICS OF PagP BY NMR A recent crystal structure of PagP in lauryldimethylamine oxide has shown that a detergent molecule is bound inside the barrel (G.G. Privé, personal commnication), very likely occupying the binding site for the phospholipid *sn*-1 acyl chain (92). Moreover, because the active site resides in the center of the barrel, the β-strands of the enzyme must come apart somewhat to allow substrate entry. Studying the conformational fluctuations in PagP is key, therefore, to understanding its function.

The main advantage of high-resolution NMR over other structural techniques lies in its ability to measure how a structure changes as a function of time. Dynamic information on the ps-ns timescale can be obtained from ^{15}N spin relaxation experiments of the sort described above for MSG. The results of such a study are color coded on the structure in Figure 11*b*. There are three highly mobile clusters in the protein with ps-ns timescale motions (indicated by red). Correlations from amino acids in yellow could not be observed in HSQC spectra, suggesting that amide positions of these residues are dynamic on a μs-ms timescale, while those residues colored in green have intensities that are attenuated, again pointing to dynamics in this time regime. Interestingly, regions with slow motions tend to localize to the extracellular loops, indicating that significant structural rearrangements are occurring here. Such conformational fluctuations may be important to allow substrate access to the active site. Still slower timescale motions can be studied from hydrogen exchange experiments. The results from all of the dynamics studies together suggest that entry of substrate

into the barrel may occur by breaking contacts between strands A and B or between strands F and G. In this regard, it is of interest that both pairs of strands have a number of proline residues and therefore contain less than the maximal number of bridging hydrogen bonds.

TROSY IN METHYL GROUPS—NEW PERSPECTIVES FOR ^{1}H-^{13}C SPECTROSCOPY OF LARGE PROTEINS AND SUPRAMOLECULAR COMPLEXES

Brief Introduction

New technologies have recently emerged that hold promise for the qualitative study of structure and dynamics in supramolecular complexes. Particularly exciting in this regard are the experiments of Wüthrich and coworkers (100, 101), based on the original work of Dalvit (102), in which polarization is transferred between ^{1}HN and ^{15}N coupled spin pairs due to interference between ^{1}H chemical shift anisotropy and ^{1}HN-^{15}N dipolar interactions. These so called CRIPT and CRINEPT pulse schemes have been used to record ^{1}HN-^{15}N correlation maps of ^{15}N-labeled GroES in a GroES-GroEL complex (900 kDa) and to obtain information about the structure and the dynamics of the assembly (29). The experiments described above focus on backbone positions of a protein. It is of considerable interest to develop complementary approaches for studying side chains, and for reasons that should come as no surprise based on the discussion in the Section on Backbone and Side Chain Assignments above, we have targeted methyl groups as probes of structure and dynamics in this context.

The Methyl-TROSY Effect—A Qualitative Explanation

A detailed analysis of the eight transitions that contribute to ^{13}C magnetization in a methyl group and the 10 ^{1}H transitions from which the ^{1}H signal is derived shows that the relaxation properties are not uniform. In the high-mol-wt limit and considering only the dominant dipolar contributions to relaxation, it can be shown that six of the eight ^{13}C lines are long lived and two relax rapidly (103, 104). Conversely, 50% of the ^{1}H signal decays with a large time constant, with the other half rapidly dephasing (105–107). The construction of a TROSY experiment for methyl groups is predicated, therefore, on designing a pulse scheme in which the fast relaxing transitions are isolated from those that relax slowly. This, of course, is the idea behind any of the amide-based TROSY pulse schemes as well. However, methyl groups are much more complex than ^{1}HN-^{15}N spin pairs, and initial attempts to design TROSY-based experiments for methyls in our group failed.

In a recent series of papers, it has been shown that the simplest of ^{1}H-^{13}C correlation experiments, the HMQC pulse sequence (108, 109), is an optimized TROSY experiment for methyl groups, with 1/2 of the resulting signal derived

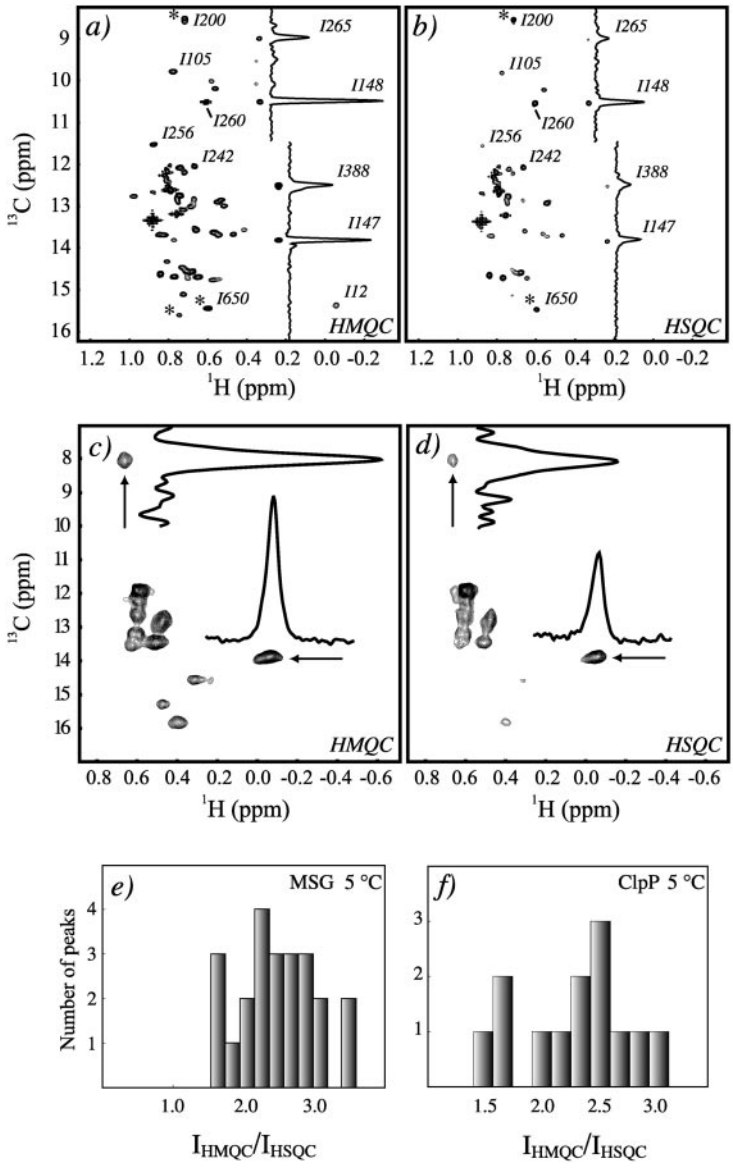


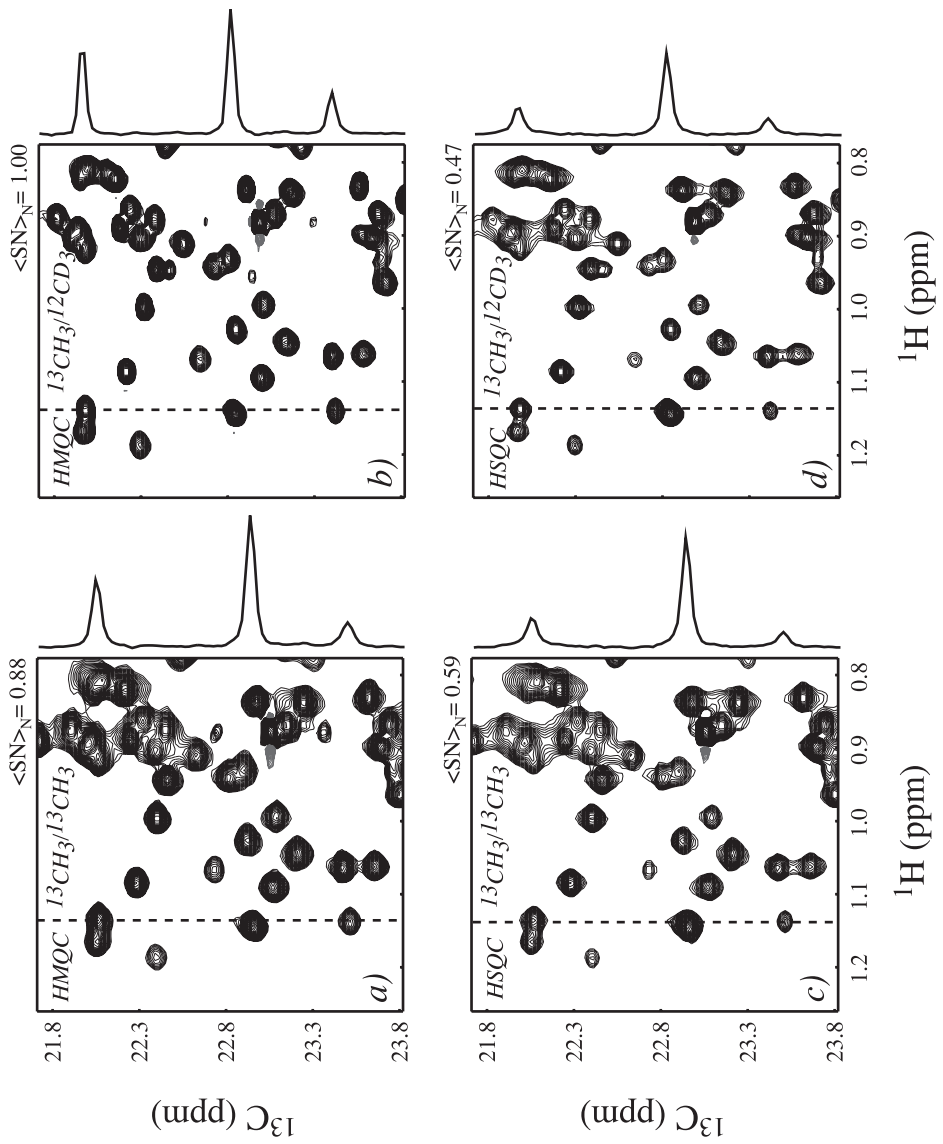
Figure 12 (*a*,*b*) A comparison of Ile-δ1 regions of ^{1}H-^{13}C correlation maps recorded at 5°C (800 MHz) on a U-[^{15}N,^{2}H], Ileδ1-[^{13}C,^{1}H] sample of MSG using (*a*) HMQC and (*b*) HSQC pulse sequences (0.8 mM, 173 min/spectrum). Peaks marked with * are aliased. (*c*,*d*) The same comparison for a U-[^{15}N,^{2}H], Ileδ1-[^{13}C,^{1}H] sample of ClpP protease at 5°C, 800 MHz (1 mM in monomer, spectra recorded in 1.5 h). (*e*,*f*) Histograms of intensity ratios in the HMQC and HSQC spectra (I_{HMQC}/I_{HSQC}) obtained for MSG (*e*) and ClpP (*f*). Adapted from Tugarinov et al. (58).

exclusively from the slowly relaxing transitions of both nuclei (58, 110). That is, the slow (50%) and fast relaxing (50%) pathways are isolated (neglecting relaxation contributions from external ^{1}H spins). Conversely, the much more popular HSQC experiment (111) intermixes fast and slowly relaxing transitions so that only 3/16 of the net signal derives from coherences that relax slowly throughout the course of the experiment. More detailed explanations are given in the literature (58, 110) and are beyond the scope of the present discussion. Suffice it to say that the TROSY effect observed in methyl groups using the HMQC experiment results from the cancellation of dipolar (both ^{1}H-^{1}H and ^{1}H-^{13}C) fields in a manner that is completely magnetic field independent, so long as the system tumbles slowly ($\omega_C \tau_M \gg 1$, in which ω_C is the ^{13}C Larmor frequency and τ_M is the tumbling time). Below, we compare spectra of regular HMQC and HSQC experiments recorded on two samples of large perdeuterated methyl-protonated proteins at 5°C and illustrate some of the benefits described above.

EXPERIMENTAL VERIFICATION WITH HIGH-MOLECULAR-WEIGHT PROTEINS As with ^{1}HN-^{15}N TROSY, the intensity of correlations in methyl-TROSY spectra are also sensitive to contributions to relaxation from external spins (58). The experiments are thus best performed using highly deuterated samples. As an initial test of the methodology, perdeuterated protein samples (dissolved in D_2O) were prepared with [^{13}C,^{1}H]-labeled Ile δ1 methyl groups. Proteins with this labeling pattern can be obtained using α-ketobutyrate [^{13}C,^{1}H]-labeled at the methyl position (compound IV in Figure 4) and [^{2}H,^{12}C]-D-glucose as carbon sources in D_2O-based minimal media growths. Restricting protonation to Ile δ1 sites ensures that the only external source of relaxation for a given methyl derives from other Ile δ1 methyls in the protein. In the case of MSG, the average sum of effective distances from a given methyl to all external protons is ∼5.5 Å using this labeling scheme, so that the majority of Ile methyls are well isolated.

Figure 12*a*,*b* shows ^{1}H-^{13}C methyl correlation maps recorded on U-[^{15}N,^{2}H], Ileδ1-[^{13}C,^{1}H] MSG in D_2O at 5°C, 800 MHz, using HMQC (panel *a*) and HSQC (panel *b*) pulse schemes. The correlation time of MSG at 5°C in D_2O is ∼120 ns (equivalent to a molecule in the 200–250 kDa range at 37°C), estimated as described in Tugarinov et al. (58). For comparison, both spectra are plotted at the same level, and it is clear that the HMQC data set has significantly higher sensitivity. A histogram of the ratios of signal to noise in HMQC and HSQC

Figure 13 A comparison of a selected region of HMQC and HSQC spectra of U-[^{15}N,^{2}H], MSG (samples are 0.71 mM), 37°C, 800 MHz, with Leu,Val methyl groups labeled as indicated. The average, normalized signal to noise ratio, $<SN>_N$, for each spectrum is indicated. All spectra are recorded and processed identically and are shown with the same contour levels. Reproduced with permission from (113).


a)
b)
c)
d)
HMQC
HSQC
$^{13}CH_3/^{13}CH_3$
$^{13}CH_3/^{12}CD_3$
<SN>$_N$= 0.88
<SN>$_N$= 1.00
<SN>$_N$= 0.59
<SN>$_N$= 0.47
^{13}C (ppm)
^{1}H (ppm)
21.8
22.3
22.8
23.3
23.8
1.2
1.1
1.0
0.9
0.8

spectra recorded at 5°C, (I_{HMQC}/I_{HSQC}) is shown in Figure 12*e*. On average, a gain in signal of a factor of 2.6 is noted in the HMQC data set.

Figure 12*c*,*d* demonstrates the improvements in sensitivity of HMQC versus HSQC methyl correlation maps that have been recorded on the U-[^{15}N,^{2}H], Ileδ1-[^{13}C,^{1}H]-labeled protease ClpP in D_2O at 5°C. This complex consists of 14 identical subunits, each of 193 residues (total mol wt 305 kDa), with 16 isoleucines/subunit (112). At 5°C, the rotational correlation time of ClpP in D_2O is estimated in the 400−450 ns range. At least 14 of the expected 16 correlations are counted in the HMQC spectrum of Figure 12*c*. As shown in the histogram of Figure 12*f*, sensitivity gains of correlations in HMQC relative to HSQC maps vary between a factor of 1.5 to 3.1 depending on the residue. Sensitivity gains lower than those theoretically predicted (58) may result from conformational heterogeneity (chemical exchange), which affects HMQC spectra to a greater extent than HSQC data sets. Indeed, significant line broadening was observed for several peaks in spectra of ClpP recorded at higher temperatures.

The high quality of HMQC spectra recorded at 5°C on the pair of systems described above suggests that methyl-TROSY will prove useful in the study of very large proteins and supramolecular complexes.

Optimal Isotopic Labeling Methods for TROSY in Methyl Groups

The labeling scheme involving selective protonation of Ile δ1 methyl groups is of somewhat limited utility because only a single type of probe is available for structural and dynamical studies, and it would clearly be advantageous to include protonation at Leu and Val methyl sites. As described above, however, the introduction of protons comes at the expense of degrading the TROSY effect because the added protons are effective sources of external relaxation. In Leu and Val residues in which two methyl groups are side by side, intraresidue ^{1}H-^{1}H dipolar interactions are especially detrimental to the TROSY effect. With this in mind, a labeling scheme has been proposed in which each Leu and Val is labeled with one $^{13}CH_3$ and one $^{12}CD_3$ methyl (using compound V of Figure 4) (113). As we show below, the sensitivity loss associated with the twofold dilution of methyl protons in Leu and Val residues is significantly more than compensated for (both in terms of increased sensitivity and resolution) by the improved relaxation properties of the remaining NMR active methyls.

Figure 13 shows selected regions of HMQC and HSQC data sets recorded on a pair of MSG samples (37°C, τ_C = 45 ns) with equal protein concentrations (0.71 ± 0.05 mM). These include (*a*) a U-[^{2}H,^{15}N] Leu,Val-[$^{13}CH_3/^{13}CH_3$] sample and (*b*) a U-[^{2}H,^{15}N] Leu,Val-[$^{13}CH_3/^{12}CD_3$] sample—both prepared using the appropriately labeled α-ketoisovalerate (compounds III and V in Figure 4). The average normalized signal-to-noise ratios ($<SN>_N$) for all well-separated correlations in the full spectra are indicated above each plot. A comparison of the panels *b* and *d* in Figure 13 establishes that the same benefits observed in HMQC data sets relative to the corresponding HSQC maps in the case of

Ile-protonated samples (Figure 12) are also seen for samples with Leu and Val methyls labeled with [$^{13}CH_3/^{12}CD_3$]. Namely, the HMQC spectrum is of significantly better resolution with more than a factor of two improvement in $<SN>_N$ relative to its HSQC counterpart.

The HMQC data sets in panels *a* and *b* of Figure 13 show that despite the loss of a factor of two in concentration of NMR active methyls in the case of [$^{13}CH_3/^{12}CD_3$] Leu/Val methyl labeling, spectra are nonetheless of improved sensitivity and resolution relative to those recorded on Leu,Val-[$^{13}CH_3/^{13}CH_3$] MSG samples. The differences between HMQC spectra recorded on the two sets of labeled samples are even more pronounced for MSG at 5°C ($\tau_C \approx 120$ ns) with an average <SN> improvement of close to a factor of two in favor of the Leu,Val-[$^{13}CH_3/^{12}CD_3$]-labeled sample. The proposed Leu,Val-[$^{13}CH_3/^{12}CD_3$] labeling scheme is, therefore, critical for the optimization of the methyl TROSY effect.

Simultaneous protonation of Ile $\delta1$ and [$^{13}CH_3/^{12}CD_3$]-labeling of Leu and Val methyls in deuterated protein samples is desirable to maximize the number of available methyl sites for NMR studies. Figure 14 shows a comparison of HMQC and HSQC correlation maps recorded on a U-[^{2}H,^{15}N] Ile$\delta1$-[$^{13}CH_3$] Leu,Val-[$^{13}CH_3/^{12}CD_3$] MSG sample at 5°C. A gain of 2.6 in <SN> and significant improvements in resolution are noted for the HMQC data set, with essentially no differences between Leu, Val correlations in spectra recorded on U-[^{2}H,^{15}N] Ile$\delta1$-[$^{13}CH_3$] Leu,Val-[$^{13}CH_3/^{12}CD_3$] and U-[^{2}H,^{15}N] Leu,Val-[$^{13}CH_3/^{12}CD_3$] MSG samples.

CONCLUDING REMARKS

Recent developments in NMR spectroscopy have been highlighted through a discussion of their application to studies of a high-mol-wt single polypeptide enzyme, malate synthase G (723 residues, 82 kDa). Backbone ^{1}HN, ^{15}N, $^{13}C^{\alpha}$, ^{13}CO, and side chain $^{13}C^{\beta}$ assignments of the protein have been obtained from a series of 4D TROSY-based triple resonance experiments—the first step in any detailed analysis of protein structure and dynamics. These assignments have facilitated a quantitative study of domain orientation and dynamics as a function of a variety of ligands that bind to the protein. Near complete side chain methyl assignments have been made using new labeling schemes and pulse sequences with improved sensitivity and resolution. Studies of the dynamics of side chains and how such dynamics change in response to ligand binding are now also possible. Advances in solution spectroscopy of water-soluble proteins can be put to good use in the study of membrane proteins, illustrated here in the context of the determination of the global fold of PagP, an eight-stranded β-barrel membrane protein. Finally, TROSY, originally developed for amide ^{1}HN, ^{15}N spin pairs has been extended to $^{13}CH_3$ methyl groups, so that both backbone and now side chain structure and dynamics can be probed in very high-mol-wt systems.

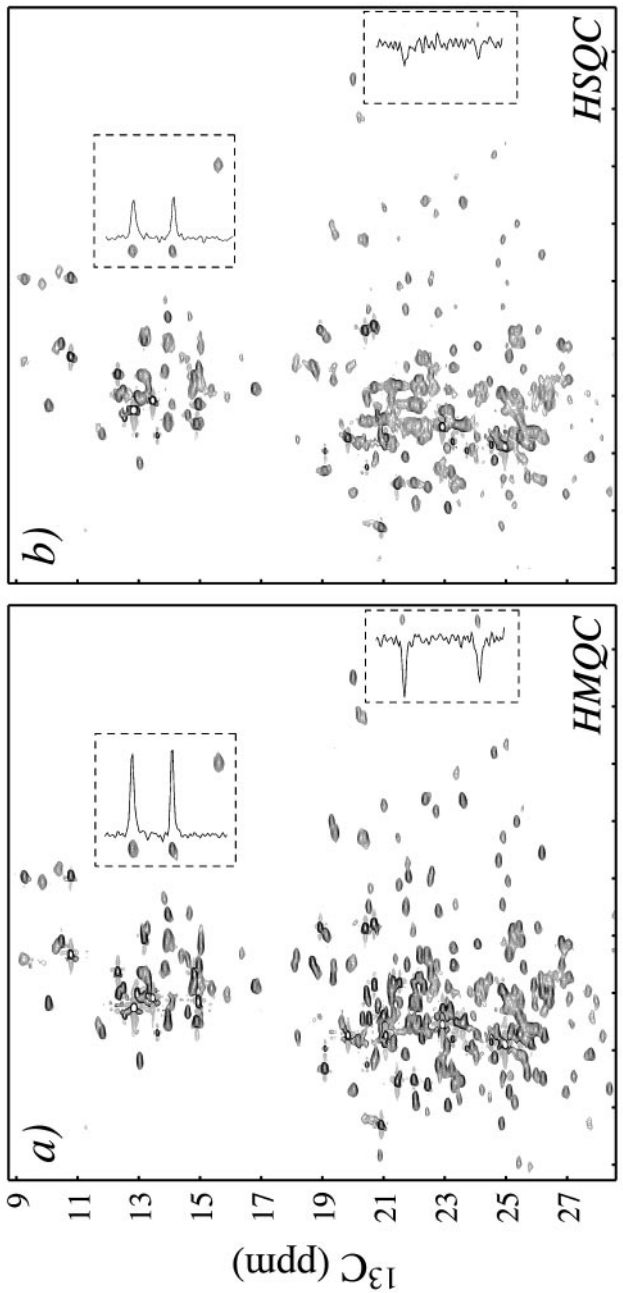


Figure 14 A comparison of HMQC (*a*) and HSQC (*b*) ^{1}H-^{13}C correlation maps recorded on a U-[^{2}H,^{15}N], Ileδ1-[$^{13}CH_3$] Leu,Val-[$^{13}CH_3/^{12}CD_3$] sample of MSG at 5°C ($\tau_c \approx 118$ns), 800 MHz (0.63 mM, net acquisition time 2.5 h/spectrum). Adapted from (113).

ACKNOWLEDGMENTS

This work was supported by a grant from the Canadian Institutes of Health Research (CIHR) to L.E.K. We thank Dr. W.-Y. Choy (University of Toronto) for providing Figure 10 of this review. V.T. acknowledges financial support from the Human Frontiers Science Program. L.E.K. holds a Canada Research Chair in Biochemistry and is a member of the Protein Engineering Network Centres of Excellence.

The *Annual Review of Biochemistry* is online at http://biochem.annualreviews.org

LITERATURE CITED

1. Bax A, Grzesiek S. 1993. *Acc. Chem. Res.* 26:131–38
2. Goto NK, Kay LE. 2000. *Curr. Opin. Struct. Biol.* 10:585–92
3. Tjandra N, Bax A. 1997. *Science* 278: 1111–14
4. Hansen MR, Mueller L, Pardi A. 1998. *Nat. Struct. Biol.* 5:1065–74
5. Clore GM, Starich MR, Gronenborn AM. 1998. *J. Am. Chem. Soc.* 120: 10571–72
6. Ruckert M, Otting G. 2000. *J. Am. Chem. Soc.* 122:7793–97
7. Sass J, Cordier F, Hoffman A, Cousin A, Omichinski JG, et al. 1999. *J. Am. Chem. Soc.* 121:2047–55
8. Bax A, Kontaxis G, Tjandra N. 2001. *Methods Enzymol.* 339:127–74
9. Cornilescu G, Marquardt J, Ottiger M, Bax A. 1998. *J. Am. Chem. Soc.* 120: 6836–37
10. Clore GM, Starich MR, Bewley CA, Cai M, Kuszewski J. 1999. *J. Am. Chem. Soc.* 121:6513–14
11. Pervushin K, Riek R, Wider G, Wüthrich K. 1997. *Proc. Natl. Acad. Sci. USA* 94:12366–71
12. Pervushin K. 2001. *J. Biomol. NMR* 20:275–85
13. LeMaster DM. 1990. *Q. Rev. Biophys.* 23:133–74
14. Clore GM, Gronenborn AM. 1991. *Science* 252:1309–99
15. Gardner KH, Kay LE. 1998. *Annu. Rev. Biophys. Biomol. Struct.* 27:357–406
16. Farmer BT, Venters RA. 1998. In *Biological Magnetic Resonance*, ed. NR Krishna, LJ Berliner, pp. 75–120. New York: Kluwer Acad./Plenum
17. Prestegard JH. 1998. *Nat. Struct. Biol. NMR Suppl.* 5:517–22
18. Wider G, Wüthrich K. 1999. *Curr. Opin. Struct. Biol.* 9:594–601
19. Kanelis V, Forman-Kay JD, Kay LE. 2001. *IUBMB Life* 52:291–302
20. Bax A. 2003. *Protein Sci.* 12:1–16
21. Howard BR, Endrizzi JA, Remington SJ. 2000. *Biochemistry* 39:3156–68
22. Hwang PM, Choy WY, Lo EI, Chen L, Forman-Kay JD, et al. 2002. *Proc. Natl. Acad. Sci. USA* 99:13560–65
23. Grzesiek S, Anglister J, Ren H, Bax A. 1993. *J. Am. Chem. Soc.* 115:4369–70
24. Yamazaki T, Lee W, Arrowsmith CH, Muhandiram DR, Kay LE. 1994. *J. Am. Chem. Soc.* 116:11655–66
25. Gardner KH, Zhang X, Gehring K, Kay LE. 1998. *J. Am. Chem. Soc.* 120: 11738–48
26. Tugarinov V, Muhandiram R, Ayed A, Kay LE. 2002. *J. Am. Chem. Soc.* 124: 10025–35
27. Salzmann M, Pervushin K, Wider G, Senn H, Wüthrich K. 2000. *J. Am. Chem. Soc.* 122:7543–48
28. Mulder FAA, Ayed A, Yang D, Arrow-

smith CH, Kay LE. 2000. *J. Biomol. NMR* 18:173–76
29. Fiaux J, Bertelsen EB, Horwich AL, Wüthrich K. 2002. *Nature* 418:207–21
30. Creighton TE. 1993. *Proteins: Structures and Molecular Properties*. New York: Freeman
31. Creighton TE, ed. 1992. *Protein Folding*. New York: Freeman
32. Fersht A. 1999. *Structure and Mechanism in Protein Science*. New York: Freeman
33. Altamirano MM, Golbik R, Zahn R, Buckle AM, Fersht AR. 1997. *Proc. Natl. Acad. Sci. USA* 94:3576–78
34. Goldberg ME, Expert-Bezançon N, Vuillard L, Rabilloud T. 1995. *Fold. Des.* 1:21–27
35. Lilie H, Schwarz E, Rudolph R. 1998. *Curr. Opin. Biotechnol.* 9:497–501
36. Fernandez C, Hilty C, Bonjour S, Adeishvili K, Pervushin K, Wüthrich K. 2001. *FEBS Lett.* 504:173–78
37. Fernandez C, Adeishvili K, Wüthrich K. 2001. *Proc. Natl. Acad. Sci. USA* 98: 2358–63
38. Arora A, Abildgaard F, Bushweller JH, Tamm LK. 2001. *Nat. Struct. Biol.* 8:334–38
39. Yang D, Kay LE. 1999. *J. Am. Chem. Soc.* 121:2571–75
40. Konrat R, Yang D, Kay LE. 1999. *J. Biomol. NMR* 15:309–13
41. Venters RA, Metzler WJ, Spicer LD, Mueller L, Farmer BT. 1995. *J. Am. Chem. Soc.* 117:9592–93
42. Grzesiek S, Wingfield P, Stahl S, Kaufman J, Bax A. 1995. *J. Am. Chem. Soc.* 117:9594–95
43. Salzmann M, Wider G, Pervushin K, Senn H, Wüthrich K. 1999. *J. Am. Chem. Soc.* 121:844–48
44. Coggins BE, Zhou P. 2003. *J. Biomol. NMR* 26:93–111
45. Wishart DS, Sykes BD. 1994. *J. Biomol. NMR* 4:171–80
46. Cornilescu G, Delaglio F, Bax A. 1999. *J. Biomol. NMR* 13:289–302
47. Chazin WJ, Kordel J, Thulin E, Hofmann T, Drakenberg T, Forsen S. 1989. *Biochemistry* 28:8646–53
48. Tugarinov V, Kay LE. 2003. *J. Mol. Biol.* 327:1121–33
49. Tugarinov V, Kay LE. 2003. *J. Am. Chem. Soc.* 125:5701–6
50. Tugarinov V, Kay LE. 2003. *J. Am. Chem. Soc.* 125:13868–78
51. Eletsky A, Moreira O, Kovacs H, Pervushin K. 2003. *J. Biomol. NMR.* 26:167–79
52. Gardner KH, Kay LE. 1997. *J. Am. Chem. Soc.* 119:7599–600
53. Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE. 1999. *J. Biomol. NMR* 13:369–74
54. Mueller GA, Choy WY, Yang D, Forman-Kay JD, Venters RA, Kay LE. 2000. *J. Mol. Biol.* 300:197–212
55. Nicholson LK, Kay LE, Baldisseri DM, Arango J, Young PE, et al. 1992. *Biochemistry* 31:5253–63
56. Mulder FAA, Mittermaier A, Hon B, Dahlquist FW, Kay LE. 2001. *Nat. Struct. Biol.* 8:932–35
57. Gardner KH, Rosen MK, Kay LE. 1997. *Biochemistry* 36:1389–401
58. Tugarinov V, Ollerenshaw J, Hwang P, Kay LE. 2003. *J. Am. Chem. Soc.* 125: 10420–28
59. Gottschalk G. 1986. *Bacterial Metabolism*. Berlin: Springer Verlag
60. Hajduk PJ, Augeri DJ, Mack J, Mendoza R, Yang JG, et al. 2000. *J. Am. Chem. Soc.* 122:7898–904
61. Gross JD, Gelev VM, Wagner G. 2003. *J. Biomol. NMR* 25:235–42
62. Fesik SW, Eaton HL, Olejniczak ET, Zuiderweg ERP, McIntosh LP, Dahlquist FW. 1990. *J. Am. Chem. Soc.* 112: 886–88
63. Montelione GT, Lyons BA, Emerson SD, Tashiro M. 1992. *J. Am. Chem. Soc.* 114:10974–75
64. Logan TM, Olejniczak ET, Xu RX, Fesik SW. 1993. *J. Biomol. NMR* 3:225–31

65. Grzesiek S, Anglister J, Bax A. 1993. *J. Magn. Reson. B* 101:114–19
66. Gardner KH, Konrat R, Rosen MK, Kay LE. 1996. *J. Biomol. NMR* 8:351–56
67. Lin Y, Wagner G. 1999. *J. Biomol. NMR* 15:227–39
68. Hilty C, Fernandez C, Wider G, Wüthrich K. 2002. *J. Biomol. NMR* 23:289–301
69. Deleted in proof
70. Bastiaan EW, MacLean C, van Zijl PCM, Bothner-By AA. 1987. *Annu. Rep. NMR Spectrosc.* 9:35–77
71. Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH. 1995. *Proc. Natl. Acad. Sci. USA* 92:9279–83
72. Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH. 1997. *Nat. Struct. Biol.* 4:292–97
73. Yang D, Venters RA, Mueller GA, Choy WY, Kay LE. 1999. *J. Biomol. NMR* 14:333–43
74. Kontaxis G, Clore GM, Bax A. 2000. *J. Magn. Reson.* 143:184–96
75. Permi P. 2000. *J. Biomol. NMR* 17: 43–54
76. Ottiger M, Tjandra N, Bax A. 1997. *J. Am. Chem. Soc.* 119:9825–30
77. Choy WY, Tollinger M, Mueller GA, Kay LE. 2001. *J. Biomol. NMR.* 21: 31–40
78. Lipsitz RS, Tjandra N. 2001. *J. Am. Chem. Soc.* 123:11065–66
79. Wu Z, Tjandra N, Bax A. 2001. *J. Am. Chem. Soc.* 123:3617–18
80. Beeckmans S, Khan AS, Kanarek L, Van Driessche E. 1994. *Biochem. J.* 303: 413–21
81. Schmid G, Durchschlag H, Biedermann G, Eggerer H, Jaenicke R. 1974. *Biochem. Biophys. Res. Commun.* 58: 419–26
82. Larsen TM, Laughlin LT, Holden HM, Rayment I, Reed GH. 1994. *Biochemistry* 33:6301–9
83. Larsen TM, Laughlin LT, Holden HM, Rayment I, Reed GH. 1997. *Arch. Biochem. Biophys.* 345:199–206
84. Remington SJ, Wiegand G, Huber R. 1982. *J. Mol. Biol.* 158:111–52
85. Sandstrom J. 1982. *Dynamic NMR Spectroscopy*. New York: Academic
86. Yang D, Kay LE. 1999. *J. Biomol. NMR* 13:3–10
87. Hwang TL, van Zijl PC, Mori S. 1998. *J. Biomol. NMR.* 11:221–26
88. Johnson PE, Brun E, MacKenzie LF, Withers SG, McIntosh LP. 1999. *J. Mol. Biol.* 287:609–25
89. Skrynnikov NR, Goto NK, Yang D, Choy WY, Tolman JR, et al. 2000. *J. Mol. Biol.* 295:1265–73
90. Evenas J, Tugarinov V, Skrynnikov NR, Goto NK, Muhandiram R, Kay LE. 2001. *J. Mol. Biol.* 309:961–74
91. Fischer MW, Losonczi JA, Weaver JL, Prestegard JH. 1999. *Biochemistry* 38: 9013–22
92. Bishop RE, Gibbons HS, Guina T, Trent MS, Miller SI, Raetz CR. 2000. *EMBO J.* 19:5071–80
93. Guo L, Lim KB, Poduje CM, Daniel M, Gunn JS, et al. 1998. *Cell* 95:189–98
94. Buchanan SK. 1999. *Curr. Opin. Struct. Biol.* 9:455–61
95. McGregor CL, Chen L, Pomroy NC, Hwang P, Go S, et al. 2003. *Nat. Biotechnol.* 21:171–76
96. Deleted in proof
97. Salzmann M, Pervushin K, Wider G, Senn H, Wüthrich K. 1998. *Proc. Natl. Acad. Sci. USA* 95:13585–90
98. Deleted in proof
99. Al-Hashimi HM, Valafar H, Terrell M, Zartler ER, Eidsness MK, Prestegard JH. 2000. *J. Magn. Reson.* 143:402–6
100. Riek R, Fiaux J, Bertelsen EB, Horwich AL, Wüthrich K. 2002. *J. Am. Chem. Soc.* 124:12144–53
101. Riek R, Pervushin K, Wüthrich K. 2000. *Trends Biochem. Sci.* 25:462–68
102. Dalvit C. 1992. *J. Magn. Reson.* 97: 645–50
103. Kay LE, Bull TE. 1992. *J. Magn. Reson.* 99:615–22

104. Kay LE, Torchia DA. 1991. *J. Magn. Reson.* 95:536–47
105. Werbelow LG, Marshall AG. 1973. *J. Magn. Reson.* 11:299–313
106. Müller N, Bodenhausen G, Ernst RR. 1987. *J. Magn. Reson.* 75:297–334
107. Kay LE, Prestegard JH. 1987. *J. Am. Chem. Soc.* 109:3829–35
108. Bax A, Griffey RH, Hawkings BL. 1983. *J. Magn. Reson.* 55:301–15
109. Mueller L. 1979. *J. Am. Chem. Soc.* 101: 4481–84
110. Ollerenshaw JE, Tugarinov V, Kay LE. 2003. *Magn. Reson. Chem.* 41:843–52
111. Bodenhausen G, Rubin DJ. 1980. *Chem. Phys. Lett.* 69:185–89
112. Wang J, Harting JA, Flanagan JM. 1997. *Cell* 91:447–56
113. Tugarinov V, Kay LE. 2004. *J. Biomol. NMR.* 28:165–72

Annual Review of Biochemistry
Volume 73, 2004

CONTENTS