Journal of Biomolecular NMR **27:** 235–259, 2003.

The influence of internuclear spatial distribution and instrument noise on the precision of distances determined by solid state NMR of isotopically enriched proteins

John D. Gehman[a,b], Eric K. Paulson[b] & Kurt W. Zilm[b,*]
[a]*Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, U.S.A.;* [b]*Department of Chemistry, Yale University, New Haven, CT 06520, U.S.A.*

Received 13 February 2003; Accepted 20 May 2003

Key words: dipolar recoupling, isotopically enriched protein, NMR structure determination, solid state NMR, spin diffusion, spin exchange

Abstract

The relative merits of different isotopic enrichment strategies that might be used in solid state NMR protein structure determinations are explored. The basis for comparison of these merits is the determination of the relative uncertainties in rates measured by a generalized dipolar recoupling experiment. The different schemes considered use ^{13}C, ^{15}N and ^{2}H labeling of ubiquitin with homonuclear magnetization-transfer type experiments under magic-angle spinning (MAS). Specific attention is given to the sensitivity of the predicted relative precisions to variation in natural nuclear density distribution and noise levels. A framework is suggested to gauge the precision of measurement of a given dipolar coupling constant, and the potential for a set of such measurements to constrain structure calculations is explored. The distribution of nuclei in homonuclear ^{15}N and ^{1}H dipolar recoupling spin-exchange experiments appear to provide the most promising tertiary structure information for uniformly labeled ubiquitin.

Introduction

NMR is a powerful spectroscopic tool by virtue of its sensitivity to the local environment of nuclei. Its conjugate weakness is a lack of sensitivity to global molecular relationships. Yet it is the nuclear dipole–dipole interactions which are a function of long range distances that are most important in solving macromolecular structure. Solid state NMR methods in combination with various isotopic labeling strategies have been used to measure such long range individual distances with unparalleled precision. Specific pairwise labeling has been particularly successful at determining important distance constraints for defining binding site geometry in enzyme-substrate (Christensen and Schaefer, 1993) and more recently receptor-ligand complexes (Watts, 1999). Related approaches have also made significant contributions to our understanding of amyloid fibril structure (Benzinger et al., 1998; Tycko, 2001). The selective labeling approach has been impressively extended to the full structure determination of the membrane resident gramicidin channel, using multiple samples to provide a sufficient number of constraints (Ketchem et al., 1996). In another landmark application, rotational echo double resonance (REDOR) of a single incorporated ^{13}C/^{19}F pair was employed to monitor a ligand-binding induced structural change of a cell signaling receptor in an intact cell membrane preparation (Murphy et al., 2001).

For macromolecular systems in which an important structural question can be addressed using a single spin pair, and in which the chemistry of this incorporation is not a major impediment, such approaches will continue to be important. However, from a practical and economic perspective, the quantity of structural information per sample is low. If solid state NMR is to become a mainstream technique and complement

*To whom correspondence should be addressed. E-mail: kurt.zilm@yale.edu

solution NMR and single crystal diffraction for protein structure work, it must evolve to make use of more extensive isotopic enrichment as is already common in solution NMR.

Several classes of important problems motivate development of such methods. By definition amyloid fibril structure and consequent understanding of protein misfolding will remain intractable by solution NMR and diffraction techniques. The fundamental interconnectedness of structure and function in membrane-associated proteins is most meaningfully addressed in their native environment, also difficult to address without solid state NMR. Of considerable practical importance are efforts to elucidate structure of bound ligands in pharmaceutically interesting complexes. Samples for solid state NMR experiments on such complexes can employ soaking techniques common to diffraction sample preparations, while obviating the need for large well-ordered crystals. While the solution NMR technique of transferred NOE has proven useful (Schalk-Hihi et al., 1998; Inooka et al., 2001) it is limited to ligands with sufficiently short occupancy times. For all the above cases, full isotopic enrichment of macromolecules, subunits, or ligands can be easily and relatively inexpensively accomplished such that many useful constraints are in theory simultaneously measurable by solid state NMR.

Realizing the potential of solid state NMR in this regard has, however, not been trivial. Lower sensitivity and resolution than solution NMR, lack of sequential assignment methods applicable to large molecules, and the absence of methods for determining large sets of structural constraints in isotopically enriched proteins have all been hurdles. Rapid progress has been made recently which addresses sensitivity, resolution, and assignment problems. The advent of solid state NMR instrumentation operating at fields over 18 Tesla has brought sample quantity requirements down to a few hundred nanomoles, similar to that of solution NMR (McDermott et al., 2000). While solid state ^{1}H NMR still does not provide particularly high resolution for large spin systems, line widths have been narrowed to the extent that its use in indirect detection methods for sensitivity enhancement of heteronuclei has been demonstrated (Ishii et al., 2001). Multinuclear sequential and side-chain assignment strategies have been developed in several laboratories, and have proven successful upon application to small proteins such as BPTI (McDermott et al., 2000), the SH3 domain of α-spectrin (Pauli et al., 2001; Rossum et al., 2001), and ubiquitin (Straus et al., 1998; Igumenova et al., 2003). Generally applicable techniques for structure determination in multiply-labeled spin systems have also been developed in a number of laboratories, with proof-of-principle experiments focusing on small model peptides (e.g., Michal and Jelinski, 1997; Nomura et al., 2000; Jaroniec et al., 2001). The state of the art is such that sensitivity, resolution and spectral assignment are no longer the insurmountable hurdles that once seemed to exclude solid state NMR as a method for structure determination of even a small protein as evidenced by the first such report (Castellani et al., 2002).

In this paper we concentrate on one of the next steps, using solid state NMR to extract large numbers of distance constraints for an entire protein. To succeed, one needs to obtain ample, accurate, precise, and structurally meaningful data while using the fewest number of samples possible. While the accuracy of distances determined in these types of solid state NMR experiments has been discussed for small spin systems (Hodgkinson and Emsley, 1999), our goal here is to explore the difficulties in the context of an entire protein. Our motivation and philosophy are generally, and in part specifically, similar to that presented by Tycko (1996) in his discussion of how resolution limits the applicability of solid state NMR to smaller macromolecules. Many aspects of the approach taken here were also inspired by related solution NMR studies, such as the work of Liu et al. (1995) in their discussion of how more realistic distance limits could be determined for nOes measured by solution NMR, and Dellwo (1994) in their discussion of error reduction in rate matrix analysis of NOESY spectra.

There are several experimental approaches to consider. Biochemically there are uniform and specific ^{13}C, ^{15}N, and ^{1}H labeling (of perdeuterated protein) protocols at one's disposal, which can be coupled with several complementary spectroscopic techniques. A few of the obvious and easily achieved labeling patterns are chosen both to serve as examples in our analysis and to characterize a few of the factors governing the precision of a given experimental approach. We focus on distance-dependent magnetization transfer rates measured by homonuclear multidimensional solid state MAS dipolar recoupling experiments analogous to the solution phase NOESY method. REDOR (Gullion and Schaefer, 1989) and TEDOR based heteronuclear distance measurements (Michal and Jelinski, 1997), while providing complimentary information and another approach to the problems to be discussed, are not readily treated in the fashion to be

introduced here. As such these techniques will not be considered further in any depth.

We begin with specific qualitative consideration of the distributions of internuclear distances and dipolar couplings in a model protein, which helps to better define the difficulties encountered with different isotopic enrichment patterns. The model system chosen for analysis is human ubiquitin. Ubiquitin is a 76-residue, 21% α-helix, and 43% β-sheet protein which has been extensively characterized by X-ray crystallography (Alexeev et al., 1994), solution NMR (Weber et al., 1987) and recently solid state NMR (Igumenova et al., 2003). We use atomic coordinates from the 1.8 Å crystal structure deposited in the PDB under 1UBI, augmented by ^{1}H coordinates calculated using the program Reduce (Word et al., 1999), and take the structure to be static. We also compare the relative precision with which each dipolar coupling might be measured by comparing a given coupling to the other couplings experienced by the same nuclei. Sorting couplings in this fashion allows us to recognize which couplings might be reliably measured and also provide tertiary structural information.

In the second section we provide a more critical comparison where the effects of the nuclear distribution and experimental noise on the precision of the measured dipolar couplings are quantitatively investigated by computation. We employ one conceivable way to assess the impact of noise and nuclear distribution on the quality of data for an arbitrary homonuclear mixing experiment. Based on a protein structure from a PDB file, magnetization exchange data is simulated using a rate matrix approach for each of several mixing times, and repeatedly seeded with Gaussian random noise values over many Monte Carlo cycles. In each cycle, the noise-seeded data is analytically transformed to an 'experimental' rate matrix from which the corresponding 'experimental' internuclear distance matrix can be determined. Analysis of the resulting distributions of each measured rate characterizes the potential for the respective spin system to provide accurate and meaningful structural constraints.

Finally we tie the two sections together by exploring the interplay of nuclear spatial distribution and finite signal-to-noise such that all can be considered at once for their impact on the precision of distances providing tertiary contact information.

Nuclear spin distributions and their implications for magnetization exchange kinetics

We begin by stepping into the position of the investigator who requires sufficiently numerous, accurate, and meaningful distance measurements with which to constrain protein structure refinement calculations, and wishes to obtain these constraints by measurement of internuclear dipolar coupling constants using solid state NMR. Of the many possibilities available we consider only a few of the most generic isotopic enrichment protocols. The most widely used are 100% enrichment of ^{13}C and/or ^{15}N, and schemes for producing a dilute proton pool. The last scheme assumes a perdeuterated protein preparation, and subsequent back-exchange of all backbone amide and labile side chain protons. This approach has been shown to provide useful resolution even in a directly detected proton dimension for small peptides (Reif et al., 2001), and has yielded good distance constraints for small protein (Reif et al., 2003). A further twist is the selective incorporation of methyl protons according to the synthesis strategies of Rosen et al. (1996) or Goto et al. (1999). In the former technique the alanine, valine, leucine, and isoleucine γ2 methyls are selectively protonated, while the latter results in valine, leucine, and isoleucine δ1 methyls being selectively protonated instead. The approaches have proven facile and beneficial to solution NMR, as per Hajduk et al. (2000). For our analysis we will consider ^{13}C-^{13}C contacts measured in uniformly ^{13}C-enriched protein, ^{15}N-^{15}N couplings in uniformly ^{15}N-enriched protein, ^{1}H-^{1}H contacts in protein where the proton pool is reduced to include exchangeable protons only, and ^{1}H-^{1}H contacts in protein where the proton pool alternatively includes both exchangeable and methyl protons (as per Rosen et al., 1996).

Comparison of the pairwise internuclear distance distribution functions for the nuclei to be recoupled in these different isotopic variants of ubiquitin show them to be qualitatively different. Figures 1A, 2A and 3A show these radial distribution functions as histograms for all ^{13}C-^{13}C, ^{15}N-^{15}N and ^{1}H-^{1}H pairings respectively. As density corresponding to predictable geometry is of little interest, the ^{13}C-^{13}C 1-bond and 2-bond contacts, ^{15}N-^{15}N distances within the guanidino and imidazole groups, and geminal ^{1}H-^{1}H contacts in amino and methyl groups are denoted. That radial density which, if known, would actually help to constrain structure beyond that which is handled by common molecular dynamics restraints, is shown

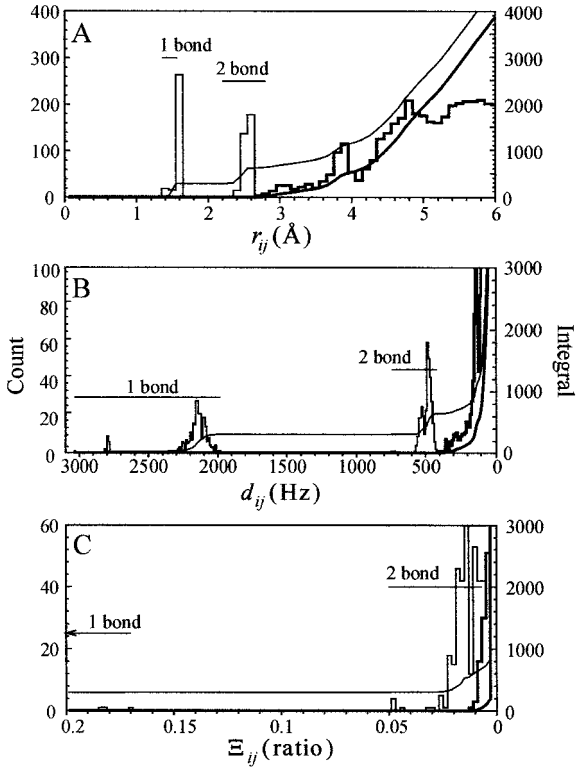

Figure 1. Distributions of calculated ^{13}C-^{13}C (A) internuclear distance, (B) dipolar coupling constants, and (C) ratios in ubiquitin. Histogram (left scale) and integral (right scale) plotted for all nuclear pairs (thin) and for pairs separated by greater than two bonds (bold). Integrals are running sums over the histograms.

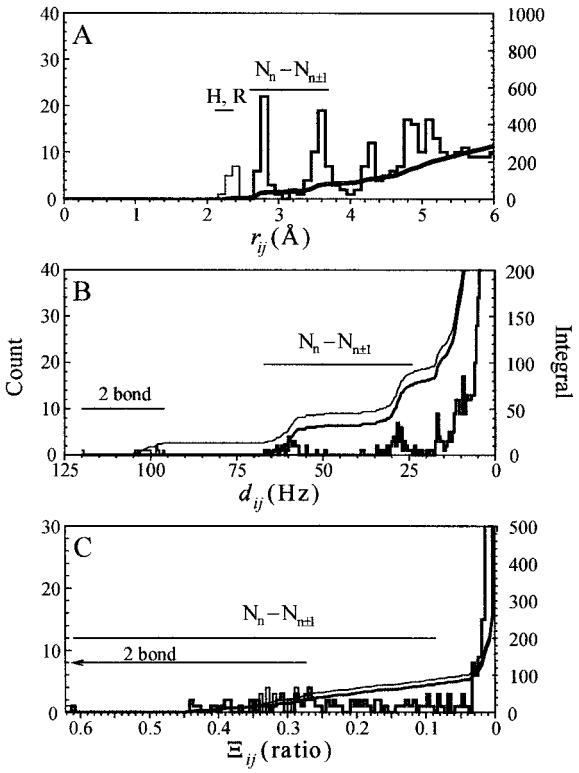

Figure 2. Distributions of calculated ^{15}N-^{15}N (A) internuclear distance, (B) dipolar coupling constants, and (C) ratios in ubiquitin. Histogram (left scale) and integral (right scale) plotted for all nuclear pairs (light) and for pairs separated by greater than two bonds (bold). Integrals are running sums over the histograms.

with thick lines. The thin versus thick integral lines provide a measure, up to any given distance, of how many internuclear contacts there are, and what fraction of them are potentially useful in defining tertiary structure, respectively.

As dipolar couplings and not internuclear distances are measured in solid state NMR, it is instructive to recast the nuclear radial distance frequency distributions as distributions of dipolar couplings. The transformation given the usual definition (in Hz) of the dipolar coupling constant d_{ij} between a spin i and a spin j is straightforward:

$$d_{ij} = \frac{\mu_o \gamma_i \gamma_j \hbar}{8\pi^2 r_{ij}^3} \text{ (Hz)}. \quad (1)$$

For a pair of protons at a separation of 1 Å this is 120.120 kHz. These distributions are shown in Figures 1B–3B for the cases under consideration. Here, one's attention should be drawn to the relative strengths of less structurally meaningful couplings, versus those that would need to be accurately measured to constrain structure calculations. Putting these numbers on a dipolar coupling scale emphasizes the compression of the scale towards zero, and that the smallest numbers are the most desirable to determine.

The distribution of internuclear dipolar couplings alone does not paint the entire picture. In a magnetization exchange experiment, transfer rates are often proportional to the square of the dipolar coupling constant (Henrichs et al., 1986; Helluy et al., 2000; Dusold et al., 1997). Further, consideration of two extremes demonstrates the need to view transfer rates in the context of their local spin systems. While the coupling for an isolated nuclear pair can be accurately measured, the coupling between a pair of spins in which each has large couplings to other nuclei is more difficult to determine. One conceivable metric of how difficult a particular coupling might be to measure is the ratio of that transfer rate to the sum of all other transfer rates experienced by the two spins involved, which we denote Ξ:

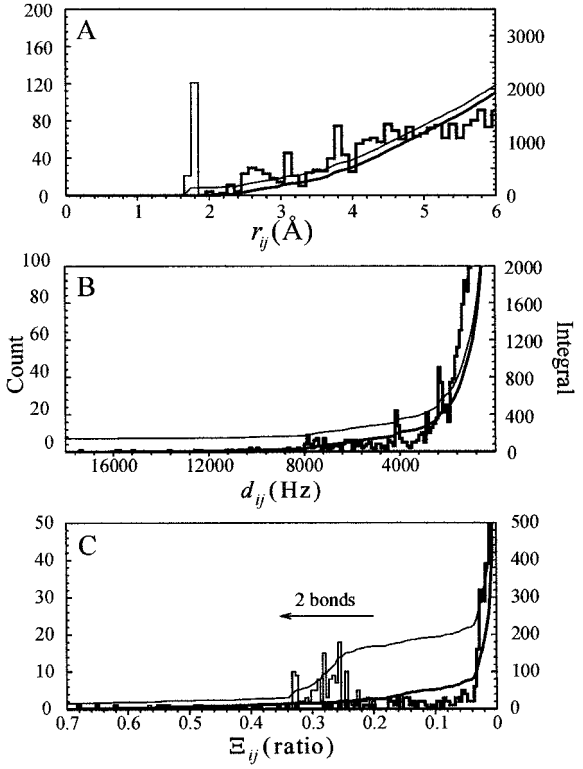

Figure 3. Distributions of calculated exchangeable plus select methyl ^{1}H-^{1}H (A) internuclear distance, (B) dipolar coupling constants, and (C) ratios in ubiquitin. Histogram (left scale) and integral (right scale) plotted for all nuclear pairs (light) and for pairs separated by greater than two bonds (bold). Integrals are running sums over the histograms.

$$\Xi_{ij} = \frac{d_{ij}^2}{d_{ij}^2 + \sum_{k \neq i,j} d_{kj}^2 + \sum_{i,j \neq k} d_{ik}^2}. \qquad (2)$$

The sums in the denominator should be restricted to some maximum distance. In practice a maximum radius can be chosen, or a single molecule or a unit cell used to limit the sum. Our conclusions are insensitive to this choice, and so a single molecule was chosen as a matter of convenience. If dipolar couplings are measured from magnetization exchange in the initial rate regime, one might expect such a ratio to correlate reasonably to the spread of the available observable magnetization over several different spin centers as the number of significant couplings increases. Expressed in this manner, the extent to which meaningful couplings are at a kinetically competitive disadvantage is then apparent. This recasting of dipolar couplings as a distribution of ratios (Ξ) is shown for comparison in Figures 1C, 2C and 3C. It will be seen later that these ratios offer a reasonable means for gauging *a priori* those couplings that one may hope to measure with useful precision for a particular enrichment pattern.

The larger the ratio, the more dominant the coupling between the two constituent nuclei. In many cases, however, structurally meaningful couplings are less dominant components of their spin systems, and occur at lower ratio values. We expect that there would be some minimum Ξ value below which a coupling will be too unreliable. Postponing justification for the moment, we use lower limits of 0.02 and 0.01 for the sake of illustration, where large typical couplings have Ξ values ranging from just under 1.0 to 0.2 depending on the particular nuclear distribution. The impact of imposing these minimum Ξ_{ij} thresholds is easily visualized using the contact maps in Figure 4. Couplings that will constrain tertiary structure appear as off-diagonal elements, while those on the diagonal define intra-residue contacts. Contacts that define secondary structure and the packing of secondary structure elements can easily be discerned in the more densely populated contact maps. Fattened stretches on the diagonal correspond to α-helix, a series of off-diagonal contacts forming a solid line segment with positive slope corresponds to parallel β-sheet, and a series of off-diagonal contacts forming a solid line segment with negative slope corresponds to anti-parallel β-sheet.

Discussion of these distributions is also aided by considering specific examples from each of the distributions. In Tables 1 and 2 the contacts for an α-helix and β-sheet sidechain carbon, backbone nitrogen, exchangeable backbone proton, and sidechain proton are specifically examined. Chosen for this purpose are lys27, the fifth residue of twelve in an α-helix, and ile13, the fourth residue of an eight residue β-strand which runs antiparallel with a neighbor strand. A relatively remote sidechain carbon is chosen in order to bias these examples towards a higher proportion of meaningful contacts, whereas a backbone nitrogen is chosen as it is backbone nitrogens which dominate the ^{15}N population. Specific contacts are given for several of the largest couplings.

^{13}C homonuclear contacts

As expected, the radial distribution of homonuclear carbon contacts in uniformly labeled ^{13}C ubiquitin is significant at distances corresponding to one and two bond separation (Figure 1A). Density increases at 3.8 Å owing to $C_{\alpha_n} - C_{\alpha_{n+1}}$ contacts. The inverse cubed

Table 1. Dipolar coupled neighbors for a selected carbon and nitrogen in an α-helix (lys27) residue of ubiquitin

	Neighbors[a]	Connectivity	d_{ij} (Hz)		Neighbors[a]	Connectivity	d_{ij} (Hz)
	^{13}C neighbors to Lys27 ^{15}N				^{13}C neighbors to Lys27 ^{13}Cγ		
	Val26 C	1 bond	1320	1	Lys27 Cδ	1 bond	2173
	Lys27 Cα	1 bond	1023	2	Lys27 Cβ	1 bond	2165
	Val26 Cα	2 bond	220	3	Lys27 Cα	2 bond	467
	Lys27 Cβ	2 bond	209	4	Lys27 Cϵ	2 bond	466
	Lys 27 C	2 bond	207	5	Lys27 C	*n*	133
	Lys27 Cγ	*n*	103	6	Leu43 Cδ1	*n*+16	127
	Val26 Cγ	*n*−1	99	7	Gln41 Cδ	*n*+14	111
	Asn25 Cγ	*n*−2	88				
	^{15}N neighbors to Lys27 ^{15}N				^{15}N neighbors to Lys27 ^{13}Cγ		
1	Ala28 N	*n*+1	60		Lys27 N	*n*	103
2	Val26 N	*n*−1	58		Lys27 Nζ	*n*	54
3	Lys29 N	*n*+2	17		Gln41 Nη2	*n*+14	30
4	Asn25 N	*n*−2	16		Ala28 N	*n*+1	23
5	Glu24 N	*n*−3	11		Val26 N	*n*−1	19
6	Ile30 N	*n*+3	11		Asn25 N	*n*−2	15
7	Gln41 Nϵ2	*n*+14	8				
	^{1}H neighbors to Lys27 ^{1}H(N)				^{1}H neighbors to Lys27 ^{1}H$_3$(Nζ)		
	Val26 H(N)	*n*−1	6119		Leu43 H(N)	n+16	2386
	Ala28 H(N)	*n*+1	6083		Ile23 H$_3$(Cγ2)	n−4	2271
	Val26 H$_3$(Cγ1)	*n*−1	2875		Arg72 H$_2$(Nη1)	n+45	2188
	Lys29 H(N)	*n*+2	1684		Asp52 H(N)	n+25	1830
	Asn25 H(N)	*n*−2	1676		Arg72 H$_2$(Nη2)	n+45	1731
	Ile23 H$_3$(Cγ2)	*n*−4	1663		Gln41 H(N)	n+14	1346
	Leu43 H$_3$(Cγ2)	*n*+16	1563		Gln49 H$_2$(Nϵ2)	n+22	1174
	Val26 H$_3$(Cγ2)	*n*−1	1356		Leu43 H$_3$(Cδ1)	n+16	1062
	Ile30 H(N)	*n*+3	1099		Arg42 H(N)	n+15	671
	Glu24 H(N)	*n*−3	1038		Leu50 H(N)	n+23	551
	Ala28 H$_3$(Cβ)	*n*+1	1028				
	Gln41 H$_2$(Nϵ)	*n*+14	708				

*Numbered nuclei correspond to peak labels in Figures 13 and 14.

relationship between internuclear distance and dipolar coupling constant causes a compression toward zero between the distributions in distance and the dipolar coupling constants (Figure 1B). Nearly every carbon in the protein has at least one other carbon at the single bond and the two bond distance. This is readily manifest upon recasting the couplings as Ξ ratios (Figure 1C). There are no meaningful homonuclear carbon contacts above a Ξ value of 0.02. Of the 4 meaningful contacts with ratio above 0.01 (all of which are less than 0.013) over 4 residue pairs (Figure 4B), all are intra-residue contacts. The maximum distance of these contacts is 2.8 Å.

Considering the specific examples in Tables 1 and 2, only the sidechain carbon of the β-strand residue has any meaningful ^{13}C-^{13}C contacts at significant relative couplings (Table 2). Meaningful contacts for the α-helix residue carbon are made at best at only 6% of that carbon's two largest couplings (Table 1).

^{15}N homonuclear contacts

In contrast to the carbon distribution, only in the arginine and histidine side chains can two nitrogen nuclei be found separated by less than three bonds (Figure 2A) in uniformly ^{15}N enriched ubiquitin. All ^{15}N-^{15}N radial density centered at 2.6 and 3.4 Å represent separation between backbone amides, which are predominantly α-helix and β-sheet, respectively. The next peak of density centered at 4.2 Å is dominated by N_n-N_{n+2} α-helix contacts, but includes tertiary contact

Table 2. Dipolar coupled neighbors for a selected carbon and nitrogen in a β-sheet (ile13) residue of ubiquitin

	Neighbors[a]	Connectivity	d_{ij} (Hz)		Neighbors[a]	Connectivity	d_{ij} (Hz)
	^{13}C neighbors to Ile13 ^{15}N				^{13}C neighbors to Ile13 ^{13}Cδ1		
	Thr12 C	1 bond	1364	1	Ile13 Cγ1	1 bond	2152
	Ile13 Cα	1 bond	1030	2	Ile13 Cβ	2 bond	474
	Thr12 Cα	2 bond	225	3	Ile13 Cγ2	*n*	306
	Ile13 C	2 bond	209	4	Lys33 Cδ	*n*+20	256
	Ile13 Cβ	2 bond	198	5	Lys33 Cγ	*n*+20	217
	Ile13 Cγ2	*n*	160	6	Glu34 Cγ	*n*+21	156
	Thr12 Cγ2	*n*−1	91	7	Ile13 Cα	*n*	130
	Thr12 Cβ	*n*−1	81		Leu15 Cδ2	*n*+2	110
	Ile13 Cγ1	*n*	63		Ile13 C	*n*	103
	Val6	*n*−8	53		Leu15 Cγ	*n*+2	93
	^{15}N neighbors to Ile13 ^{15}N				^{13}C neighbors to Ile13 ^{13}Cδ1		
1	Thr12 N	n−1	30		Thr14 N	n+1	30
2	Thr14 N	n+1	28		Ile13 Nζ	n	27
3	Val5 N	n−8	16		Lys33 Nζ	n+20	24
4	Lys6 N	n−7	11		Glu34 N	n+21	21
5	Lys11 Nζ	n−2	8		Ile30 N	n+17	18
6	Thr7 N	n−6	7		Leu15 N	n+2	16
	^{1}H neighbors to Ile13 ^{1}H(N)				^{1}H neighbors to Ile13 ^{1}H$_3$(Cγ2)		
	Ile13 H$_3$(Cγ2)	n	13348		Ile13 H(N)	*n*	13348
	Val5 H(N)	n−8	4706		Val5 H$_3$(Cγ1)	*n*−8	3726
	Val5 H$_3$(Cγ1)	n−8	1696		Val5 H(N)	*n*−8	2293
	Thr12 H(N)	n−1	1409		Lys11 H$_3$(Nζ)	*n*−2	2251
	Thr7 H(N)	n−6	1376		Ile30 H$_3$(Cγ2)	*n*+17	2097
	Thr14 H(N)	n+1	1315		Val5 H$_3$(Cγ2)	*n*−8	1290
	Lys6 H(N)	*n*−7	1048		Thr7 H(N)	*n*−6	1170
	Val5 H$_3$(Cγ2)	*n*−8	929		Leu69 H$_3$(Cδ1)	*n*+56	1167
	Thr12 H(Cγ1)	*n*−1	923		Thr7 H(Oγ1)	*n*−6	1111
	Lys11 H$_3$(Nζ)	*n*−2	721		Thr12 H(N)	*n*−1	1068
	Thr7 H(Oγ1)	*n*−6	714		Thr14 H(N)	*n*+1	891

*Numbered nuclei correspond to peak labels in Figures 13 and 14.

of no particular classification, as does the rest of the range. The distribution of calculated dipolar couplings looks similar (Figure 2B), except for the contraction of the distribution towards zero.

The bimodal distribution of couplings calculated for N_n to $N_{n\pm1}$ contacts seen in Figures 2A and 2B indicates the expected correlation between a residue's ϕ and ψ dihedral angles and the N_n to $N_{n\pm1}$ dipolar coupling, analogous to the relative coupling intensities in the NOESY fingerprint (H(N) vs. H(C_α)) region. If one constructs such a map using protein databank coordinates for a few representative proteins, the expected correlations are quite clearly observed. Rotation about ϕ has little bearing on the $N_n - N_{n+1}$ coupling, while rotation about ψ does have a reasonably strong affect. Although the $N_n - N_{n+1}$ coupling does not determine a unique combination of dihedral angles, it does discriminate between α-helices and β-sheets.

For backbone nitrogens, the strongest couplings are with the preceding and following nitrogen along the chain, and these couplings figure prominently against competing couplings in their spin systems. All but one of the 75 sequential contacts are found above $\Xi_{ij} = 0.09$ where the strongest tertiary contact is found. There are several tertiary contacts below $\Xi_{ij} = 0.09$ but still above 0.01 and 0.02 which would provide significant tertiary structure constraints while at the same time being quite measurable (at least from the standpoint of limited interference from spin diffusion).

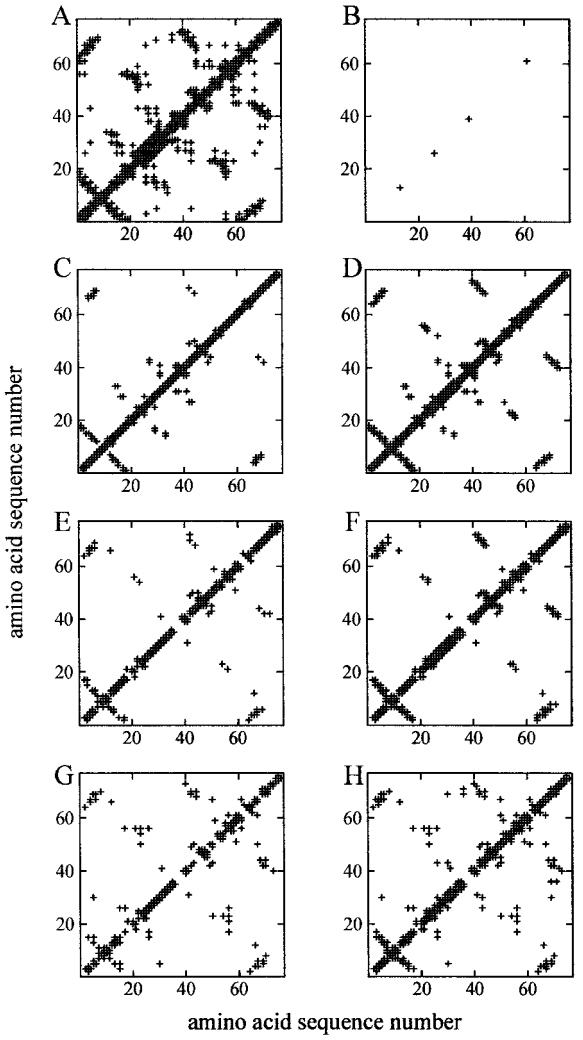

Figure 4. (A) Reference contact map showing inter-residue correlations corresponding to constituent carbons with internuclear distances of 5Å or less. (B–H) contact maps indicating residue-residue correlations corresponding to constituent internuclear contact ratios above (B) 0.01 for ^{13}C-^{13}C, (C and D) contact ratios above 0.02 and 0.01 respectively for ^{15}N-^{15}N, (E and F) contact ratios above 0.02 and 0.01 for exchangeable ^{1}H-^{1}H, and (G and H) contact ratios above 0.02 and 0.01 for exchangeable plus select methyl ^{1}H-^{1}H.

Above a ratio threshold of 0.02 (Figure 4C) there are 42 contacts making 42 non-sequential inter-residue contacts, and at a ratio threshold of 0.01 (Figure 4D) there are 98 couplings making 92 non-sequential inter-residue contacts. The maximum distance corresponding to any of the contacts above a ratio of 0.01 is 7.4 Å.

The numerous non-sequential inter-residue contacts help to define the packing of all secondary structure elements, even at ratios above 0.02. Considering the specific examples in Tables 1 and 2, it is seen that the strongest several ^{15}N couplings define the α-helix (Table 1). At lower Ξ_{ij} are tertiary contacts that indicate the packing of the helix with other secondary structures. For the β-strand (Table 2), the strand's register with other strands is indicated at fairly large couplings relative to the strongest couplings to sequential backbone nitrogens. The utility of homonuclear ^{15}N couplings to help define tertiary structure for the case of β-sheet was recently demonstrated by Castellani et al. (2002).

^{1}H homonuclear contacts

The distributions of exchangeable plus select methyl protons with respect to internuclear distance (Figure 3A), dipolar couplings (Figure 3B), and Ξ_{ij} ratios (Figure 3C) are somewhat similar (on relative scales) to the homonuclear ^{15}N case. While some ranges of couplings and ratios are dominated by one general category of contact, e.g. 4845–9360 Hz, which are primarily $H(N)_n - H(N)_{n+1}$ couplings, tertiary contacts are spread throughout the full range of dipolar constants and ratios. While the largest tertiary contact ratio is 0.68, the bulk of tertiary contacts occur at ratios less than 0.20. Of the 161 contacts above $\Xi_{ij} = 0.02$ (Figure 4G), 48 are internal to 29 residues, while there are 45 sequential and 68 non-sequential contacts among 90 unique residue pairs. For the 258 couplings above $\Xi_{ij} = 0.01$ (Figure 4H), there are 56 intra-residue contacts among 30 residues, and 74 sequential plus 128 non-sequential contacts among 145 unique residue pairs. The maximum distance corresponding to any of the contacts above a ratio of 0.01 is 5.5 Å. The numerous non-sequential inter-residue contacts help to define the packing of all secondary structure elements, even at ratios above 0.02.

If the proton pool is reduced to exchangeable protons only, the 123 contacts above a ratio of 0.02 (Figure 4E) are comprised of 14 internal to 14 residues, and 63 sequential and 46 non-sequential contacts among 99 unique residue pairs. For the 175 couplings above ratio 0.01 (Figure 4F), there are 15 internal contacts in 14 residues, and 75 sequential plus 85 non-sequential contacts among 160 unique residue pairs. The maximum distance corresponding to any of the contacts above a ratio of 0.01 is 6.2 Å.

As with ^{15}N-^{15}N couplings, the numerous non-sequential contacts help to define the packing of all secondary structure elements, even at ratios above 0.02. For homonuclear ^{1}H contacts among exchange-

able and methyl protons, we see couplings to the α-helix amide proton (Table 1) mostly define secondary structure, while many strong couplings of comparable strength to the lysine amine protons truly constrain tertiary structure. For the β-strand residue, contacts to the amide proton appear to help define the residue's relationship with another β-strand more so than its own strand (Table 2). The side chain methyl protons show these tertiary contacts as well as others, which help pack this strand against a neighbor helix.

Consequences for structure determination by solids NMR

On the basis of the Ξ_{ij} ratios, experimental approaches which employ magnetization transfer under MAS conditions among nitrogen in uniformly ^{15}N enriched protein, and ^{1}H-^{1}H spin exchange in appropriately deuterated protein, appear to offer the largest number of dipolar couplings that will accurately constrain tertiary structure, at least from the viewpoint of limited interference from spin diffusion. ^{13}C-^{13}C spin exchange experiments have serious difficulties on this qualitative basis with providing tertiary structural constraints. However their weakness in this regard is a strength in resonance assignment experiments where long range couplings would in fact be a hindrance.

Choosing among the various labeling schemes under discussion depends upon additional factors. ^{1}H-^{1}H spin exchange experiments will suffer from interference by dipolar couplings to and relay of magnetization through water, as well as poor resolution in the ^{1}H spectral dimension. This is not a problem for ^{15}N-^{15}N dipolar recoupling methods, but such experiments instead suffer from low sensitivity if ^{15}N detection is used, and perhaps more importantly, technical challenges from recoupling the small dipolar couplings one wishes to measure of only tens of Hz in size. Which set of limitations provides the lesser handicap to a large extent will depend upon the system and the questions asked.

One of course can use the qualitative analysis just presented to suggest and contrast additional isotope enrichment schemes. Dilution of ^{13}C to select sites (Castellani et al., 2002) is one of many obvious ways for preserving the sensitivity and dispersion of ^{13}C solid state NMR, while ameliorating the detrimental effects, such as dipolar truncation (Hohwy et al., 2002), that accompany uniform ^{13}C enrichment. ^{13}C/^{15}N heteronuclear dipolar coupling measurements, while not being affected by spin diffusion per se, do however suffer from dilution of available magnetic moment between interesting and uninteresting coupling partners, and this can be studied in a similar fashion. As of this writing many different isotope labeling schemes are being evaluated in a number of different laboratories, as well as efforts to use selective magnetization transfers in an effort to design pulse sequences to effectively make subsets of spins behave as more magnetically dilute than they actually are. Our attention in the next sections will instead be focused on ways to compare the relative accuracy of such competing approaches. As such we will continue to consider only the labeling patterns just discussed. Of the many experimental factors that can be involved, we will limit our attention to the influence of finite signal-to-noise and spin diffusion on the precision of the distances that may be derived from NMR magnetization exchange data in solids.

Theoretical background for a simple computational model

The kinetics of magnetization exchange in a multi-spin system under MAS conditions is largely a function of the set of internuclear couplings. Measurement of these couplings provides distance constraints that help determine the parent molecular structure. Ideally one would incorporate a full quantum mechanical simulation of the mixing or dephasing process of a specific solid state NMR experiment into a structure refinement routine. Such a procedure can be specified, and would include relaxation, motional dynamics, relevant tensor interactions, and radio frequency field inhomogeneity, among other factors. Impressive progress has been made in this respect (Bak et al., 2000). However, as a randomly ordered powder sample requires consideration of thousands of individually oriented crystallites and nuclear centers, such a complete treatment is currently difficult for even a small protein. We proceed by adapting a computational framework which captures many of the important aspects of the relevant physics, while remaining a practical approach for dealing with as much as an entire protein. Our goal is to provide a realistic analysis of the errors to be expected in measurement of distances that could be accessed in a homonuclear spin exchange experiment.

Our primary assumption supplies an efficient means to handle the direct determination of distances from complicated magnetization exchange kinetics. Insofar as the exchange process is first-order, the

problem may be modeled similarly to cross-relaxation processes of a NOESY experiment. Although a considerable simplification, the utility of the rate matrix approach has been demonstrated in application to modeling homonuclear magnetization transfer in rotating powdered solids (Dusold et al., 1997; Helluy et al., 2000; Goobes and Vega, 2002). The assumption is most accurate in the initial rate regime and with magic angle spinning rotor periods much shorter than the exchange time. Modeling the time evolution of the complex spin dynamics of a powder sample by a single orientation-independent effective trajectory is not without precedent. Very good results have also been obtained for the related problem of simulating TEDOR magnetization transfer (Mueller, 1995) between ^{13}C and ^{15}N spins in powder samples of isotopically enriched peptides (Jaroniec et al., 2002).

When the dipolar couplings become comparable to the available MAS rate such approximations are no longer tenable, and the exchange rates strongly depend upon both dipolar and shielding tensor orientations. Therefore, while we might expect this approach to reflect the spin exchange dynamics for ^{13}C nuclei under MAS with rotation rates of tens of kilohertz, ^{1}H-^{1}H spin exchange is not expected to be modeled nearly as well, unless the ^{1}H are diluted by extensive deuteration as we assume here. In our experience the simulation of ^{1}H-^{1}H spin exchange in fully protonated material requires explicit consideration of the variation of dipolar coupling with rotor position, and a summation over crystallite orientations in a powder sample. At present this is not feasible to include in the computational studies here on an entire protein molecule.

A series of secondary assumptions assigns values to the rate matrix $\mathbf{R}$ elements R_{ij}. Our calculations will assume that the exchange rate $R_{i\neq j}$ between a pair of spins i and j is simply proportional to the square of the dipolar coupling constant d_{ij}, i.e. that $R_{i\neq j} = \xi d_{ij}^2$ where ξ is a phenomenological scaling factor dependent on the recoupling efficiency, among other factors (Henrichs et al., 1986; Helluy et al., 2000; Dusold et al., 1997). If ξ is assumed to be the same for all spins involved, it simply scales the time scale of the spin dynamics. As the dynamic range of the time scale of our simulations is dictated by the largest couplings, ξ is set to a characteristic large coupling $1/d_{\max}$ for the spin system under analysis:

$$R_{i\neq j} = \frac{d_{ij}^2}{d_{\max}}. \tag{3}$$

For homonuclear carbon, we choose a one-bond 2250 Hz coupling, for homonuclear nitrogen the largest sequential backbone amide coupling of 64.5 Hz, and for homonuclear protons the largest non-intra-methyl or amino coupling of 16 kHz. We also then use the constant to express mixing time as a dimensionless quantity. A ^{13}C-^{13}C 200μs numeric mixing time then is specified as 0.45 on the τ_{2250} scale.

The diagonal elements R_{ii} are a combination of the sum of all rates between site i and sites j, and the longitudinal relaxation rate $R_{1i} = 1/T_{1i}$. If T_{1i} is long, the rates are negligible, and therefore

$$R_{ii} = -\frac{1}{d_{\max}} \sum_j d_{ij}^2. \tag{4}$$

This is reasonably accurate for all nuclei considered here except methyl protons.

These assumptions are motivated by the need for a computational model that will be accurate enough to be relevant, yet fast enough to be repeated tens of thousands of times on a system as large as small protein for the purposes of Monte Carlo analysis of errors. For situations in which these assumptions are overly simplistic, such as those caveats offered above, more parameters and more complicated treatment would be necessary. Our treatment for those cases must then be regarded as a best case scenario as additional complexity will result in even less precision than that which we estimate.

For our simulations we adapt the relaxation matrix theory of Macura and Ernst (1980) to the description of a generic 2D homonuclear solid state magnetization exchange experiment of the type $(\pi/2)_{\varphi} - t_1 - (\pi/2)_{\pm\varphi} - \tau_m - (\pi/2)$. For a solution sample this would be a standard NOESY experiment, while in a rotating solid some sort of dipolar recoupling is implied during τ_m. The purpose of the model is to relate 2D peak intensities to magnetization exchange rates and therefore the structure of the spin system. We first define an initial z-polarization vector $\mathbf{M}_z(0)$ and an equilibrium vector $\mathbf{M}_0$ as

$$\mathbf{M}_z(0) = \begin{pmatrix} M_{z_1} \\ M_{z_2} \\ \vdots \\ M_{z_n} \end{pmatrix} \text{ where } M_{z_i} = b_i M_{o_i}$$

and

$$\mathbf{M}_0 = \begin{pmatrix} M_{o_1} \\ M_{o_2} \\ \vdots \\ M_{o_n} \end{pmatrix} \text{ where } M_{o_i} = c_i M_0. \quad (5)$$

For each of n chemically distinct sites there are c_i magnetically equivalent spins, and each spin is taken to have a thermal equilibrium magnetization M_0. It is convenient to account for magnetically equivalent spins individually, and set all $c_i = 1$. The individual factors b_i account for the possibility that the spins can be variably polarized, such as when cross polarization of ^{13}C or ^{15}N is used. The $(\pi/2)_\varphi - t_1 - (\pi/2)_{\pm\varphi}$ portion of the sequence serves to produce frequency labeled z-magnetization. Assuming that all the transverse magnetization components are dephased by a gradient or otherwise suppressed after the second $\pi/2$ pulse, the z-polarization vector has real components

$$M_z^{\pm}(t_1)_i = \pm M_z(0)_i(\cos\omega_i t_1)\mathrm{e}^{-t_1/T_2}, \quad (6)$$

where it is assumed that transverse relaxation during t_1 can be accounted for by a single T_2 for each type of spin i.

This frequency-labeled magnetization mixes according to the partial reinstatement of the dipolar coupling during τ_m to an extent that depends upon the specific pulse sequence employed, in addition to other factors. For the isotope enrichment patterns investigated here, exchange due to the scalar coupling is not included as it is either not important, only becomes so for small couplings which are not reliably measured, or can be suppressed by suitable modification of the pulse sequence. The evolution of the magnetization during τ_m is best cast in terms of a difference-magnetization vector defined as $\Delta\mathbf{M}_z(t_1, \tau_m) = \mathbf{M}_z(t_1, \tau_m) - \mathbf{M}_o$. This vector follows the usual system of coupled differential equations(Macura and Ernst, 1980)

$$\frac{\mathrm{d}\Delta\mathbf{M}_z}{\mathrm{d}t} = -\mathbf{R}\cdot\Delta\mathbf{M}_z. \quad (7)$$

The solution to Equation 7 can be written in the usual fashion as

$$\Delta\mathbf{M}_z(t_1, \tau_m) = \mathrm{e}^{-\mathbf{R}\tau_m}\cdot\Delta\mathbf{M}_z(t_1, 0) = \mathbf{P}(\tau_m)\cdot\Delta\mathbf{M}_z(t_1, 0). \quad (8)$$

The difference-magnetization vector for the set of spins evolves according to the spin-exchange propagator $\mathbf{P}(\tau_m)$, which is the indicated exponential function of the relaxation matrix $\mathbf{R}$. Reading the magnetization out at the end of the mixing period with a y-phase $\pi/2$ pulse, the signal is proportional to $\mathbf{M}_x(t_1, \tau_m, t_2)$ which has components

$$M_x^{\pm}(t_1, \tau_m, t_2)_i = M_z^{\pm}(t_1, \tau_m)_i(\cos\omega_i t_2)\mathrm{e}^{-t_2/T_2}, \quad (9)$$

where

$$M_z^{\pm}(t_1, \tau_m)_i = \pm\sum_j P(\tau_m)_{ij}M_z(t_1)_j - \sum_j P(\tau_m)_{ij}M_{o_j} + M_{o_i}. \quad (10)$$

The $\pm$ corresponds to the experiments where the projection pulse is of phase $-\phi$ or $+\phi$. As long as we assume this phase alternation is performed with alternate addition and subtraction of the resulting signals, the last two terms in Equation 10, which would produce axial peaks, are eliminated. The resulting 2D free induction decay $\mathrm{h}(t_1, \tau_m, t_2)$, normalized to unit intensity for a magnetization of magnitude, M_0, has n components given by

$$\mathrm{h}(t_1, \tau_m, t_2)_i = \left(\frac{M_{x_i}^+ - M_{x_i}^-}{2M_{o_i}}\right) = (\cos\omega_i t_2)\mathrm{e}^{-t_2/T_{2i}}\times \left\{\sum_j c_j b_j \cdot P(\tau_m)_{ij}\cdot(\cos\omega_j t_1)\mathrm{e}^{-t_1/T_{2j}}\right\}. \quad (11)$$

Upon 2D Fourier transformation, the 2D spectrum is a product of 1D lines, $g_{ji}(\omega_{1j}, \omega_{2i}) = g_1(\omega_j)g_2(\omega_i)$, where the line shape $g_1(\omega_j)$ is centered in ω_1 at ω_j, with a peak height and width determined by T_2. The integrated intensity $S(\omega_j, \tau_m, \omega_i)$ of the peak at the frequency coordinates $(\omega_1, \omega_2) = (\omega_j, \omega_i)$ is therefore

$$S(\omega_j, \tau_m, \omega_i) = c_j b_j \cdot P(\tau_m)_{ij}. \quad (12)$$

If one indexes the spins individually so that all the c_j are unity, and takes experimental care to ensure that the spins are uniformly polarized so that all the $b_j = b$, there is a one-to-one correspondence between matrix elements of the propagator and the integrated intensities of the peaks in the 2D NMR spectrum. Writing the matrix of all the auto-peak (S_{ii}) and cross-peak (S_{ij}) integrated intensities as $\mathbf{S}(\tau_m)$, we see that the 2D exchange spectrum experimentally determines the propagator to within a multiplicative factor at the time τ_m

$$\frac{1}{b}\mathbf{S}(\tau_m) = \mathbf{P}(\tau_m). \quad (13)$$

Thus, for individually counted nuclei of equal initial polarization and which experience negligible T_1 relaxation, the 2D spectrum and the corresponding $\mathbf{S}(\tau_m)$ matrix will be symmetric and proportional to $\mathbf{P}(\tau_m)$.

It is also important to note that for all longitudinal relaxation rates R_{1_i} equal to a single value R_1, explicit series summation gives

$$\sum_j P(\tau_m)_{ij} = e^{-R_1\tau_m}. \tag{14}$$

Thus for negligible R_1 the sum of all elements across each row vector in $\mathbf{P}(\tau_m)$ is one, and the sum over all matrix elements in $\mathbf{P}(\tau_m)$ is n.

Just as in solution NMR refinement procedures which incorporate relaxation matrix analysis (Olejniczak et al., 1986; Borgias and James, 1988, 1990; Shriver and Edmondson, 1994), a simple analytic transformation exists to convert $\mathbf{S}(\tau_m)$ to the experimentally-derived relaxation matrix $\mathbf{R}$ if the NMR data has been normalized to satisfy $\mathbf{S}(\tau_m) = \mathbf{P}(\tau_m)$. Operationally this entails finding the magnetization vector eigenbasis to diagonalize $\mathbf{S}(\tau_m)$, dividing the natural logarithm of the eigenvalues by τ_m, and transforming the resultant matrix back to the original magnetization vector basis:

$$\mathbf{R} = \frac{1}{\tau_m}\ln[\mathbf{S}(\tau_m)]. \tag{15}$$

Connection to experiment

When $\mathbf{S}(\tau_m)$ is derived from actual experimental data, this inversion suffers from finite signal-to-noise (S/N) which introduces errors into the determination of rate matrix elements R_{ij}, and subsequent distances r_{ij}. We incorporate this noise into our analysis by adding a matrix of finite random noise $\mathbf{N}$ to the magnetization matrix which has theoretically infinite signal-to-noise (bounded only by machine precision):

$$\mathbf{S}^N(\tau_m) = \mathbf{P}(\tau_m) + \mathbf{N}. \tag{16}$$

Computationally it is convenient to compute the propagator $\mathbf{P}(\tau_m)$ from a structure, and to add Gaussian random noise specified by a root mean square (r.m.s.) amplitude to generate a realistic 'experimental' data set $\mathbf{S}^N(\tau_m)$. Since our assumptions in essence put the intensity of all auto peaks to 1 at $\tau_m = 0$, the r.m.s. amplitude of the added noise specifies the signal to noise in the 2D spectrum.

The next step to making connection with experimental reality is the relationship between the peak-to-peak signal-to-noise (S/N_{ptp}) measured in a 1D spectrum, and the computationally more convenient r.m.s. signal-to-noise (S/N_σ) in the 2D spectrum. We will assume that the 2D data set is a collection of M free induction decays (FID), acquired using a t_1 increment of Δt. Fourier transform of the directly detected t_2 dimension yields a series of M spectra. For simplicity we consider a spectrum with a single line at zero frequency which decays in t_1 as e^{-t_1/T_2}, and that $S/N_{ptp} = \alpha/\beta$ at $t_1 = 0$. In the indirect dimension the signal-to-noise of the slice through this peak at zero frequency is simple to determine, remembering that the Fourier transform amplitude at zero frequency is the integral of the time domain signal. Summing the noise gives a noise amplitude $\sqrt{M}$ larger, while summing the signal gives a result dependent upon T_2. The resultant S/N_{ptp} of the slice in the 2D spectrum is thus

$$(S/N_{ptp})_{slice} = \frac{\alpha\sum_{n=0}^{M-1}(e^{-\Delta t/T_2})^n}{\beta\sqrt{M}} = \frac{\alpha}{\beta\sqrt{M}} = \frac{(e^{-\Delta t/T_2})^M - 1}{e^{-\Delta t/T_2} - 1}, \tag{17}$$

where the summation has been evaluated using the familiar rule for the sum of a geometric series. This can be cast in a more compact form by expressing the t_1 acquisition time in units of T_2 using $M\Delta t = QT_2$ to give

$$\frac{S}{N_{ptp}} = \frac{\alpha}{\beta\sqrt{M}}\frac{e^{-Q} - 1}{e^{-Q/M} - 1} \approx \frac{\alpha}{\beta}\frac{(1 - e^{-Q})\sqrt{M}}{Q}. \tag{18}$$

The last step follows the reasonable assumption that M is many times Q. Thus the S/N_{ptp} of a slice in a 2D spectrum acquired using M points in t_1 is $\sqrt{M}$ times the signal-to-noise α/β of the spectrum in the first t_1 point, multiplied by a correction factor of $(1 - e^{-Q})/Q$ to account for the decay of the signal during t_1. For a typical experiment Q might equal 3 and $M = 1024$, the slice through the $\tau_m = 0$ auto peak has a $S/N_{ptp} \sim 10$ times that of the spectrum from the first t_1 point. Hence, if a signal for a single line in the first t_1 point has a S/N_{ptp} of 10, the S/N_{ptp} of the auto peak will be ~100 with 1024 t_1 values, which we take to be a reasonable experimental S/N_{ptp} level to achieve.

As a point of calibration we quote some numbers for solid state NMR sensitivity from our experience with a modern 800 MHz instrument. Using a sample occupying only 6.5 µl, i.e., less than 4 mg of protein, the ^{13}C resonance for a single methyl group at *natural*

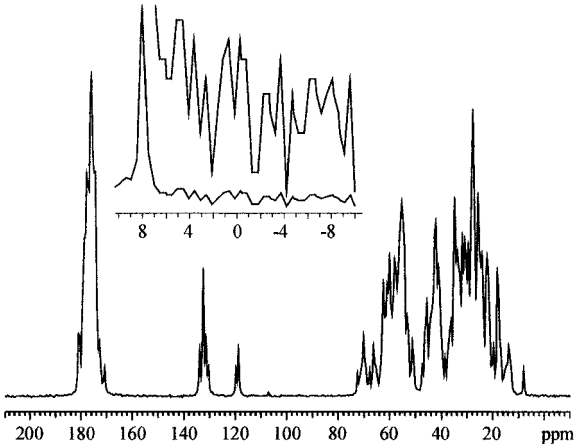

Figure 5. Signal-to-noise of 10 for 1D ^{13}C spectrum of ubiquitin. Inset: magnification of the ile23 methyl carbon at 7.8 ppm. The noise to the right of this peak is shown in the inset on a 10× scale.

abundance in hen egg white lysozyme (15 kD) can be observed in a 1D experiment with a S/N_{ptp} of 3:1 in 80 000 scans (33 h of signal acquisition, line width ~0.3 ppm). This translates to a S/N_{ptp} of 100:1 in the 2D plane for a 100% ^{13}C-enriched sample in a similar amount of time using 86 scans per each of 1024 t_1 points with 1.5 s recycle delay.

To connect to our simulations the S/N_{ptp} must be recast in terms of an r.m.s. description of the noise, S/N_σ. When the fluctuation of noise values is specified as a statistical standard deviation of a Gaussian distribution, it is empirically observed that the two formulations are related by a factor of five (Ernst et al., 1992):

$$N_{ptp} = 5 \times N_\sigma. \tag{19}$$

Therefore when a signal-to-noise value specified as S/N_σ of 500 is used in our computations, the reader may correlate this to an experimental S/N_{ptp} through a slice in the 2D spectrum of 100, or approximately 10 in the 1D spectrum corresponding to the first point in t_1. Figure 5 shows a simulated 1D ^{13}C spectrum of ubiquitin at this noise level.

Simulations, measured rates, and corresponding distances

The calculations described below require repeated transformation of the magnetization matrix into and out of its eigenbasis to yield a rate matrix (Equation 15), a computation that scales as n^3. In some instances we have reduced the set of atoms used in the simulations to those of representative molecular segments to achieve more manageable matrix sizes. For the ^{15}N and exchangeable-only proton cases, all 105 nitrogen nuclei and all 137 such protons respectively in ubiquitin are used. In the ^{13}C case the calculations include all 101 carbon nuclei from residues 1 to 20 of ubiquitin (MQIFVKTLTGKTITLEVEPS), a molecular segment which folds as two anti-parallel β-strands linked by a Type I turn. Simulations on exchangeable plus methyl protons use the 95 such protons from residues 1 to 30 (1-20+DTIENVAKIQ) which adds an α-helix with close contacts to the previously described β-sheet. For expediency methyls are treated as static triads of distinct protons.

To investigate the effect of finite signal-to-noise and internuclear distribution on the precision of distances derived from an experimental 2D exchange spectrum, simulations were carried out using the above computational model, summarized in Figure 6, on the molecular segments noted above. All code and simulations were written in C/C++ and Perl with a heavy dependence on the Perl Data Language package run under Redhat Linux 7.1. A PDB structure file (including H atoms generated as described earlier) is used as input to routines that generate the rate matrix **R**. The rate matrix is used to compute the propagator $\mathbf{P}(\tau_m)$ for each of a series of mixing times at several noise levels. Experimental data is simulated by adding scaled Gaussian random noise. This is achieved by generating a Gaussian random number (gran) with a normal probability distribution (mean = 0, standard deviation = 1) and scaling it to fit the desired S/N_σ:

$$N_{ij} = \frac{\text{gran}}{S/N_\sigma}. \tag{20}$$

Having been corrupted by the addition of noise, the simulated data must be scaled as if from an authentic experiment. We first average about the diagonal to render the matrix symmetric, and renormalize the result $\mathbf{S}^\dagger(\tau_m)$ according to earlier discussion so that the sum of all elements equals the number of spins n:

$$S^\dagger(\tau_m)_{ij} = \frac{n}{2} \cdot \frac{S^N(\tau_m)_{ij} + S^N(\tau_m)_{ji}}{\sum_{r=1}^{n}\sum_{s=1}^{n} S^N(\tau_m)_{rs}}. \tag{21}$$

The 'measured' rate matrix $\mathbf{R}^\dagger$ is then calculated as per Equation (15) from the adjusted simulated data:

$$\mathbf{R}^\dagger = \frac{1}{\tau_m} \ln[\mathbf{S}^\dagger(\tau_m)]. \tag{22}$$

Experimental rates are translated to experimental distances according to Equations 1 and 3, and relative fractional differences from true rates and true internuclear distance are calculated by

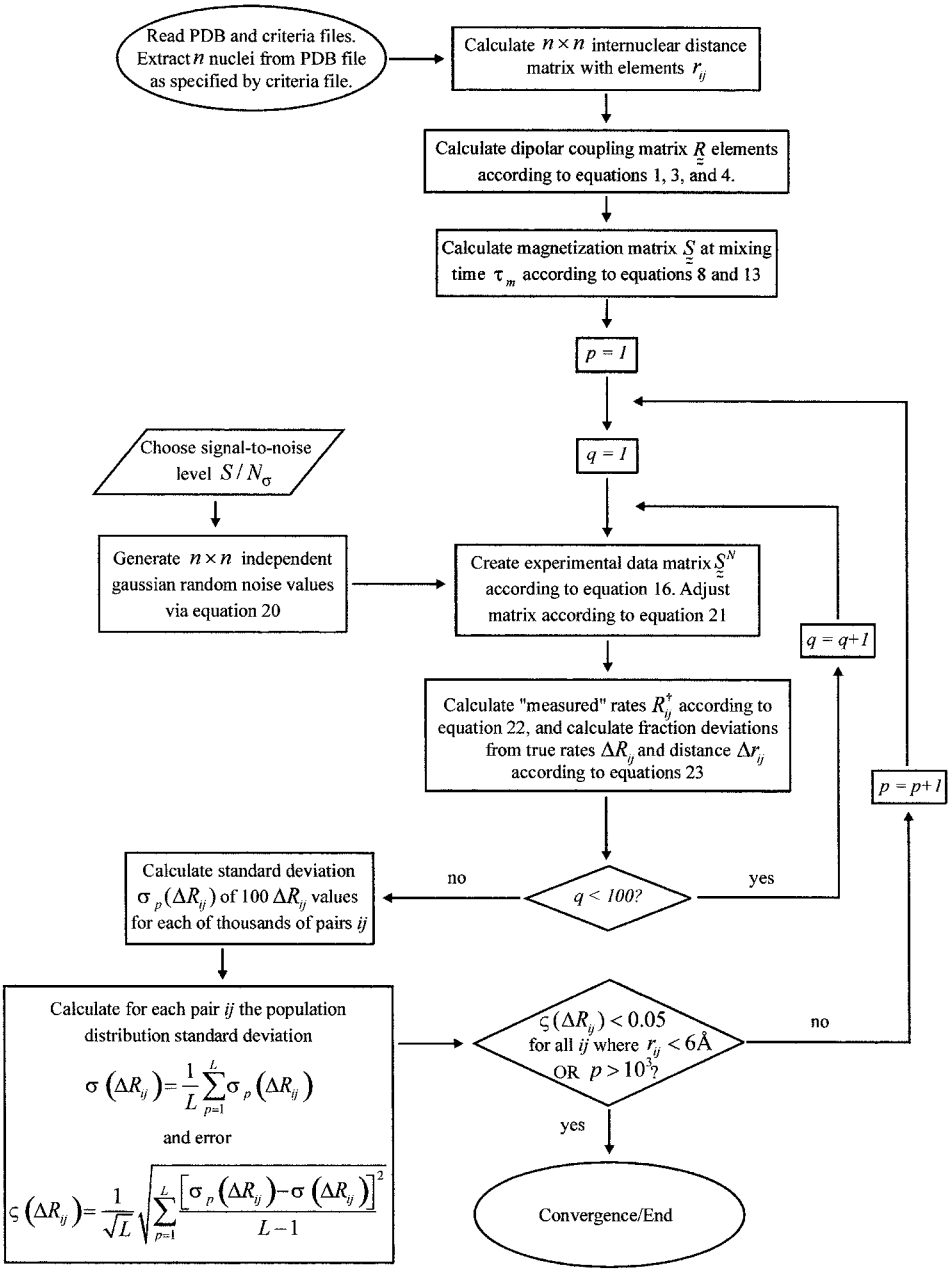

Figure 6. Flowchart outlining the computational procedure used in Monte Carlo determination of the precision of dipolar coupling rate measurements.

$$\Delta R_{ij} = \frac{R_{ij} - R_{ij}^{\dagger}}{R_{ij}} \quad \text{and} \quad \Delta r_{ij} = \frac{r_{ij} - r_{ij}^{\dagger}}{r_{ij}}. \quad (23)$$

Evaluating how signal-to-noise affects the precision of rates and the derived distances from the preceding computation is achieved by performing this calculation repeatedly, reseeding the added noise on each repetition in a Monte Carlo fashion. This results in a distribution of values ΔR_{ij} and Δr_{ij} for each pair ij. Monte Carlo cycles are run 10^5 times, unless the error $\varsigma(\Delta R_{ij})$ (see Figure 6) in the standard deviation for all distributions $\sigma(\Delta R_{ij})$ corresponding to contacts where $r_{ij} < 6$ Å falls below 0.05 sooner.

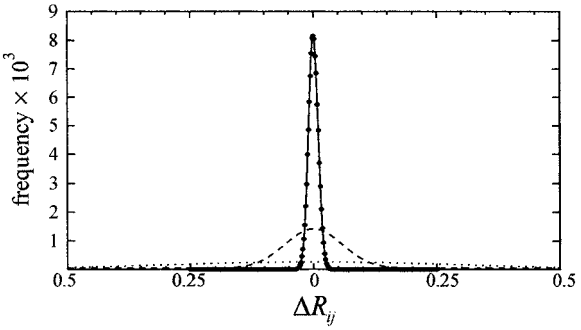

Figure 7. Sample distributions of ΔR_{ij} corresponding to three different internuclear distances r_{ij} for ^{15}N-^{15}N at $S/N_\sigma = 500$ and mixing time of 0.48 $\tau_{64.5}$. Data and fit for 2.93 Å (– • –), and fits only for 3.96 Å (- - -), and 5.06 Å (· · ·).

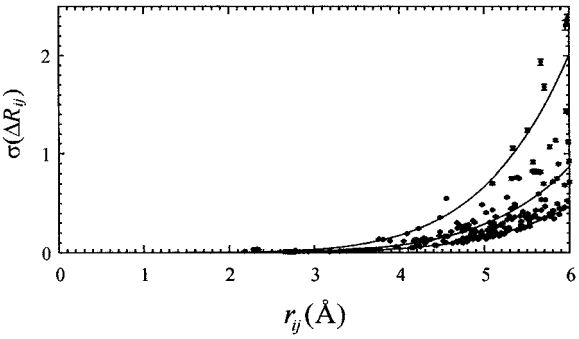

Figure 8. Standard deviation in distribution of relative measured rates $\sigma(\Delta R_{ij})$ vs. internuclear distance r_{ij} for ^{15}N-^{15}N at $S/N_\sigma = 500$ and mixing time of 0.48$\tau_{64.5}$. Curves are fit empirically to scaling factors η_{avg} for all data, η_{low} to the data selected as defining the lower boundary, and η_{up} to the data selected as defining the upper boundary. Error bars are $\pm\varsigma(\Delta Rij)$ (see Figure 6).

Computational results

The data generated consists of a distribution of 'measured' fractional rate differences ΔR_{ij} and corresponding fractional distance differences Δr_{ij} for each of several thousand nuclear pairs ij. We make no advance assumptions regarding the probability distributions, but empirically find, as one might expect from classical error analysis (Macura, 1995) based upon analytic matrix gradients (Dellwo and Wand, 1993), that the distributions fit gaussian profiles centered essentially at zero with standard deviations $\sigma(\Delta R_{ij})$ equal to that converged upon as per the flow chart in Figure 6:

$$f(\Delta R_{ij}) = \frac{1}{\sigma(\Delta R_{ij})\sqrt{2\pi}} e^{\frac{1}{2}\left(\frac{\Delta R_{ij}}{\sigma(\Delta R_{ij})}\right)^2}. \qquad (24)$$

This is true in general for the entire range of spin systems investigated. Figure 7 shows some examples of these distributions for homonuclear ^{15}N contacts of approximately 3, 4 and 5 Å. In accord with previous observations in solution experiments, pairs of spins separated by longer distances and having smaller cross peak intensities correspond to larger fractional errors in measured rates (Borgias and James, 1990; Suri and Levy, 1995).

The standard deviations $\sigma(\Delta R_{ij})$ provide a convenient single parameter with which to comprehensively characterize each empirical $f(\Delta R_{ij})$ distribution. We therefore focus upon $\sigma(\Delta R_{ij})$ in the continuing analysis, but briefly note the correspondence of these qualities in the Δr_{ij} dimension. The analytic relationship between a given $\sigma(\Delta R_{ij})$ and the corresponding $\sigma(\Delta r_{ij})$ holds up to $\sigma(\Delta R_{ij})$ of approximately 0.25:

$$\sigma(\Delta r_{ij}) = \tfrac{1}{6}\sigma(\Delta R_{ij}) + \tfrac{7}{72}\sigma(\Delta R_{ij})^2, \qquad (25)$$

where standard error propagation has been employed with a higher order term (Bevington and Robinson, 1991). At larger error values, the asymmetry of the Δr_{ij} distributions inherent in $R_{ij} \propto r_{ij}^{-6}$ translates to a departure from Gaussian with respect to both mean and standard deviation.

It is informative to plot the correlation of $\sigma(\Delta R_{ij})$ to internuclear distances for all nuclear pairs at each signal-to-noise and mixing time value. Regardless of the particular nuclear distribution, mixing time, or noise level under study, the result is a plot similar to that shown in Figure 8. Again, we make no assumptions that this correlation should take any particular functional form, but the data is found to empirically fit the simple function $\eta_{\text{avg}} r_{ij}^6$ over short to intermediate internuclear distances. The residuals of $\sigma(\Delta R_{ij})$ values about the average become larger as r_{ij} increases, an important quality not accounted for by η_{avg}. Therefore scaled r_{ij}^6 curves are also fit to data selected to define the upper and lower boundaries, with proportionality constants η_{up} and η_{low} respectively. A data point is identified as defining the lower boundary if it is the minimum of the following ten data points inclusive, and is identified as defining the upper boundary if it is the maximum of the previous ten points inclusive. Alternatively, on a log/log plot, this data forms a linear band with uniform width and slope of ~6. The three η scaling factors then correlate to the y-intercept of the three parallel lines fit to all data, and the upper and lower edges of the data band.

Closer inspection reveals that the data which defines the upper versus lower boundaries of the data

envelope correspond respectively to contacts between nuclei in relatively dense (e.g., C_α) versus dilute (e.g., carbonyl and termini of side chains) local densities. The broadening of the data envelope, i.e., the difference in $\sigma(\Delta R_{ij})$ between dilute nuclei versus tightly clustered nuclei, is more profound at longer mixing times and lower signal-to-noise levels. These three parameters fit least well at very early mixing time when the data envelope is most narrow. As mixing time increases, the upper boundary becomes less well defined and is hence less well characterized by the single parameter η_{up}. This characterization nevertheless provides a reasonable means to condense the effects of noise and mixing time on the precisions of thousands of exchange-matrix elements resulting from the Monte Carlo analysis.

Figures 9–11 show in panels C the time-evolution of η_{avg}, η_{low} and η_{up} aligned with the time-evolution of selected auto peaks (panels A) and cross peaks (panels B) for ^{13}C-^{13}C, ^{15}N-^{15}N and ^{1}H-^{1}H homonuclear spin exchange experiments respectively. The more rapid cross peak build-up and concomitant rapid autopeak decay curves correspond to the most tightly coupled pairs of spins. Bunching of the magnetization kinetics curves is pronounced for the ^{13}C-^{13}C plot (Figure 9) and the ^{1}H-^{1}H (Figure 11) which mirror the discrete clustering of nuclear distribution in Figures 1A and 3A. The corresponding time courses for ^{15}N-^{15}N magnetization kinetics are more evenly distributed. While the plot in Figure 10B would indicate gaps in the ^{15}N build-up kinetics, this is an artifact of the particular set of nuclei chosen and is not seen when a larger sample of ^{15}N centers are chosen. The distinct gap seen in Figure 9B persists no matter how large the set of ^{13}C nuclei. At longer mixing times the magnetization matrix elements approach a common value as spin diffusion evenly distributes all magnetization. The eigenvalues of $\mathbf{S}^{\dagger}(\tau_m)$ then become ill-defined and a unique rate matrix becomes impossible to determine by Equation 22.

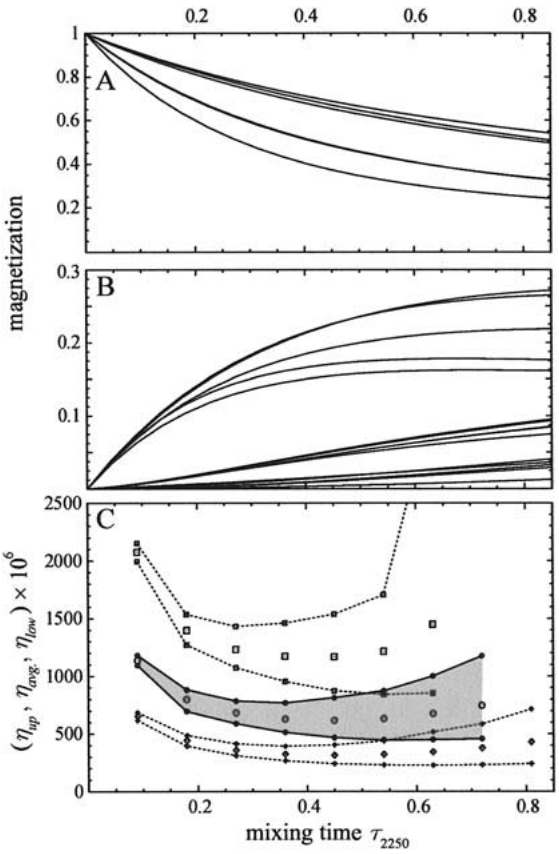

Figure 9. ^{13}C-^{13}C magnetization curves and precision of couplings. (A) Autopeak magnetization and (B) crosspeak magnetization for carbon in ile13, and (C) precision of meaningful couplings vs. duration of mixing time for all carbons in residues 1–20. Scaling factors η for all mixing durations and noise levels used: (□) $S/N_\sigma = 250$, (○) $S/N_\sigma = 500$ (shaded), and (◊) $S/N_\sigma = 1000$. Upper (η_{up}) and lower (η_{low}) boundary scaling factors are plotted with lines drawn through the respective data style.

Consequences for structure determination by solids NMR

Interpretation of the Monte Carlo results and analysis provide a more substantial basis upon which to compare the merits of different isotopic labeling schemes than the qualitative assessment offered in the first section of this paper. How does variability in nuclear spatial distribution and finite signal-to-noise affect the measurement of rates, and how can we approximate these effects for other labeling schemes without weeks of computational work?

With respect to the raw data, the solution of Equation 22 frequently results in negative values for 'measured' rates for nuclear pairs with low signal-to-noise. While mathematically valid, these values correspond to non-physical $f(\Delta R_{ij})$ density greater than 1, and do not correspond to real Δr_{ij}. Similarly, ΔR_{ij} values less than -1 correspond to 'measured' rates that are greater than double their true rates. Distributions with standard deviations $\sigma(\Delta R_{ij})$ over 0.25 have significant density outside of the -1 to 1 limits, and therefore correspond to rates not reliably measured.

For nuclear pairs with $\sigma(\Delta R_{ij})$ less than 0.25, characterization of each Gaussian distribution $f(\Delta R_{ij})$ by its standard deviation $\sigma(\Delta R_{ij})$ permits convenient

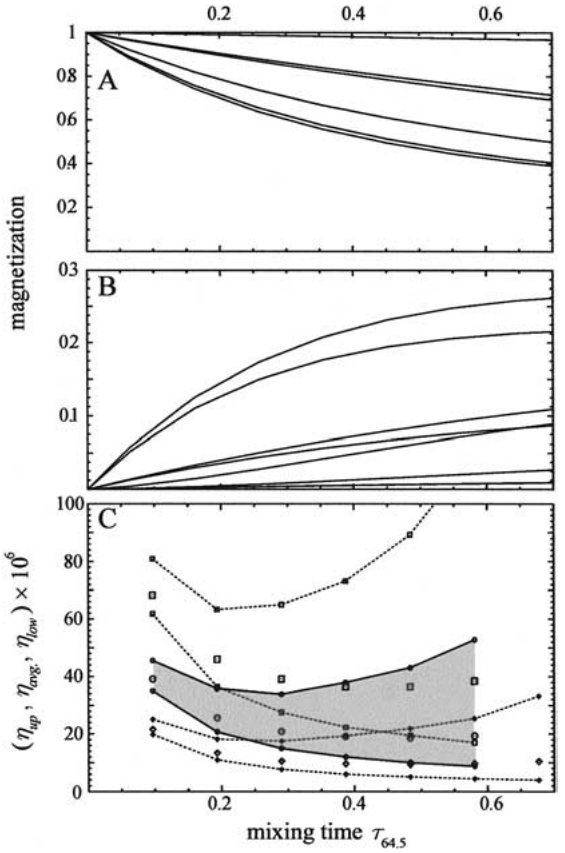

Figure 10. ^{15}N-^{15}N magnetization curves and precision of couplings. (A) Autopeak magnetization and (B) crosspeak magnetization for nitrogen in residues 9–13; and (C) precision of meaningful couplings vs. duration of mixing time for all nitrogen centers. Scaling factors η for all mixing durations and noise levels used: (□) $S/N_\sigma = 250$, (○) $S/N_\sigma = 500$ (shaded), and (◊) $S/N_\sigma = 1000$. Upper (η_{up}) and lower (η_{low}) boundary scaling factors are plotted with lines drawn through the respective data style.

use of the error function to apply confidence levels to different finite deviation limits. The measured rate for an internuclear contact with a true rate R_{ij} has 68% probability of lying within $\pm\sigma(\Delta R_{ij})R_{ij}$ and 95% probability of lying within $\pm 2\sigma(\Delta R_{ij})R_{ij}$. For example, at 500 S/N_σ level and $0.48\tau_{64.5}$ mixing time, a homonuclear nitrogen contact at 4 Å might have a $\sigma(\Delta R_{ij})$ value of 0.1, and measured distances would be 68% likely to lie between 3.94 and 4.07 Å, and 95% likely to lie between 3.88 and 4.15 Å, with the expectation value being essentially the true value. Meanwhile, a homonuclear carbon contact at the same signal-to-noise, analogous mixing time ($0.45\tau_{2250}$), and same internuclear distance might have $\sigma(\Delta R_{ij})$ equal to 2.5, and would not likely be observed or reliably measured.

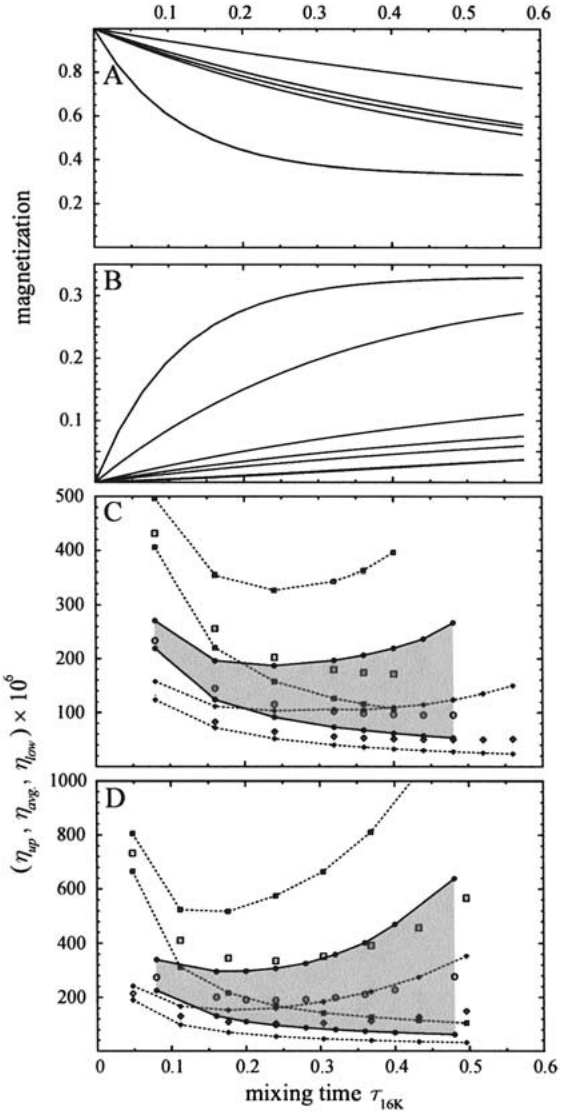

Figure 11. Exchangeable ^{1}H-^{1}H magnetization curves and precision of couplings. (A) Autopeak magnetization and (B) crosspeak magnetization for exchangeable protons in residues 9–11; precision of meaningful couplings vs. duration of mixing time for (C) all exchangeable protons and (D) all exchangeable plus select methyl protons in residues 1–30. Scaling factors η for all mixing durations and noise levels used: (□) $S/N_\sigma = 250$, (○) $S/N_\sigma = 500$ (shaded), and (◊) $S/N_\sigma = 1000$. Upper (η_{up}) and lower (η_{low}) boundary scaling factors are plotted with lines drawn through the respective data style. Note that the protons whose magnetization curves are plotted in (A) and (B) are common to both the pools of protons and molecular segments used, and the build-up curves for these nuclei are nearly identical for both cases.

Figures 9C, 10C, 11C and 11D summarize the hundred analyses exemplified by Figure 8 over all noise levels, mixing times, and chosen labeling schemes. We see, in accord with analogous observations in solution NMR (Olejniczak et al., 1986), that at short mixing durations limited signal-to-noise blurs the variation in crosspeak intensities which hold all the information and leads to large errors in measured rates. Long range, small couplings are the most affected, where the variation in precision becomes even greater.

At longer mixing times, data imprecision increases with an attendant increase in difference between η_{up} and η_{low}. Several observations show that these phenomena are a direct consequence of spin diffusion. First is the correlation between the broadening and the degree to which indirectly transferred magnetization becomes an increasing component of a given nuclear pair's build-up of magnetization at longer mixing times, as indicated by a departure from initial linearity in Figures 9–11, panel B. This correlation is further substantiated by identifying the couplings which define the upper and lower boundaries of the $\sigma(\Delta R_{ij})$ vs. r_{ij} correlations (Figure 8). The scaling factors η_{up} (plotted as the upper line of each S/N_σ set in Figures 9C, 10C, 11C and 11D) reach a minimum at a relatively early mixing time, consonant with the initial onset of relayed magnetization for more densely packed nuclear pairs. Comparatively, the factors η_{low} (plotted as the lower line of each S/N_σ set in Figures 9C, 10C, 11C and 11D) reach a minimum at longer mixing time, consonant with a later onset of relayed magnetization for more dilute nuclear pairs. Indeed, the phenomenon is most pronounced by noting proton precisions and timings when methyl protons are included (Figure 11D) in contrast to when only exchangeable protons are used (Figure 11C). Not only does the entire set of measured distances gain precision when methyl protons are left out, but the difference between the upper and lower boundaries is less, and mixing time can be extended to longer duration.

The average scaling factors η_{avg} (see Figures 9C, 10C, 11C and 11D) then appear to be a function of both signal to noise, as well as the degree to which indirectly transferred magnetization contributes to a given internuclear contact's crosspeak growth. At early mixing times the precision is poor, owing to low signal to noise of experimental data, whereas at longer mixing times crosspeaks are more intense but the precision is poor owing to spin diffusion. Precision is best between the two regimes where a balance between the two factors is reached, and this is deemed the optimum mixing time. These optimum mixing times are $0.45\tau_{2250}$ for homonuclear carbon, $0.48\tau_{64.5}$ for homonuclear nitrogen, $0.4\tau_{16k}$ for homonuclear exchangeable only protons, and $0.24\tau_{16k}$ for homonuclear exchangeable plus select methyl protons. However, as discussed by Macura (1994), it may be appropriate, particularly when signal to noise is low, to perform mixing experiments at several mixing times so as to achieve the lowest error possible for all nuclei in a spin system.

This leads to one of our principle conclusions. A comparison across Figures 9C, 10C, 11C and 11D reveals the relative precisions of measurement afforded by each isotopic labeling scheme. Considered at the optimum mixing time and common noise levels, η_{avg} serves as an indicator of relative precisions of rate measurement for all nuclear pairs at short ($\sim$2Å) to intermediate (4–6 Å) distance. At each signal-to-noise level used, η_{avg} for homonuclear ^{15}N is 34-fold lower than η_{avg} for the homonuclear ^{13}C analysis. Similarly η_{avg} for exchangeable protons is 6.7-fold lower than that for ^{13}C, and η_{avg} for exchangeable plus select methyl protons is only 3.5-fold lower.

It is precisely the differential local densities and their effect on the precision of rate measurement that we wished to address with the recasting of internuclear contacts as the previously defined ratios Ξ. We indeed find a tighter, albeit nonlinear correlation (plotted in Figure 12), between the calculated $\sigma(\Delta R_{ij})$ and these ratios, than we do with either the internuclear distances (such as in Figure 8) or the coupling constants (not shown). This is likely grounded in the fact that the ratios are more closely correlated to the dispersal of magnetization among coupled groups of spins than either the internuclear distance or dipolar coupling constant alone indicate. In fact, the better correlations in Figure 12 show the more densely grouped points define a lower boundary $\sigma(\Delta R_{ij}) = (\Xi \cdot S/N_\sigma)^{-1}$.

This correlation is best for the uniformly labeled ^{13}C and ^{15}N homonuclear experiments, and weakest for the homonuclear experiments among exchangeable protons, improving somewhat when select methyl protons are added to the pool. Closer inspection of the exchangeable proton pairs above a ratio Ξ of 0.02 and above $\sigma(\Delta R_{ij})$ of 0.25 in Figure 12C, indicates that these contacts are either arginine side chain protons which are not included in the reduced molecular segment used to generate Figure 12D, or are exchangeable protons which are pushed to much lower ratio values when select methyls are included (most dramatically

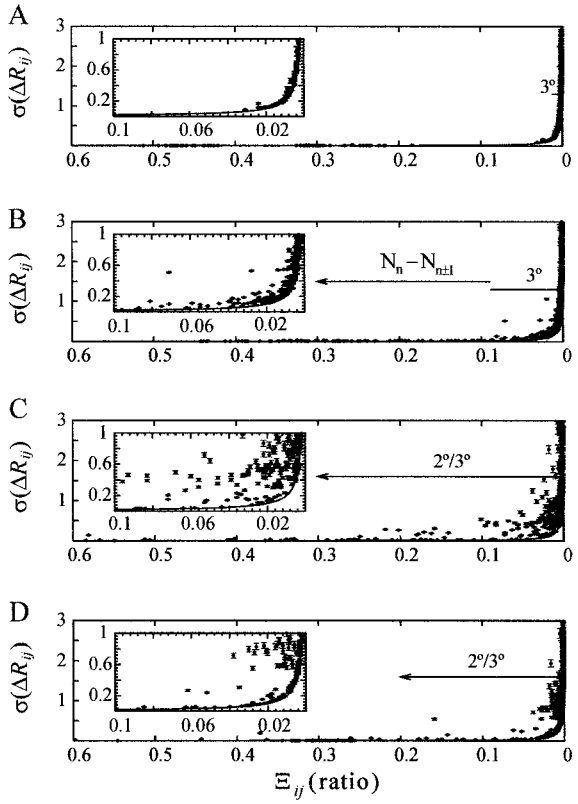

Figure 12. Correlations of $\sigma(\Delta R_{ij})$ to dipolar coupling ratio Ξ_{ij} at $S/N_\sigma = 500$ for (A) ^{13}C at $0.45\tau_{2250}$, (B) ^{15}N at $0.48\tau_{64.5}$, (C) exchangeable protons at $0.44\tau_{16k}$, and (D) exchangeable plus select methyl protons at $0.24\tau_{16k}$. Annotations indicate Ξ ranges dominated by structurally meaningful internuclear contacts. Error bars are $\pm\varsigma(\Delta R_{ij})$, and curves are $\sigma(\Delta R_{ij}) = (S/N_\sigma \cdot \Xi)^{-1}$.

for contacts involving a valine backbone proton). The implications of Figure 12 are explored further in the discussion.

The impact of nuclear distribution-dependent spin diffusion on rate measurement is perhaps made most clear by the 2D spectral simulations for full-size ubiquitin in Figures 13 and 14. In addition, these spectra help to illustrate the complications arising from limited resolution and spectral dispersion, which we have so far ignored in our analysis. Shifts are based upon ubiquitin shifts from solution, but were manually adjusted such that all crosspeaks in the featured slices represent exchange with a single autopeak. The two dimensional plots assume an 800 MHz ^{1}H frequency, line widths of 0.5 ppm for carbon and 0.25 ppm for nitrogen, and incorporate the 2D S/N_σ level of 500 as discussed. The indicated slices are shown below the 2D spectrum with this noise level and line width. For each simulated slice a companion plot is provided using line widths a factor of 4 more narrow and having no noise.

By cross-referencing the labeled peaks in the featured slices with Tables 1 and 2, it is clear that it will be difficult to measure meaningful couplings among generally densely distributed nuclei like carbon, whereas meaningful couplings among less dense nitrogen nuclei appear more accessible. Of the numbered carbon peaks (Figure 13), only peaks 6 and 7 (not shown) on the lys27 slice, and 4, 5 and 6 on the ile13 slice, correspond to distant tertiary contacts. Meanwhile, all of the labeled nitrogen peaks in Figure 14 are meaningful secondary and tertiary contacts, with all but peaks 1 and 2 in each slice being non-sequential. Figure 13 also shows a consequence of mixing for too long in addition to that shown in Figures 12–14. Not only do measured distances lose precision because of spin-diffusion, but many more peaks gain significant signal to noise in an already-crowded spectrum, making accurate measurements even more difficult.

Discussion

Previous studies have focused upon only very small model spin systems native to protein to analyze specific recoupling methods in careful detail, or used a collection of spins more representative of full protein for the purpose of exploring more general concerns such as the effect of finite resolution on solid state NMR resonance assignment (Tycko, 1996). Our approach has been intermediate to these two extremes, analyzing a generalized spin exchange experiment in the context of large spin systems native to protein to explore the effect of nuclear distribution and noise on solid state NMR structure data.

Ubiquitin contains 378 carbons, 105 nitrogens, 137 exchangeable protons, and 105 select methyl protons. Six angstroms is a distance regarded by some workers as the limit of measurement by magnetization-transfer in solution NMR experiments (Clore and Schwieters, 2002). Within this restriction there are 3781 homonuclear carbon pairs, 274 homonuclear nitrogen pairs, 391 exchangeable proton pairs and 1883 exchangeable plus methyl proton pairs separated spatially by less than 6 Å and covalently by more than two bonds.

It is not enough, however, to judge the potential contribution of distance constraints by a given spin system to structure calculations simply by enumeration of these meaningful contacts. The ratios Ξ_{ij} provide a reasonable means to gauge the expected pre-

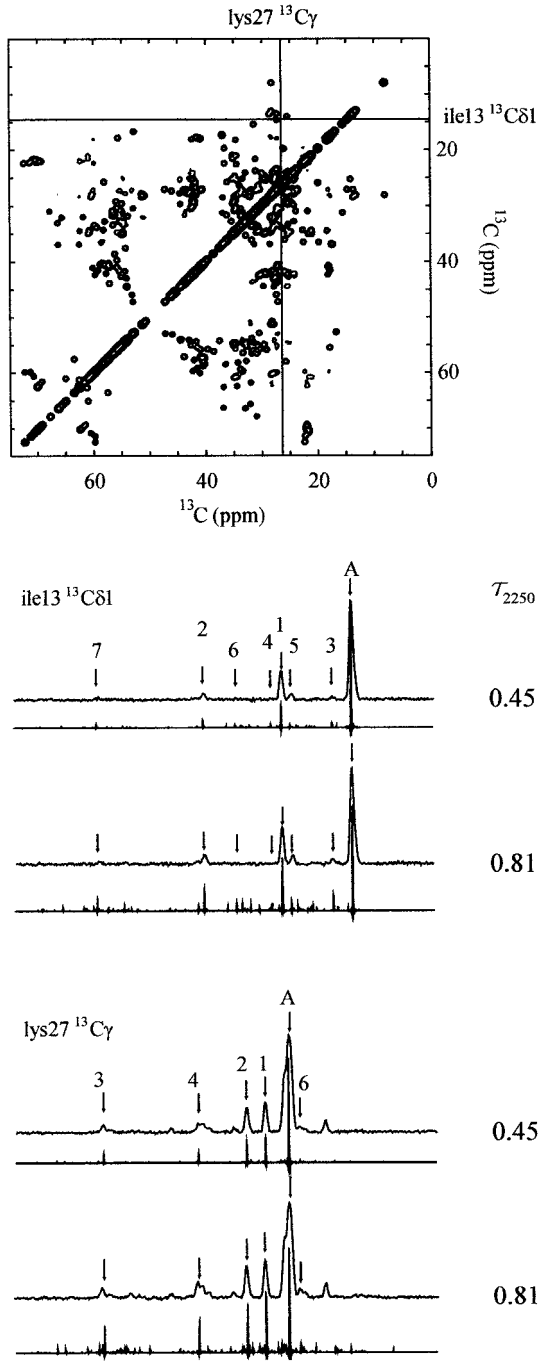

Figure 13. ^{13}C -^{13}C spectral simulations. Top: 2D spectrum at optimum mixing time for full protein (0.45 τ_{2250}). Middle and bottom: indicated slices at the optimum and a longer mixing time (0.81 τ_{2250}). 'A' indicates the autopeak, and numbered crosspeaks correspond to nuclei listed in Tables 1 and 2. Slices with realistic line widths are extracted from the 2D spectra, slices with artificially narrowed line widths are provided simply for illustration.

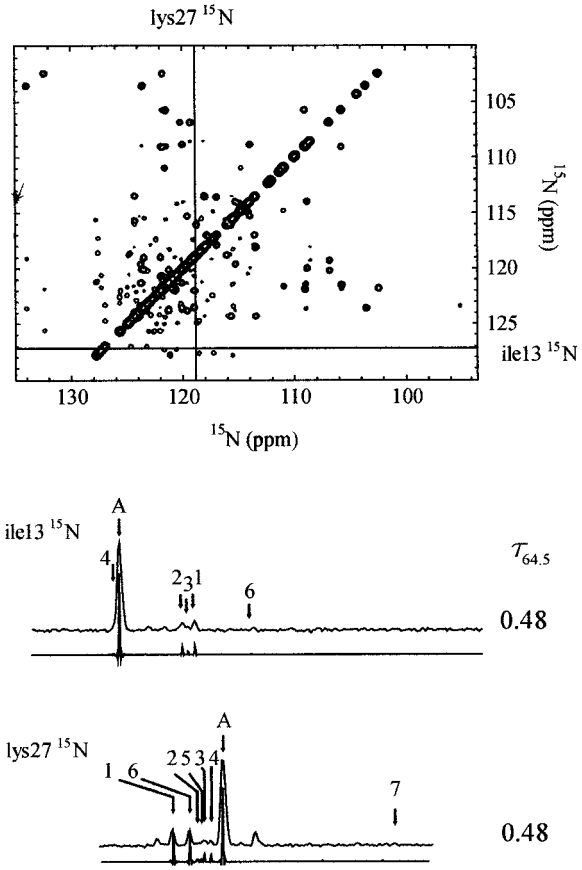

Figure 14. ^{15}N –^{15}N spectral simulation. Top: 2D spectrum at optimum mixing time for full protein ($0.48_{64.5}$). Bottom: Slices taken through the indicated amide nitrogen resonances. 'A' indicates the autopeak, and numbered crosspeaks correspond to nuclei listed in Tables 1 and 2. Slices with realistic line widths are extracted from the 2D spectra, slices with artificially narrowed line widths are provided simply for illustration.

cision of rate measurement for a given internuclear contact. The limits of relevance of these ratios is demonstrated in Figure 12 following the more quantitative assessment of the import of local nuclear distributions on solid state NMR structural data provided by the Monte Carlo analysis.

This relevance appears to hinge upon the nature of the spatial distribution of nuclei in the spin system. Homonuclear carbon and nitrogen in respectively uniformly labeled protein both have reasonably homogenous spatial distributions, albeit at very different densities. This is apparently important for the tight empirical relationship between the numerically determined $\sigma(\Delta R_{ij})$ value and calculated ratio Ξ_{ij} plotted in Figures 12A and 12B.

The homonuclear proton experiments considered present an interesting variation on spin system distribution. If exchangeable backbone and single exchangeable side-chain protons only were to be considered, results similar to nitrogen would be expected. However, in the case of arginine, lysine, asparagine, and glutamine, the side chain protons are present in very tight clusters relative to the rest of the distribution. The same is true of the select methyl protons when they are included. In this light, the inclusion of the select methyl group protons has interesting consequences. On one hand the additional proton clusters actually help to provide a more uniform spatial distribution of nuclei over exchangeable protons only, leading to a better $\sigma(\Delta R_{ij})$ vs. Ξ_{ij} correlation in plots 12D compared to 12C. On the other hand the additional methyl protons, while ostensibly providing more tertiary contacts, serve to exacerbate spin diffusion problems, at least when treated in the simplistic fashion employed here. This is reflected in the increase of analogous η scaling factors between Figures 11C and 11D. Hence, the recasting of dipolar coupling constants to Ξ ratios as in Figures 1C, 2C and 3C, to the extent that the spatial distribution renders the ratios relevant, provides a qualitative measure for a given internuclear pair's potential to contribute a constructive distance constraint.

Next we seek the means to judge different ratio values. How large must a coupling's ratio be, or how narrow must a coupling's distribution of experimentally derived distances be to be useful as a structure-refinement constraint? We pass no definite judgment ourselves, but do explore the considerations that govern the decision. One would certainly wish to set the bar of confidence high, but not so high that too many calculated distance constraints are so loose that they contribute negligibly in structure refinement. The $\sigma(\Delta R_{ij})$ values of 0.25 have already been discussed as being the upper limit whereby finite internuclear-distance constraint limits may be expected, and more stringent limits are imposed by accepting only contacts that demonstrate even lower values.

The correlations shown in Figure 12 indicate that at the $S/N_\sigma = 500$ level, a $\sigma(\Delta R_{ij})$ of 0.25 can only be realized by internuclear contacts with Ξ values of approximately 0.01 (or higher for the least homogenous of nuclear distributions, in this case exchangeable only protons), and a $\sigma(\Delta R_{ij})$ of 0.1 correlates to a ratio of 0.02 (or higher for exchangeable only protons). Thus ratios of 0.02 translate to distance precisions of approximately r_{ij} $+1.8\%/-1.6\%$ at 68% confidence level and r_{ij} $+3.8\%/-3.0\%$ at 95% confidence level. Ratios of 0.01 correspond to distance precisions of approximately r_{ij} $+4.9\%/-3.7\%$ at 68% confidence level and r_{ij} $+12.2\%/-6.5\%$ at the 95% confidence level. These $\sigma(\Delta R_{ij})$ values are of course specific to the noise level used. If higher signal-to-noise data is acquired lower Ξ values can be accessed while reliably remaining within these confidence intervals.

Concomitant with the significantly asymmetric r_{ij} confidence interval limits are expectation values which differ significantly from their true values, introducing inaccuracy as well as imprecision to structure calculations. Some perspective of these ranges is available by noting that precisions on isolated spin pair distance measurements have been made using solids NMR at $r_{ij} \pm 2\%$ for distances as long as 5 Å (Costa et al., 1997), and arguably even up to 8 Å (Holl et al., 1992). The current wisdom for CNS-type refinements (Brunger et al., 1998) in both liquids (Clore and Gronenborn, 1998) and solids (Castellani et al., 2002) NMR is to broadly bin cross-peaks into three or four classes. With flat potential across the range, in the above terms these classes span $r_{ij} \pm 20\%$ for 'strong' couplings to $r_{ij} \pm 54\%$ for 'very weak' couplings.

If the couplings with ratios below the 0.02 or 0.01 Ξ thresholds should not be depended upon, we must assess how much structural information is available from the remaining reliable data. For this we return to the discussion presented in the first half of this paper, the distribution of ratios featured in Figures 1–3, panels C, and the contact maps in Figure 4, all of which anticipated these threshold Ξ values. Additionally, Figure 15 highlights three residues in each of two β-strands with meaningful contacts above a ratio of 0.01 drawn explicitly. The implications of these Ξ thresholds for each specific isotopic labeling scheme and the spatial distribution that it represents follows.

In the homonuclear carbon case, all contacts that meet even the lower ratio threshold (0.01) are intra-residue (Figure 4B). While the explicit contact shown in Figure 15A and the few others like it may help to constrain side chain conformation, it is difficult to imagine that these couplings together with the much stronger one- and two-bond dipolar couplings would constrain the segment into even the correct secondary structure. It is even more difficult to imagine folding of the strands into the correct tertiary structure on the basis of these constraints alone. Should the reader wish to find another threshold value more appropriate, the individual contacts given for selected nuclei in Tables 1 and 2 may be useful. However, even though

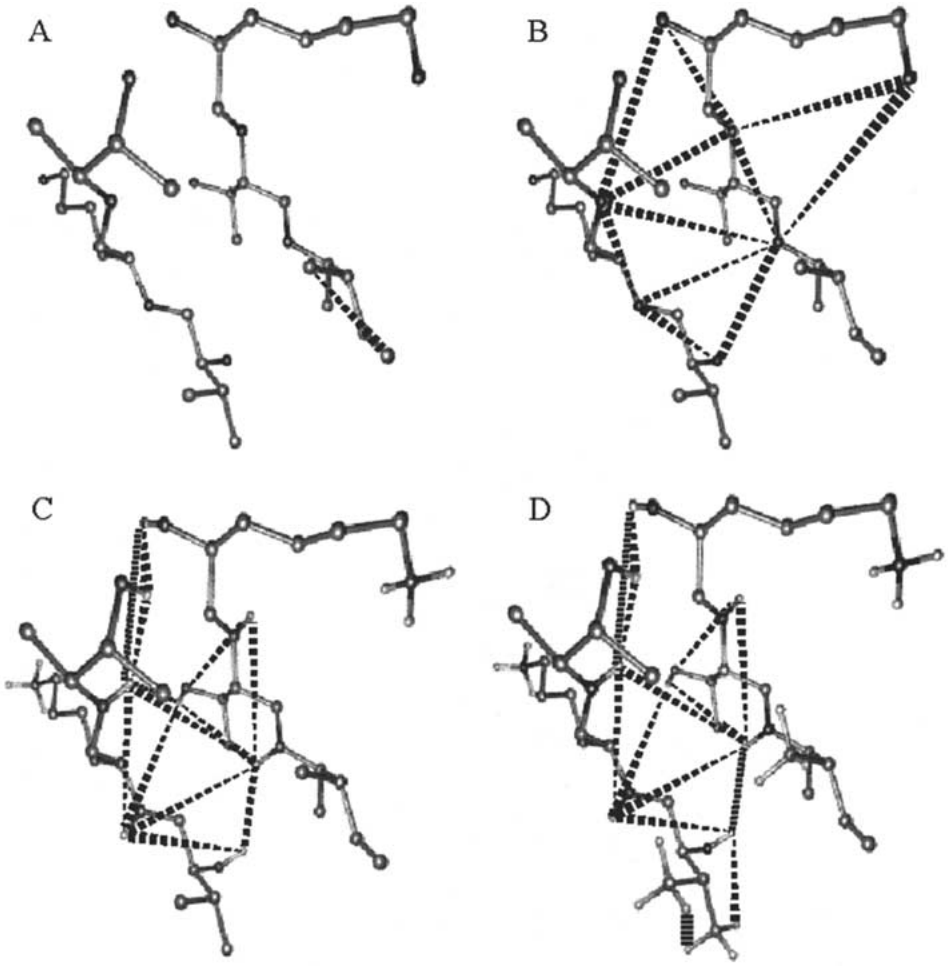

Figure 15. Molecular excerpts. Three residues from each of two anti-parallel β-strands (val5 lys6 thr7 – lys11 thr12 ile13) are depicted together with internuclear contacts of greater than two-bonds covalent separation and dipolar ratio greater than 0.01. (A) ^{13}C, (B) ^{15}N, (C) exchangeable protons, and (D) exchangeable plus methyl protons.

the residues selected were those with sidechains directed inwards from the protein surface, the α-helical sidechain carbon makes the most meaningful contacts only at very low couplings. The β-sheet sidechain carbon does not fare substantially better.

The homonuclear nitrogen case, in contrast, offers many more structurally meaningful couplings at ratio values even above 0.02, at which point the strongest couplings (between sequential backbone amides) are themselves useful. The impact of this on structure refinement is suggested by the contact plots in Figures 4C and 4D, as well as the explicitly indicated contacts for the molecular segment in Figure 15B. It is easier to believe that there are enough sufficiently precise ^{15}N contacts to constrain this segment into the correct secondary and tertiary structure. It should be noted that the absolute value of the homonuclear nitrogen couplings are very small, which would require a mixing time substantially longer than one's general experience with ^{13}C. Our experience with nitrogen however suggests that the ^{15}N relaxation times in protein are sufficiently long, being the order of a minute, that such experiments are feasible. Measurements of this type have in fact been used to provide a handful of ^{15}N-^{15}N distance constraints in the recent report of the structure of an SH3 domain (Castellani et al., 2002) by solid state NMR.

The accuracy of the homonuclear experiment conducted on select protons appears to depend on whether or not methyl protons are included. One might expect that the inclusion of methyl protons would benefit structure efforts, as it is residues with aliphatic side-

chains, frequently those with methyls, which pack the protein core. However, we see in Figures 11C and 11D that the inclusion of methyls causes an apparent loss in precision of all inter-proton distances, such that the data offer precisions closer to homonuclear carbon than nitrogen. When only exchangeable protons are used, the precision of the data is more similar to homonuclear nitrogen. Indeed, mixing among exchangeable protons alone has yielded useful distance constraints as demonstrated recently by Reif et al. (2003). It is possible then that the more precise, albeit fewer internuclear contacts afforded by a mixing experiment using only exchangeable protons is a more attractive prospect than the same experiment which adds the select methyl protons to the spin system. To some extent this inference is a consequence of our simplistic handling of methyl groups. It remains to be examined whether a more realistic model for the protons in a rotating methyl group, which recognizes them as on average being isochronous and magnetically equivalent, would come to a different conclusion. Given the potential tertiary contacts these spins could provide, the development of better treatments for methyl groups in proteins will be an important next step in solid state NMR of proteins.

Conclusion

The approach presented in this paper is useful for evaluating the suitability of an isotopic enrichment scheme for obtaining distance constraints by solids NMR, and for estimating the signal-to-noise required to achieve a desired level of precision. It is clear from our results that dipolar truncation effects aside (Hohwy et al., 2002), the isotopic distribution needs to be dilute enough so that the length scale one wishes to probe is comparable to the shorter inter-spin distances in the distribution. If this constraint is not satisfied by the nuclear distribution, the relay of magnetization by spin diffusion coupled with instrumental noise rapidly degrades the precision with which the important distances can be measured. This makes ^{15}N or exchangeable ^{1}H the most attractive isotopic enrichment patterns to use in proteins for homonuclear distance measurements. Only by extensive dilution can ^{13}C be made to compete in this regard, as demonstrated recently (Castellani et al., 2002). In this type of an approach an additional uncertainty may be introduced in that the isotopic incorporation can have significant residue dependent variability. Application of the computational method developed in this paper would be a good entry to estimating the additional uncertainty introduced into distances determined with this type of sample.

The recasting of dipolar coupling constants as Ξ ratios offers an optimistic estimation of anticipated precision of coupling rate measurement at a given signal to noise. The number and precision of structurally useful dipolar contacts above a selected ratio threshold helps to identify those nuclear distributions which hold the greatest potential utility to solid state NMR structure determination. While our approach is generally applicable, our conclusions regarding the specific isotopic variations considered are qualitative only; we explicitly leave to future work a quantitative assessment of the sensitivity of structure refinement upon the number, precision, and accuracy of measured distance constraints. The approach should find use in providing a realistic estimate of the signal to noise required to obtain a desired precision, and thereby aid the experimenter in differentiating feasible experiments from those which are not. With the proper calibration this computational approach can also be used to identify the range of optimum mixing times for spin exchange.

While our results would indicate that homonuclear solid state dipolar exchange experiments using ^{15}N-^{15}N couplings or ^{1}H-^{1}H couplings in deuterated back-exchanged protein will provide the best precision in distance measurement, this conclusion is based on equivalent signal to noise. Comparison between ^{13}C detect, ^{1}H detect and ^{15}N detect experiments is not as straightforward in solid state NMR as in solution NMR. For the moment ^{15}N is still largely directly detected when resolution is at a premium. Although the apparent precision afforded by the ^{15}N dipolar exchange experiments is high, achieving the signal to noise required is not straightforward for small samples. However, if lower precision is satisfactory, the requirements on signal to noise become significantly less stringent. As indicated in figures 10 and 11, the precision of the ^{15}N exchange experiment is about 10 times that of the ^{1}H-^{1}H exchange experiments. The two experiments can however be made equivalent if the signal to noise of the ^{15}N data is likewise reduced by a factor of 10. This is a regime in which such experiments are quite feasible even with direct ^{15}N detection. The determination of whether ^{15}N or ^{1}H exchange will ultimately be more useful then is an open question, one whose answer will depend on a number of very different experimental factors. Regardless of the outcome, the work presented here

demonstrates how to optimize these sorts of experiments to achieve the highest precision that can be experimentally achieved.

Although not specifically addressed in the present work, heteronuclear measurements using ^{13}C-^{15}N dipolar couplings have some advantages in the realm of providing distance constraints for proteins and peptides. While the spin dynamics are quite different, and not directly amenable to the present analysis, some qualitative lessons may be transferred. In uniformly labeled samples we observe that the problem of spin diffusion is as much one of dilution of the available magnetization, as it is of magnetization relay short circuiting transfer via desired long-range pathways. In any experiment where magnetization is transferred between ^{15}N and ^{13}C nuclei, for instance by TEDOR (Michal and Jelinski, 1997; Jaroniec et al., 2002), always having a directly bonded nearest neighbor will have a negative impact on the precision of the experiment. While relay transfer is not a concern, reduction of the available magnetic moment for transfer to more distant coupling partners by leakage into less interesting nearest neighbor sites can be a problem for such techniques. The advantages of modifications of the isotopic labeling pattern, or use of selective transfers to circumvent these problems, can then be evaluated on the basis of analogous coupling ratios and contact maps as described in the first half of this article.

New isotopic enrichment schemes and solid state NMR methods are being applied to small proteins at a rapid pace, each with its own merits and detractions. In comparing different approaches one has to address not only issues of sensitivity and resolution and whether through-space contacts can be observed, but also how accurate the distance constraints so obtained are. The Monte Carlo method used herein provides a means of critically comparing competing techniques insofar as simple spin exchange is an accurate model of the spin dynamics. Continued advances in computing power and techniques for simulating large complex spin systems will eventually make it possible to adapt this general approach to address the precision of distance measurements using more sophisticated mixing schemes which are both more complex and more realistic.

References

Alexeev, D., Bury, S.M., Turner, M. A., Ogunjobi, O. M., Muir, T. W., Ramage, R. and Sawyer, L. (1994) *Biochem. J.*, **299**, 159–163.

Bak, M., Rasmussen, J.T. and Nielsen, N.C. (2000) *J. Magn. Reson.*, **147**, 296–330.

Benzinger, T.L., Gregory, D.M., Burkoth, T.S., Miller-Auer, H., Lynn, D.G., Botto, R.E. and Meredith, S.C. (1998) *Proc. Natl. Acad. Sci. USA*, **95**, 13407–12.

Bevington, P.R. and Robinson, D.K. (1991) *Data Reduction and Error Analysis for the Physical Sciences*, WCB/McGraw-Hill, New York, NY.

Borgias, B.A. and James, T.L. (1988) *J. Magn. Reson.*, **79**, 493–512.

Borgias, B.A. and James, T.L. (1990) *J. Magn. Reson.*, **87**, 475–487.

Brunger, A., Adams, P., Clore, G., DeLano, W., Gros, P., Grosse-Kunstleve, R., Jiang, J., Kuszewski, J., Nilges, M., Pannu, N., Read, R., Rice, L., Simonson, T. and Warren, G. (1998) *Acta Crystallogr.*, **D54**, 905–921.

Castellani, F., Rossum, B.V., Diehl, A., Schubert, M., Rehbein, K. and Oschkinat, H. (2002) *Nature*, **420**, 98–102.

Christensen, A.M. and Schaefer, J. (1993) *Biochemistry*, **32**, 2868–2873.

Clore, G.M. and Gronenborn, A.M. (1998) *Proc. Natl. Acad. Sci. USA*, **95**, 5891–5898.

Clore, G.M. and Schwieters, C.D. (2002) *Curr. Opin. Struct. Biol.*, **12**, 146–153.

Costa, P.R., Sun, B. and Griffin, R.G. (1997) *J. Am. Chem. Soc.*, **119**, 10821–10830.

Dellwo, M.J. and Wand, J. (1993) *J. Biomol. NMR*, **3,** 205–214.

Dellwo, M.J., Schneider, D.M. and Wand, A.J. (1994) *J. Magn. Reson.*, **B103**, 1–9.

Dusold, S., Kummerlen, J., Schaller, T., Sebald, A. and Dollase, W.A. (1997) *J. Phys. Chem.*, **B101**, 6359–6366.

Ernst, R.R., Bodenhausen, G. and Wokaun, A. (1992) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions,* Oxford University Press, Oxford.

Goobes, G. and Vega, S. (2002) *J. Magn. Reson.*, **154**, 236–251.

Goto, N.K., Gardner, K.H., Mueller, G.A., Willis, R.C. and Kay, L.E. (1999) *J. Biomol. NMR*, **13**, 369–374.

Gullion, T. and Schaefer, J. (1989) *J. Magn. Reson.*, **81**, 196–200.

Hajduk, P.J., Augeri, D.J., Mack, J., Mendoza, R., Yang, J.G., Betz, S.F. and Fesik, S.W. (2000) *J. Am. Chem. Soc.*, **122**, 7898–7904.

Helluy, X., Marichal, C. and Sebald, A. (2000) *J. Phys. Chem.*, **B104**, 2836–2845.

Henrichs, P.M., Linder, M. and Hewitt, J.M. (1986) *J. Chem. Phys.*, **85**, 7077–1086.

Hodgkinson, P. and Emsley, L. (1999) *J. Magn. Reson.*, **139**, 46–59.

Hohwy, M., Rienstra, C.M. and Griffin, R.G. (2002) *J. Chem. Phys.*, **117**, 4973–4987.

Holl, S.M., Marshall, G. R., Beusen, D.D., Kociolek, K., Redlinski, A.S., Leplawy, M.T., McKay, R.A., Vega, S. and Schaefer, J. (1992) *J. Am. Chem. Soc.*, **114**, 4830–4833.

Igumenova, T.I., McDermott, A.E., Zilm, K.W., Martin, R.W. and Paulson, E. K. (2003) *J. Biomol. NMR*, submitted.

Inooka, H., Ohtaki, T., Kitahara, O., Ikegami, T., Endo, S., Kitada, C., Ogi, K., Onda, H., Fujino, M. and Shirakawa, M. (2001) *Nat. Struct. Biol.*, **8**, 161–165.

Ishii, Y., Yesinowski, J.P. and Tycko, R. (2001) *J. Am. Chem. Soc.*, **123**, 2921–2922.

Jaroniec, C.P., Filip, C. and Griffin, R. G. (2002) *J. Am. Chem. Soc.*, **124**, 10728–10742.

Jaroniec, C.P., Tounge, B.A., Herzfeld, J. and Griffin, R.G. (2001) *J. Am. Chem. Soc.*, **123**, 3507–3519.

Ketchem, R.R., Lee, K.C., Huo, S. and Cross, T.A. (1996) *J. Biomol. NMR*, **8,** 1–14.

Liu, H., Spielmann, H.P., Ulyanov, N.B., Wemmer, D.E. and James, T.L. (1995) *J. Biomol. NMR*, **6**, 390–402.

Macura, S. (1994) *J. Magn. Reson.*, **B104**, 168–171.

Macura, S. (1995) *J. Magn. Reson.*, **A112**, 152–159.
Macura, S. and Ernst, R. R. (1980) *Mol. Phys.*, **41**, 95–117.
McDermott, A., Polenova, T., Bockmann, A., Zilm, K.W., Paulsen, E.K., Martin, R.W. and Montelione, G.T. (2000) *J. Biomol. NMR*, **16**, 209–219.
Michal, C.A. and Jelinski, L.W. (1997) *J. Am. Chem. Soc.*, **119**, 9059–9060.
Mueller, K.T. (1995) *J. Magn. Reson.*, **A113**, 81–93.
Murphy, O.J., Kovacs, F.A., Sicard, E.L. and Thompson, L.K. (2001) *Biochemistry*, **40**, 1358–1366.
Nomura, K., Takegoshi, K., Terao, T., Uchida, K. and Kainosho, M. (2000) *J. Biomol. NMR*, **17**, 111–123.
Olejniczak, E.T., Grampe, R.T. and Fesik, S.W. (1986) *J. Magn. Reson.*, **67**, 28–41.
Pauli, J., Baldus, M., Rossum, B. van, Groot, H. de and Oschkinat, H. (2001) *CHEMBIOCHEM*, **2001**, 272–281.
Reif, B., Jaroniec, C.P., Rienstra, C.M., Hohwy, M. and Griffin, R.G. (2001) *J. Magn. Reson.*, **151**, 320–327.
Reif, B., Rossum, B. van, Castellani, F., Rehbein, K., Diehl, A. and Oschkinat, H. (2003) *J. Am. Chem. Soc.*, **125**, 1488–1489.
Rosen, M.K., Gardner, K.H., Willis, R.C., Parris, W.E., Pawson, T. and Kay, L.E. (1996) *J. Mol. Biol.*, **263**, 627–636.
Rossum, B.-J. van, Castellani, F., Rehbein, K., Pauli, J. and Oschkinat, H. (2001) *CHEMBIOCHEM*, **2**, 906–914.
Schalk-Hihi, C., Zhang, Y.Z. and Markham, G.D. (1998) *Biochemistry*, **37**, 7608–7616.
Shriver, J.W. and Edmondson, S. (1994) *Meth. Enzymol.*, **240**, 415–438.
Straus, S.K., Bremi, T. and Ernst, R.R. (1998) *J. Biomol. NMR*, **12**, 39–50.
Suri, A.K. and Levy, R.M. (1995) *J. Magn. Reson.*, **B106**, 24–31.
Tycko, R. (1996) *J. Biomol. NMR*, **8**, 239–251.
Tycko, R. (2001) *Annu. Rev. Phys. Chem.*, **52**, 575–606.
Watts, A. (1999) *Curr. Opin. Biotechnol.*, **10**, 48–53.
Weber, P.L., Brown, S.C. and Mueller, L. (1987) *Biochemistry*, **26**, 7282–7290.
Word, J.M., Lovell, S.C., Richardson, J.S. and Richardson, D.C. (1999) *J. Mol. Biol.*, **285**, 1735–1747.