Journal of Biomolecular NMR, 8 (1996) 239–251
ESCOM

J-Bio NMR 383

Prospects for resonance assignments in multidimensional solid-state NMR spectra of uniformly labeled proteins

Robert Tycko

Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, 5 Center Drive, Bethesda, MD 20892-0520, U.S.A.

Received 28 May 1996
Accepted 22 July 1996

Keywords: Solid-state NMR; Magic angle spinning; Chemical shift correlation; Resonance assignment

Summary

The feasibility of assigning the backbone ^{15}N and ^{13}C NMR chemical shifts in multidimensional magic angle spinning NMR spectra of uniformly isotopically labeled proteins and peptides in unoriented solid samples is assessed by means of numerical simulations. The goal of these simulations is to examine how the upper limit on the size of a peptide for which unique assignments can be made depends on the spectral resolution, i.e., the NMR line widths. Sets of simulated three-dimensional chemical shift correlation spectra for artificial peptides of varying length are constructed from published liquid-state NMR chemical shift data for ubiquitin, a well-characterized soluble protein. Resonance assignments consistent with these spectra to within the assumed spectral resolution are found by a numerical search algorithm. The dependence of the number of consistent assignments on the assumed spectral resolution and on the length of the peptide is reported. If only three-dimensional chemical shift correlation data for backbone ^{15}N and ^{13}C nuclei are used, and no residue-specific chemical shift information, information from amino acid side-chain signals, and proton chemical shift information are available, a spectral resolution of 1 ppm or less is generally required for a unique assignment of backbone chemical shifts for a peptide of 30 amino acid residues.

Introduction

Multidimensional chemical shift correlation spectra are the basis for resonance assignments in liquid-state NMR spectra of uniformly isotopically labeled proteins (Wüthrich, 1986; Ernst et al., 1987; Bax and Grzesiek, 1993). The assignment of chemical shifts to the majority of the nuclei in a protein backbone using multinuclear, multidimensional spectra is a prerequisite for structure determination. For a protein of approximate molecular weight 20 kDa, typical resonance line widths for ^{1}H, ^{13}C, and ^{15}N signals in liquid-state spectra range approximately from 10 to 30 Hz. The relatively narrow lines allow nearly all of the signals in appropriate combinations of three-dimensional (3D) chemical shift correlation spectra of uniformly labeled proteins of this size to be resolved and sequentially assigned.

Recently, analogs of liquid-state chemical shift correlation spectroscopy techniques have been developed for solid-state NMR spectroscopy of multiply labeled organic and biochemical compounds (Menger et al., 1986; Raleigh et al., 1987; Bennett et al., 1992; Ok et al., 1992; Baldus et al., 1994; Griffiths et al., 1994; Boender et al., 1995; Fujiwara et al., 1995; Sun et al., 1995). These analogs apply to noncrystalline and polycrystalline solids with a random distribution of molecular orientations. The solid-state NMR techniques depend on magic angle spinning (MAS) and high-power proton decoupling to narrow the lines in ^{13}C and ^{15}N spectra of isotopically labeled samples. An important motivation for the development of these chemical shift correlation techniques has been the goal of assigning the ^{13}C and ^{15}N signals in multidimensional MAS spectra of uniformly isotopically labeled peptides and proteins (or other biopolymers) as a prelude to structural investigations of peptides and proteins in the solid state. Such an approach may be useful in studies of peptides in membrane-bound form, in insoluble aggregates, in lyophilized or precipitated solids, and in frozen solution.

Based on published results from a number of labora-

tories, it is clear that multidimensional solid-state chemical shift correlation spectra can be obtained and interpreted for relatively small, multiply labeled or uniformly labeled compounds (Bennett et al., 1992; Baldus et al., 1994; Boender et al., 1995; Fujiwara et al., 1995). What is less clear is whether the corresponding spectra of much larger, uniformly labeled peptides and proteins would be interpretable. The feasibility of solid-state chemical shift correlation spectroscopy in larger, uniformly labeled biopolymers depends on factors related to sensitivity, such as spin relaxation rates, polarization transfer efficiencies, sample quantities and concentrations, and the efficacy of signal enhancement techniques (Barrett et al., 1994; Zysmilich and McDermott, 1994; Gerfen et al., 1995; Tycko et al., 1995). But in addition to the issue of sensitivity, an upper limit on the tractable size of a uniformly labeled peptide or protein is inevitably imposed by the spectral resolution, i.e., the NMR line widths. The reported line widths in solid-state NMR spectra of specifically labeled proteins depend on the details of the system and the sample preparation, but they are invariably larger than in liquid-state NMR spectra of proteins. The impact of line widths on multidimensional solid-state NMR spectroscopy of uniformly labeled peptides and proteins has not yet been investigated systematically.

This paper reports the results of a series of simulations carried out as an initial exploration of the limits on molecular size in multidimensional solid-state chemical shift correlation spectra of uniformly labeled peptides and proteins imposed by the ^{13}C and ^{15}N line widths. In these simulations, 3D spectra for artificial peptides of variable length are constructed from liquid-state NMR chemical shift data for uniformly ^{13}C- and ^{15}N-labeled ubiquitin (Wang et al., 1995), a well-characterized soluble protein. Only NMR signals from ^{13}C and ^{15}N nuclei in the peptide backbone are considered. For various values of the spectral resolution, chemical shift assignments that are consistent with the 3D spectra to within the assumed resolution are found by a random search algorithm. In this way, the peptide length beyond which a unique assignment of the spectra no longer exists, and its dependence on the spectral resolution are estimated. The results of these simulations are expected to be important as a guide to the development of multidimensional solid-state NMR techniques and to future solid-state NMR investigations of uniformly labeled peptides and proteins.

It should be emphasized at the outset that the simulations and discussion below apply to unoriented solid samples, i.e., samples with an isotropic distribution of molecular orientations. Multidimensional solid-state NMR techniques for applications to uniaxially oriented systems have also been developed (Jelinek et al., 1995; Ramamoorthy et al., 1995). The strategies for resonance assignment and structure determination that apply to oriented systems may be qualitatively different from the strategies that apply to unoriented systems. The problem of resonance assignment in oriented systems is not treated in this paper.

Idealized solid-state chemical shift correlation spectroscopy

The simulations described below are based on seven artificial 3D heteronuclear chemical shift correlation data sets that would result from measurements with radio-frequency (rf) pulse sequences of the general forms shown in Fig. 1. The contents of these data sets are summarized in Table 1. The pulse sequence in Fig. 1a represents a

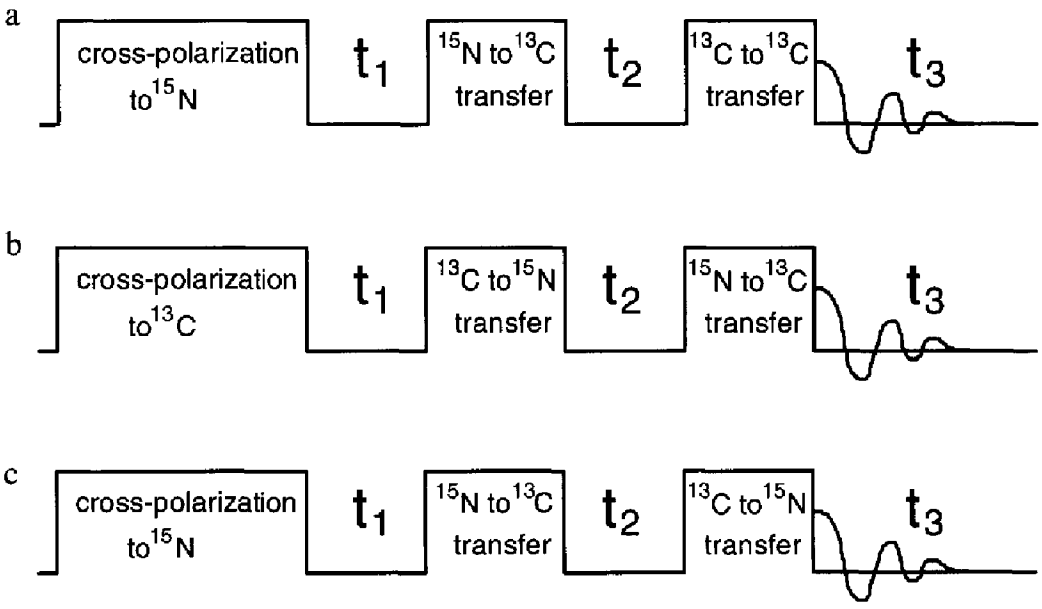

Fig. 1. Idealized rf pulse sequences for 3D chemical shift correlation measurements on ^{15}N- and ^{13}C-labeled peptides and proteins.

TABLE 1
CONTENTS OF 3D SPECTRAL DATA SETS USED IN ASSIGNMENT SIMULATIONS

Data set	j=1	j=2	j=3
SA(i,j), $1 \le i \le L$	N(i)	C^{α}(i)	CO(i)
SB(i,j), $1 \le i \le L-1$	C^{α}(i)	CO(i)	N(i+1)
SC(i,j) $1 \le i \le L-1$	CO(i)	N(i+1)	C^{α}(i+1)
SD(i,j), $1 \le i \le L-1$	CO(i+1)	N(i+1)	CO(i)
SD(i,j), $L \le i \le 2L-2$	CO(i–L+1)	N(i–L+2)	CO(i–L+1)
SE(i,j), $1 \le i \le L-1$	C^{α}(i)	N(i+1)	C^{α}(i+1)
SE(i,j), $L \le i \le 2L-1$	C^{α}(i–L+1)	N(i–L+1)	C^{α}(i–L+1)
SF(i,j), $1 \le i \le L-1$	N(i+1)	C^{α}(i)	N(i)
SF(i,j), $L \le i \le 2L-1$	N(i–L+1)	C^{α}(i–L+1)	N(i–L+1)
SG(i,j), $1 \le i \le L-1$	N(i)	CO(i)	N(i+1)
SG(i,j), $L \le i \le 2L-2$	N(i–L+2)	CO(i–L+1)	N(i–L+2)

N(i), C^{α}(i), and CO(i) are the chemical shifts of the amide nitrogen, α-carbon, and carbonyl carbon of residue i in the backbone of a peptide comprising L amino acid residues.

generic 3D ^{15}N-^{13}C-^{13}C chemical shift correlation technique in solid-state NMR. Nuclear spin polarization is prepared initially on ^{15}N sites by cross-polarization from protons. After the t_1 evolution period, polarization is transferred to ^{13}C nuclei, and after the t_2 evolution period, transferred further to other ^{13}C nuclei. ^{13}C NMR signals are detected during the t_3 period. The pulse sequence in Fig. 1b represents a generic 3D ^{13}C-^{15}N-^{13}C chemical shift correlation technique. Nuclear spin polarization is prepared initially on ^{13}C sites. After the t_1 evolution period, polarization is transferred to ^{15}N nuclei, and after the t_2 evolution period, transferred back to ^{13}C nuclei. ^{13}C NMR signals are detected during the t_3 period. The pulse sequence in Fig. 1c represents a generic 3D ^{15}N-^{13}C-^{15}N chemical shift correlation technique. Nuclear polarization is prepared initially on ^{15}N sites. After the t_1 evolution period, polarization is transferred to ^{13}C nuclei, and after the t_2 evolution period, transferred back to ^{15}N nuclei. ^{15}N NMR signals are detected during the t_3 period. High-speed MAS and high-power proton decoupling are assumed to be employed during t_1, t_2, and t_3 in Fig. 1, so that the nuclear spin evolution during these periods is determined primarily by the isotropic chemical shifts of the ^{15}N and ^{13}C nuclei.

The polarization transfers in Fig. 1 may be effected, at least in principle, by a variety of techniques that are based on homonuclear or heteronuclear magnetic dipole–dipole couplings or scalar couplings (Gullion and Schaefer, 1989; Tycko and Dabbagh, 1990,1991; Bennett et al., 1992,1994; Hing et al., 1993; Tycko and Smith, 1993; Baldus et al., 1994; Kolbert et al., 1994; Tomaselli et al., 1994; Tycko, 1994; Gregory et al., 1995; Sun et al., 1995). The details of these techniques are unimportant in the present simulations, but the following simplifying assumptions are made regarding the polarization transfers:

(1) ^{13}C-^{13}C, ^{13}C-^{15}N and ^{15}N-^{13}C polarization transfers can be limited to transfers between directly bonded nuclei, i.e., one-bond polarization transfers. This assumption permits signals from the peptide backbone to be considered exclusively, allowing potential problems associated with the overlap of backbone and side-chain resonances to be dismissed. In the case of polarization transfers mediated by dipole–dipole couplings, the amount of transferred polarization is proportional to r^{-6} in the short-time limit, where r is the internuclear distance. One-bond polarization transfers are then favored over more distant transfers by a factor > 10. Sequential one-bond transfers can be suppressed by employing short polarization transfer periods. Alternatively, polarization transfer techniques that are intrinsically selective for the chemical shifts of the nuclei that act as the sources and destinations of polarization could be designed. Thus, the assumption of one-bond polarization transfers is not incompatible with experiments.

(2) Only polarization transfers among nuclei in the peptide backbone are considered. Amino acid side chains and conformations are assumed not to affect the backbone one-bond polarization transfer efficiencies significantly. The intensities of signals in a 3D chemical shift correlation spectrum that arise from one-bond transfers are then roughly equal to one another.

(3) If desired, ^{13}C-^{15}N and ^{15}N-^{13}C two-bond polarization transfers can be performed, in particular between t_1 and t_2 in Figs. 1b and c, as explained below. Two-bond polarization transfers may occur in a single step or as sequential ^{13}C-^{13}C and ^{13}C-^{15}N or ^{15}N-^{13}C one-bond transfers. No assumption is made regarding the uniformity of two-bond polarization transfers.

If only one-bond, non-frequency-selective polarization transfers occur, the application of the pulse sequence in Fig. 1a to a uniformly ^{13}C- and ^{15}N-labeled polypeptide leads to a 3D spectrum which contains signals that correlate the chemical shifts of the amide nitrogen N(i) (in t_1), the α-carbon C^{α}(i) (in t_2), and the carbonyl carbon CO(i) (in t_3) of each amino acid in the peptide sequence, where i is the number of the amino acid. The N(i)-C^{α}(i)-CO(i) 3D chemical shift correlation data set is called SA. The same 3D spectrum also contains signals that correlate the

chemical shifts of $C^{\alpha}(i)$ (in t_3), CO(i) (in t_2), and N(i + 1) (in t_1). The $C^{\alpha}(i)$-CO(i)-N(i + 1) chemical shift correlation data set is called SB. The pulse sequence in Fig. 1b leads to a 3D spectrum which contains signals that correlate the chemical shifts of CO(i) (in t_1), N(i + 1) (in t_2), and $C^{\alpha}(i + 1)$ (in t_3). The CO(i)-N(i + 1)-$C^{\alpha}(i + 1)$ chemical shift correlation data set is called SC.

If both one-bond and two-bond ^{13}C-^{15}N polarization transfers occur between t_1 and t_2 in Fig. 1b, two additional data sets that are useful in the assignment algorithm may be derived from the resulting 3D spectrum. The first of these data sets, called SD, contains signals that correlate the chemical shifts of CO(i + 1) (in t_1), N(i + 1) (in t_2), and CO(i) (in t_3) as well as signals that correlate the chemical shifts of CO(i) (in t_1), N(i + 1) (in t_2), and CO(i) (in t_3). The second of these data sets, called SE, contains signals that correlate the chemical shifts of $C^{\alpha}(i)$ (in t_1), N(i + 1) (in t_2), and $C^{\alpha}(i + 1)$ (in t_3) as well as signals that correlate the chemical shifts of $C^{\alpha}(i)$ (in t_1), N(i) (in t_2), and $C^{\alpha}(i)$ (in t_3).

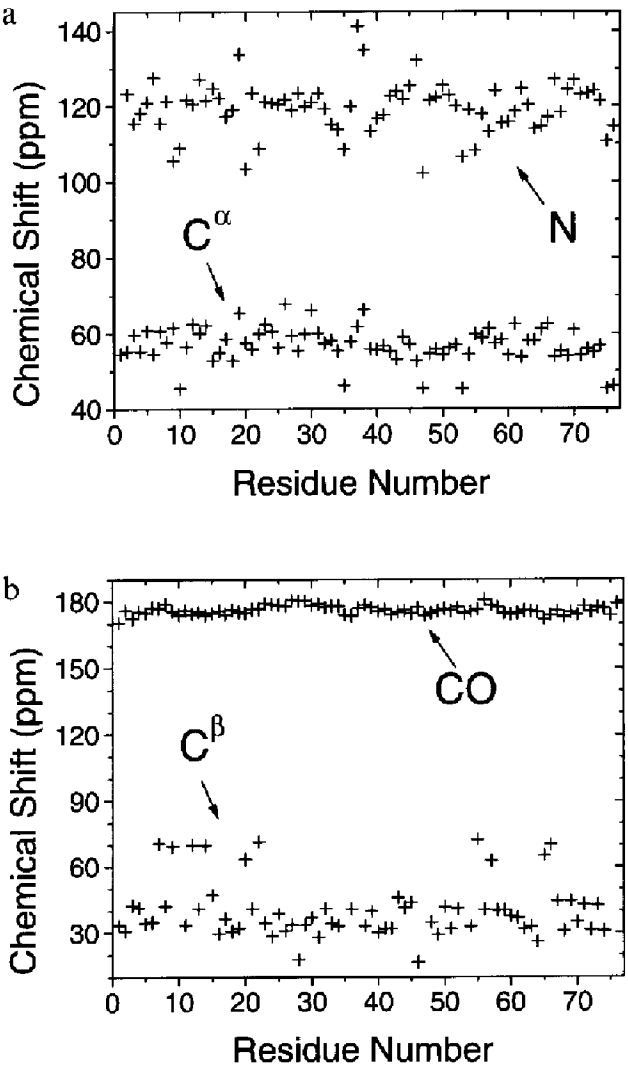

Fig. 2. Chemical shifts for (a) backbone amide nitrogens and α-carbons; and (b) backbone carbonyl carbons and β-carbons in ubiquitin, taken from liquid-state NMR measurements (Wang et al., 1995).

If one-bond and two-bond ^{15}N-^{13}C polarization transfers occur between t_1 and t_2 in Fig. 1c, two additional data sets may be derived from the resulting 3D spectrum. The first of these data sets, called SF, contains signals that correlate the chemical shifts of N(i + 1) (in t_1), $C^{\alpha}(i)$ (in t_2), and N(i) (in t_3) as well as signals that correlate the chemical shifts of N(i) (in t_1), $C^{\alpha}(i)$ (in t_2), and N(i) (in t_3). The second of these data sets, called SG, contains signals that correlate the chemical shifts of N(i) (in t_1), CO(i) (in t_2), and N(i + 1) (in t_3) as well as signals that correlate the chemical shifts of N(i + 1) (in t_1), CO(i) (in t_2), and N(i + 1) (in t_3).

As summarized in Table 1, data sets SA through SG are represented in the simulations by two-dimensional arrays of chemical shift values, with three columns and M rows. A signal in a 3D spectrum is represented by one row in the corresponding array. For a peptide of length L, M is equal to L for SA, L–1 for SB and SC, 2L–2 for SD and SG, and 2L–1 for SE and SF. For example, SB(i,1) is the chemical shift of $C^{\alpha}(i)$, SB(i,2) is the chemical shift of CO(i), and SB(i,3) is the chemical shift of N(i + 1). Signals in SA, SB, and SC are assumed to be quantitative, i.e., of approximately equal intensities. This assumption means that rows in SA, SB, and SC can be eliminated from further consideration after they are used once in the assignment algorithm.

In order to approximate the chemical shift distributions of real proteins, the chemical shift values in data sets SA through SG are taken from liquid-state chemical shift data for the 76-residue protein ubiquitin (Wang et al., 1995). The relevant chemical shift values are plotted in Fig. 2. The reported assignments for L consecutive amino acids in the ubiquitin sequence, starting with an initial residue number I, are used to generate artificial 3D chemical shift correlation data sets for peptides of length L. The chemical shift of the terminal amino nitrogen N(1) is set to zero in every peptide, reflecting the fact that amino ^{15}N signals are well resolved from amide ^{15}N signals in experimental spectra.

The association of a chemical shift value with each nucleus in the peptide backbone constitutes a complete 'assignment'. The assignment algorithm is the process of deriving an assignment from given 3D chemical shift correlation data sets. In the simulations described below, all 3D spectra are assumed to have the same spectral resolution, specified by the variable R. In a real experimental spectrum, the NMR line widths are likely to vary somewhat from nucleus to nucleus and from site to site in a manner that is currently not well characterized. R therefore represents an average spectral resolution. Possible assignments then fall into four categories. An assignment is considered to be *consistent* with the given 3D data sets if the chemical shift values of all signals in the 3D

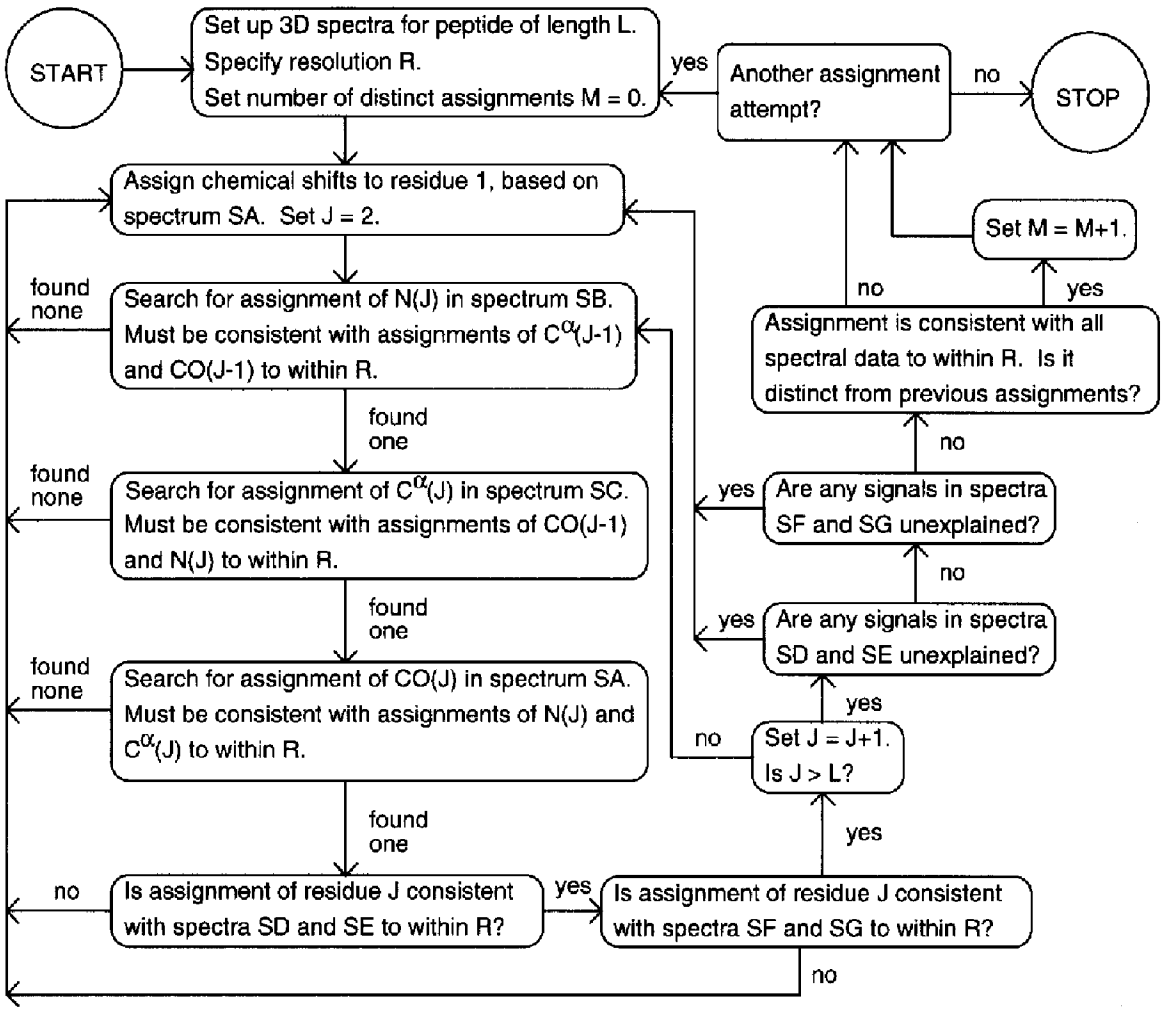

Fig. 3. Summary of the sequential assignment algorithm.

data sets differ from the values that would be predicted from the assignment by amounts less than or equal to the specified R in all three dimensions. An assignment is considered to be *inconsistent* with the given 3D data sets if at least one of the chemical shift values in the 3D data sets differs from the values that would be predicted from the assignment by an amount greater than R. An assignment is considered to be *correct* (to within the specified spectral resolution) if the chemical shift values assigned to each nucleus in the peptide backbone differ from the corresponding values in residues I to I+L−1 of ubiquitin by amounts less than or equal to R (with the exception of the terminal amino ^{15}N chemical shift). An assignment is considered to be *incorrect* if at least one of the chemical shift values assigned to the backbone nuclei differs from the corresponding value in ubiquitin by an amount greater than R. Under these definitions, for $R > 0$, a correct assignment for an artificial peptide is not necessarily exactly the same as the assignment of the corresponding segment of ubiquitin. A consistent assignment is not necessarily a correct assignment, but a correct assignment is necessarily a consistent assignment.

For nonzero values of R, assignments that are consistent with SA, SB, and SC, but are not correct, may exist. Data sets SD through SG may then be used as additional constraints on the assignments. Since signals in SD through SG arise from both one-bond and two-bond polarization transfers, the signals are likely to vary significantly in intensity. Signals in these data sets are therefore not assumed to be quantitative. These data sets are used to constrain the assignments by imposing the criteria that all signals predicted by a tentative assignment be present in the data sets and that signals not predicted by a tentative assignment be absent from the data sets, to within the specified spectral resolution.

Sequential assignment algorithm

Simulations are carried out for artificial peptides of variable length L and with variable spectral resolution R. Three-dimensional chemical shift correlation data sets SA through SG are constructed from liquid-state chemical shift data for ubiquitin (Fig. 2). The sequential assignment algorithm is summarized in the flow chart in Fig. 3.

The algorithm begins by setting SA(1,1) to zero and assigning SA(1,1), SA(2,1), and SA(3,1) to N(1), $C^{\alpha}(1)$, and CO(1). Sequential assignments of chemical shifts to subsequent backbone nuclei are then attempted by randomly directed searches for signals in data sets SA, SB, and SC. Starting with a consistent assignment for the first J–1 amino acids, SB is used to assign N(J), SC is used to assign $C^{\alpha}(J)$, and SA is used to assign CO(J). The assignment is then checked for consistency with data sets SD through SG. If further sequential assignment becomes impossible before the assignment is complete (J = L), a new assignment attempt is initiated. The sequential assignment algorithm consists of the following steps:

(1) Select a peptide of length L from contiguous residues in the amino acid sequence of ubiquitin. Using the reported assignment of the backbone ^{15}N and ^{13}C NMR signals from ubiquitin, set up the seven 3D spectral data arrays SA through SG as summarized in Table 1. The ^{15}N signal from the amino terminus is assumed to be well resolved from all of the amide nitrogen signals, so SA(1,1), SE(L,2), SF(1,3), SF(L,1), SF(L,3), and SG(1,1) are set to 0.0.

(2) Specify a value for the assumed chemical shift resolution R.

(3) Begin the current assignment attempt. Set up an array A that will contain the tentative resonance assignment, with A(i,1), A(i,2), and A(i,3) equal to the assigned chemical shifts of N(i), $C^{\alpha}(i)$, and CO(i) for $1 \le i \le L$. Since the ^{15}N signal from the amino terminus is assumed to be resolved, set A(1,m) = SA(1,m) for $1 \le m \le 3$. Set J = 2, where J is the number of the next residue to be sequentially assigned.

(4) Search for a signal K_B in SB such that $|SB(K_B,1) - A(J-1,2)| \le R$ and $|SB(K_B,2) - A(J-1,3)| \le R$. Trial values of K_B are obtained with a random number generator, so that all possible orderings of the signals are sampled with equal a priori probabilities. If no such signal exists, go to step 3. If such a signal is found, set $A(J,1) = SB(K_B,3)$. Eliminate signal K_B from further consideration in the current assignment attempt.

(5) Search randomly for a signal K_C in SC such that $|SC(K_C,1) - A(J-1,3)| \le R$ and $|SC(K_C,2) - A(J,1)| \le R$. If no such signal exists, go to step 3. If such a signal is found, set $A(J,2) = SC(K_C,3)$. Eliminate signal K_C from further consideration in the current assignment attempt.

(6) Search randomly for a signal K_A in SA such that $|SA(K_A,1) - A(J-1,2)| \le R$ and $|SB(K_A,2) - A(J-1,3)| \le R$. If no such signal exists, go to step 3. If such a signal is found, set $A(J,3) = SA(K_A,3)$. Eliminate signal K_A from further consideration in the current assignment attempt.

(7) At this point, there is a tentative assignment for the first J residues. Check that the assignment for residue J is consistent with spectra SD and SE by searching for signals K_D and K_E such that $|SD(K_D,1) - A(J,3)| \le R$, $|SD(K_D,2) - A(J,1)| \le R$, $|SD(K_D,3) - A(J-1,3)| \le R$,

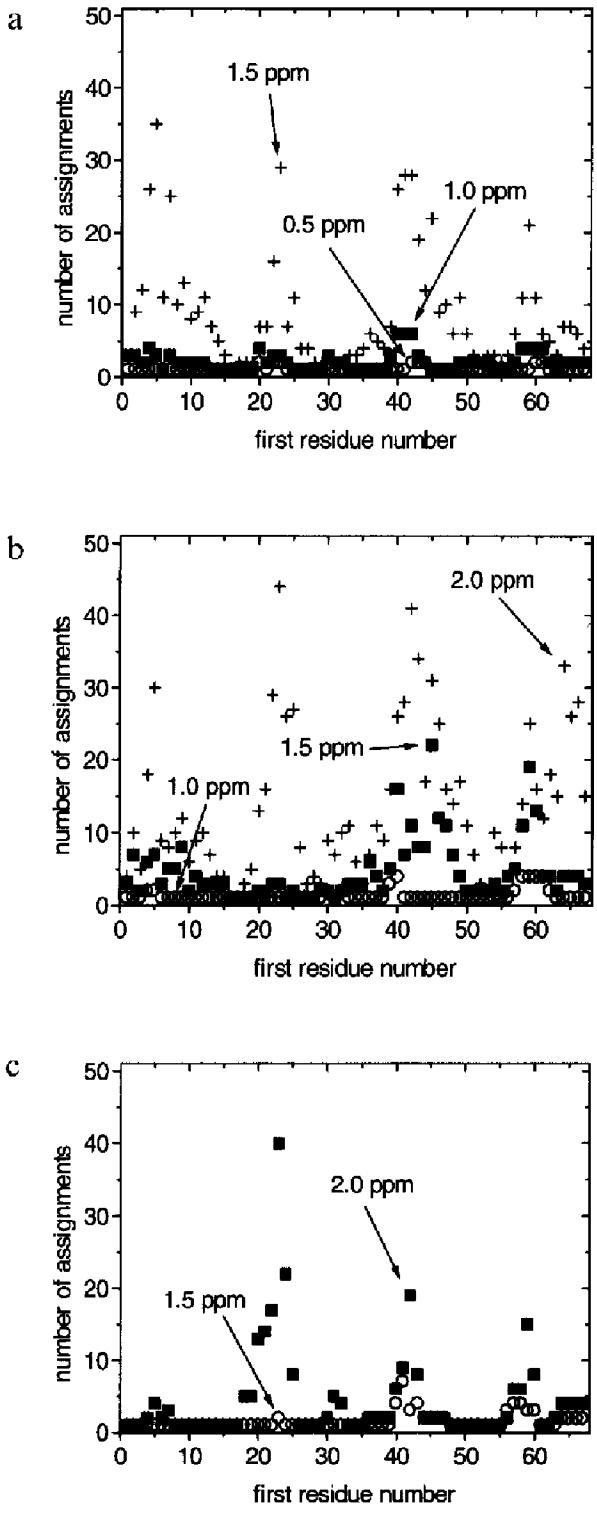

Fig. 4. Results of chemical shift assignment simulations for artificial peptides of length L = 10, derived from the ubiquitin sequence. (a) Number of assignments consistent with 3D data sets SA, SB, and SC as a function of the number of the first residue in the ubiquitin sequence for spectral resolutions R = 0.5 ppm (open circles), 1.0 ppm (filled squares), and 1.5 ppm (crosses). (b) Number of assignments consistent with 3D data sets SA, SB, SC, SD, and SE for R = 1.0 ppm (open circles), 1.5 ppm (filled squares), and 2.0 ppm (crosses). (c) Number of assignments consistent with 3D data sets SA, SB, SC, SD, SE, SF, and SG for R = 1.5 ppm (open circles) and 2.0 ppm (filled squares).

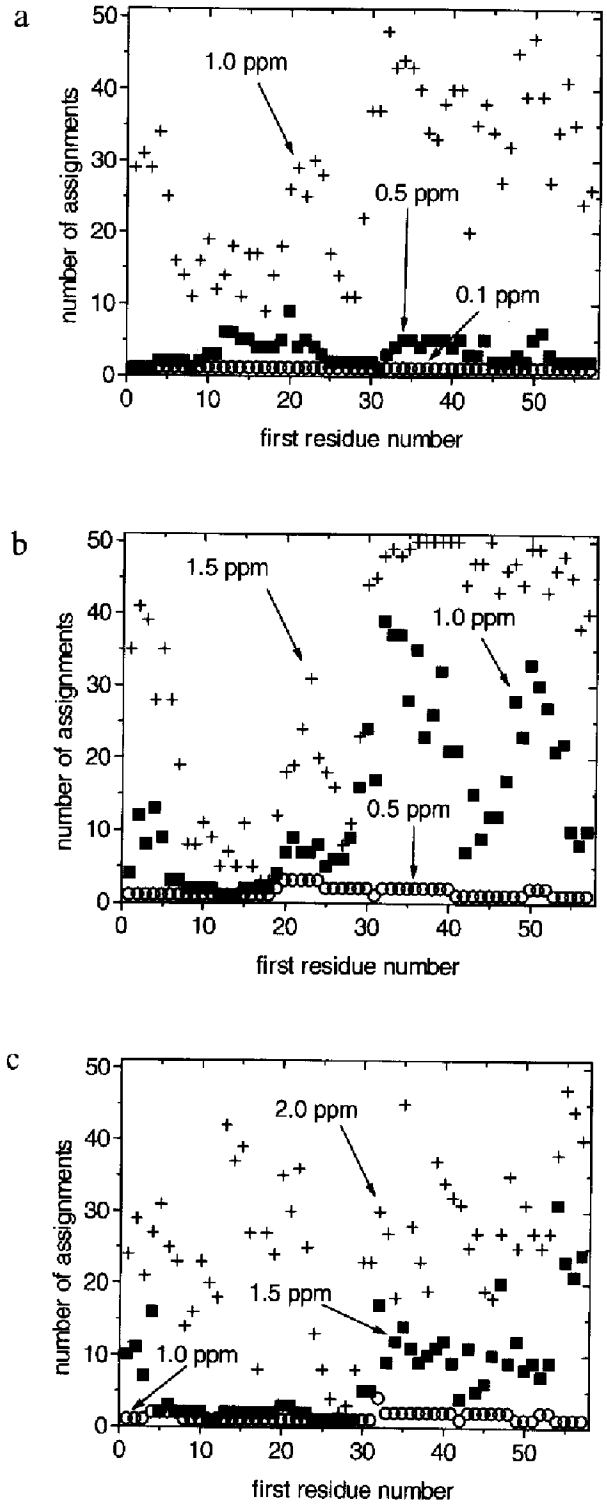

Fig. 5. Results of chemical shift assignment simulations for peptides of length L = 20, derived from the ubiquitin sequence. (a) Number of assignments consistent with 3D data sets SA, SB, and SC as a function of the number of the first residue in the ubiquitin sequence for spectral resolutions R = 0.1 ppm (open circles), 0.5 ppm (filled squares), and 1.0 ppm (crosses). (b) Number of assignments consistent with 3D data sets SA, SB, SC, SD, and SE for R = 0.5 ppm (open circles), 1.0 ppm (filled squares), and 1.5 ppm (crosses). (c) Number of assignments consistent with 3D data sets SA, SB, SC, SD, SE, SF, and SG for R = 1.0 ppm (open circles), 1.5 ppm (filled squares), and 2.0 ppm (crosses).

$|SE(K_E,1) - A(J-1,2)| \le R$, $|SE(K_E,2) - A(J,1)| \le R$, and $|SE(K_E,3) - A(J,2)| \le R$. If no such signals exist, go to step 3.

(8) Check that the assignment for residue J is consistent with spectra SF and SG by searching for signals K_F and K_G such that $|SF(K_F,1) - A(J,1)| \le R$, $|SF(K_F,2) - A(J-1,2)| \le R$, $|SF(K_F,3) - A(J-1,1)| \le R$, $|SG(K_G,1) - A(J-1,1)| \le R$, $|SG(K_G,2) - A(J-1,3)| \le R$, and $|SG(K_G,3) - A(J,1)| \le R$. If no such signals exist, go to step 3.

(9) Increment J by one. If $J \le L$, go to step 4.

(10) At this point, there is a tentative complete assignment for the peptide backbone. Check that there are no unexplained signals in spectra SD and SE, as follows. For each signal i in SD with $1 \le i \le 2L-2$, search for a residue number j such that either $|SD(i,1) - A(j+1,3)| \le R$, $|SD(i,2) - A(j+1,1)| \le R$, and $|SD(i,3) - A(j,3)| \le R$, or $|SD(i,1) - A(j,3)| \le R$, $|SD(i,2) - A(j+1,1)| \le R$, and $|SD(i,3) - A(j,3)| \le R$. For each signal i in SE with $1 \le i \le 2L-1$, search for a residue number j such that either $|SE(i,1) - A(j,2)| \le R$, $|SE(i,2) - A(j+1,1)| \le R$, and $|SE(i,3) - A(j+1,2)| \le R$, or $|SE(i,1) - A(j,2)| \le R$, $|SE(i,2) - A(j,1)| \le R$, and $|SE(i,3) - A(j,2)| \le R$. If any of these criteria cannot be satisfied, go to step 3.

(11) Check that there are no unexplained signals in spectra SF and SG, as follows. For each signal i in SF with $1 \le i \le 2L-1$, search for a residue number j such that either $|SF(i,1) - A(j+1,1)| \le R$, $|SF(i,2) - A(j,2)| \le R$, and $|SF(i,3) - A(j,1)| \le R$, or $|SF(i,1) - A(j,1)| \le R$, $|SF(i,2) - A(j,2)| \le R$, and $|SF(i,3) - A(j,1)| \le R$. For each signal i in SG with $1 \le i \le 2L-2$, search for a residue number j such that either $|SG(i,1) - A(j,1)| \le R$, $|SG(i,2) - A(j,3)| \le R$, and $|SG(i,3) - A(j+1,1)| \le R$, or $|SG(i,1) - A(j+1,1)| \le R$, $|SG(i,2) - A(j,3)| \le R$, and $|SG(i,3) - A(j+1,1)| \le R$. If any of these criteria cannot be satisfied, go to step 3.

(12) At this point, the assignment in A is complete and is *consistent* with all of the 3D spectral data to within the specified resolution. Check whether this assignment is also *correct* to within the specified resolution by comparing A with SA. The assignment is correct if $|A(i,1) - SA(i,1)| \le R$, $|A(i,2) - SA(i,2)| \le R$, and $|A(i,3) - SA(i,3)| \le R$ for all i with $1 \le i \le L$.

(13) Check whether the assignment in A is distinct within the specified resolution from previously determined consistent assignments A' for the same peptide. Two assignments are different if $|A(i,j) - A'(i,j)| > R$ for any i and j.

(14) Record the number of correct assignments and distinct consistent assignments that have been found thus far. If another assignment attempt is required, go to step 3.

If signals that result from two-bond polarization transfers are not used as constraints on the assignment, steps 7, 8, 10, and 11 are omitted.

A FORTRAN program was written to implement the sequential assignment algorithm and was executed on a

Sun Microsystems Sparcstation 20 Model 61 computer. Execution times depend strongly on the values of L, R, and I and on the number of 3D data sets used as constraints. As an example, simulations that use all data sets SA through SG with L = 30 and R = 1.5 ppm require an average of 22 s of processor time to find one complete consistent assignment.

In principle, the algorithm described above and in Fig. 3 will find all possible consistent assignments if enough attempts are made. However, different consistent assignments are generally found with different frequencies. If the path to a particular complete, consistent assignment involves many steps at which there are alternative assignment choices (some of which eventually lead to complete assignments and some to dead ends), then the frequency will be low. Thus, the algorithm necessarily generates a lower bound on the actual number of distinct, consistent assignments.

Simulation results

The algorithm described above is used to find resonance assignments for peptides of various lengths L and with various assumed spectral resolution values R. For each value of L, up to 77–L peptides may be considered by choosing the initial residue number I in the ubiquitin sequence between 1 and 77–L. For each peptide and each R value, 50 independent determinations of a complete assignment consistent with the 3D spectral data are carried out. The 50 assignments are compared with one another to determine the number of distinct consistent assignments (according to the criterion in step 13 of the assignment algorithm). Representative results of these simulations are presented in Figs. 4–6 as plots of the number of distinct assignments versus the initial residue number. The results for L = 10 (Fig. 4), 20 (Fig. 5), and 30 (Fig. 6) are shown. Simulations were carried out with the inclusion of only data sets SA through SC (parts a of Figs. 4–6), only data sets SA through SE (parts b), and data sets SA through SG (parts c).

Except in very rare cases, when only one consistent assignment is found for given values of L, R, and I, that assignment is also a correct assignment. When more than one distinct, consistent assignment is found, no more than one of these assignments is correct. As explained above, the numbers plotted in Figs. 4–6 represent lower bounds on the actual numbers of distinct, consistent assignments. However, when the number of distinct, consistent assignments discovered in 50 attempts is small (< 5, roughly speaking), the actual number of distinct, consistent assignments is unlikely to be much larger. On the other hand, when the number of distinct, consistent assignments discovered in 50 attempts is large (> 20), the actual number of distinct, consistent assignments may well be significantly larger.

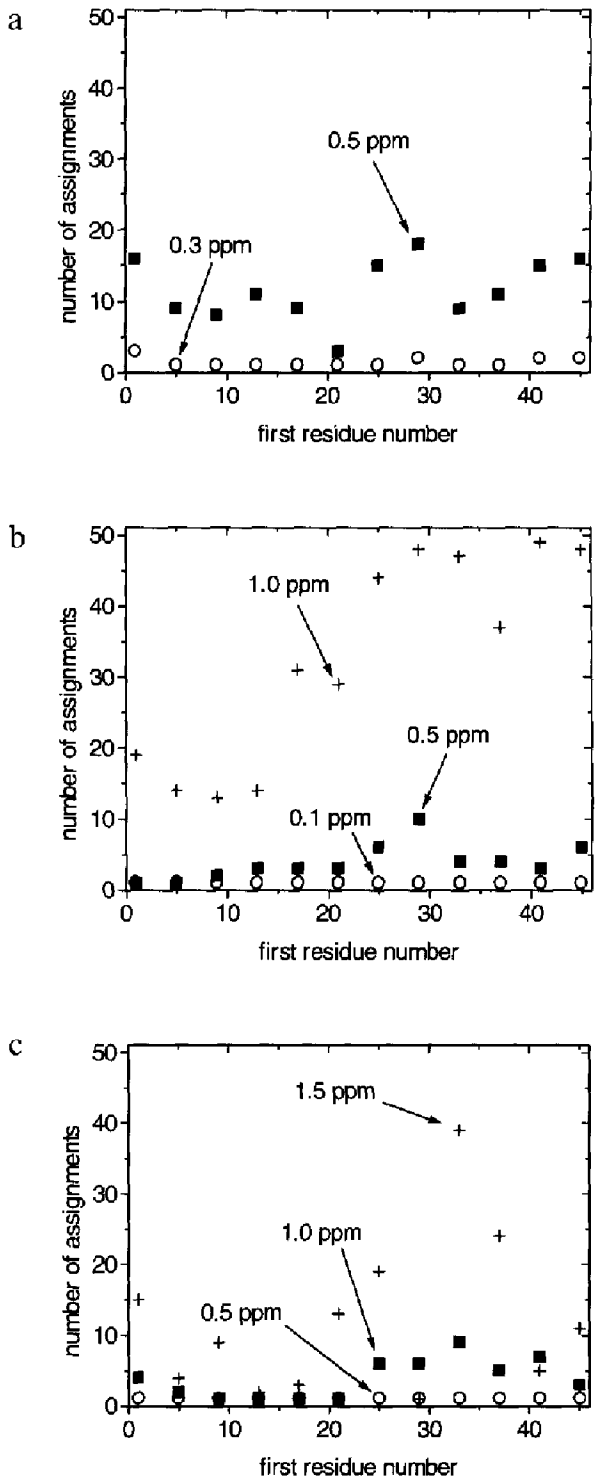

Fig. 6. Results of chemical shift assignment simulations for peptides of length L = 30, derived from the ubiquitin sequence. (a) Number of assignments consistent with 3D data sets SA, SB, and SC as a function of the number of the first residue in the ubiquitin sequence for spectral resolutions R = 0.3 ppm (open circles) and 0.5 ppm (filled squares). (b) Number of assignments consistent with 3D data sets SA, SB, SC, SD, and SE for R = 0.1 ppm (open circles), 0.5 ppm (filled squares), and 1.0 ppm (crosses). (c) Number of assignments consistent with 3D data sets SA, SB, SC, SD, SE, SF, and SG for R = 0.5 ppm (open circles), 1.0 ppm (filled squares), and 1.5 ppm (crosses). In these simulations, the first residue number is incremented by four, rather than one, in order to reduce the total computation time.

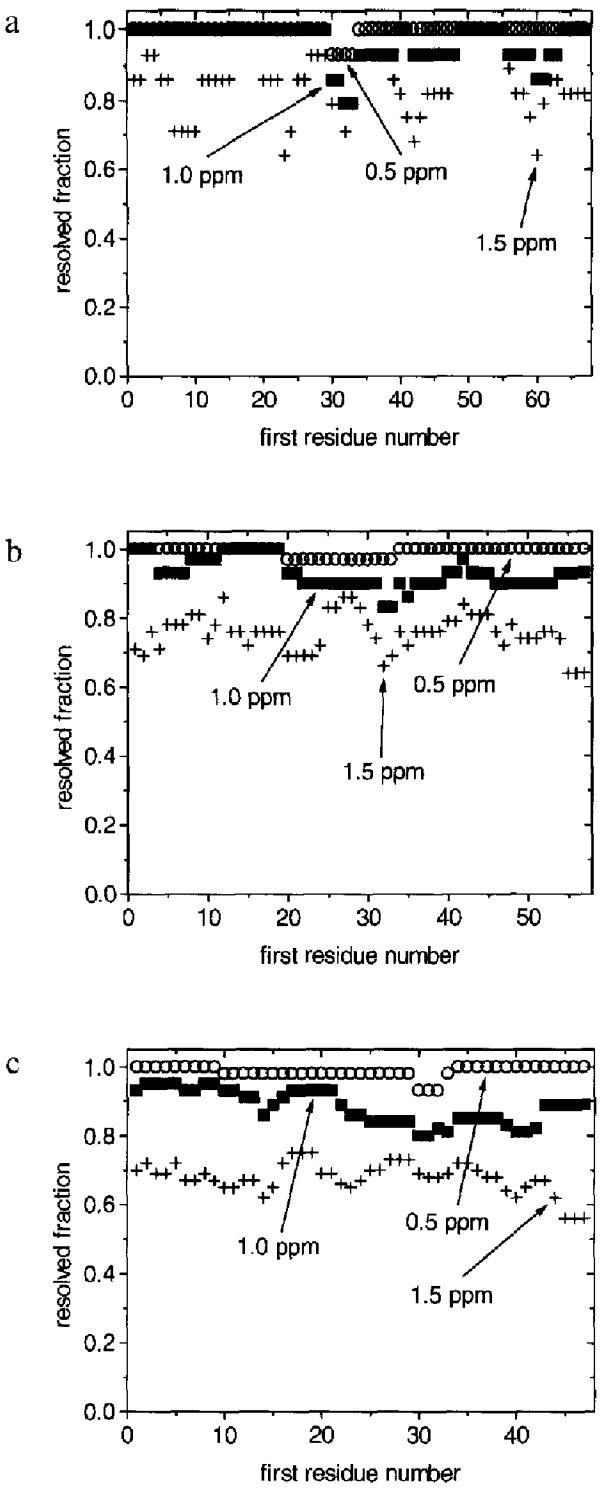

Fig. 7. Fraction of signals that are resolved in 3D data sets SA, SB, and SC as a function of the first residue in the ubiquitin sequence. (a) L = 10; (b) L = 20; (c) L = 30. Spectral resolution R = 0.5 ppm (open circles), 1.0 ppm (filled squares), and 1.5 ppm (crosses).

Figures 4–6 display three expected qualitative trends: (i) for a given value of L, the number of distinct, consistent assignments increases as R increases; (ii) for a given value of R, the number of distinct, consistent assignments increases as L increases; and (iii) for given values of L and R, the number of distinct, consistent assignments decreases as more 3D data sets are included as constraints on the assignments.

In more quantitative terms, the results in Figs. 4–6 demonstrate the existence of an approximate threshold for R, denoted by R_t, above which the number of distinct, consistent assignments becomes greater than one for many choices of the initial residue number. R_t is approximately the maximum value of the ^{13}C and ^{15}N NMR line widths that would permit the unique assignment of the backbone resonances in the series of 3D experiments considered in these simulations. R_t depends on L and on the number of 3D data sets that are included as constraints. For L = 10, R_t is roughly 0.5–1.0 ppm when only data sets SA through SC are used, roughly 1.0–1.5 ppm when only data sets SA through SE are used, and roughly 1.5–2.0 ppm when data sets SA through SG are used. For L = 20, R_t is less than 0.5 ppm when only data sets SA through SC are used, roughly 0.5 ppm when only data sets SA through SE are used, and roughly 1.0 ppm when data sets SA through SG are used. For L = 30, R_t is roughly 0.3 ppm when only data sets SA through SC are used, roughly 0.5 ppm when only data sets SA through SE are used, and roughly 1.0 ppm when data sets SA through SG are used.

Certain regions of the ubiquitin sequence appear to be more difficult to assign than others. This observation is particularly striking in Fig. 4c, where the number of distinct, consistent assignments with R = 2.0 ppm is greater than one only when the initial residue number is in the intervals 4–7, 18–25, 30–32, 36–47, 56–60, and 63–67. These intervals do not correspond in an obvious way to the secondary structure elements of ubiquitin (Vijay-Kumar et al., 1987).

Discussion

Limits imposed by spectral resolution

The principal conclusion to be drawn from the results in Figs. 4–6 is that, under the assumptions employed in these simulations, the length of a peptide for which a complete, unique assignment of ^{13}C and ^{15}N backbone resonances in solid-state NMR spectra is feasible is strongly limited by the spectral resolution. A resolution better than 1.0–2.0 ppm is generally required for a unique assignment for a peptide of 10 amino acid residues, depending on the number of 3D data sets that are used as constraints. For peptides of 20 or 30 amino acid residues, resolutions better than 0.5–1.5 ppm or 0.5–1.0 ppm, respectively, are generally required.

As an indication of the impact of the spectral resolution on the 3D spectra themselves, Fig. 7 presents plots of the fraction of signals in 3D data sets SA, SB, and SC that are resolved from all other signals for L = 10, 20, and 30 and for R = 0.5, 1.0, and 1.5 ppm, as functions of the

initial residue number I. The criterion for a signal i in SA to be resolved, for example, is that at least one of the inequalities $|SA(i,1) - SA(j,1)| \geq R$, $|SA(i,2) - SA(j,2)| \geq R$, or $|SA(i,3) - SA(j,3)| \geq R$ be satisfied for all $j \neq i$. In other words, a resolved signal must differ from all other signals by at least R in at least one chemical shift dimension. The fraction of resolved signals is simply the ratio of the number of resolved signals in SA, SB, and SC to the total number of signals. The plots in Fig. 7 permit the formulation of quantitative estimates of the fraction of resolved signals that is required for a unique assignment. Interestingly, if only data sets SA, SB, and SC are used in the assignment algorithm, there need not be a unique consistent assignment even if all the signals in SA, SB, and SC are resolved, as may be seen by a comparison of Figs. 7a and 4a. On the other hand, if all 3D data sets are used, there may be a unique assignment even if the fraction of signals in data sets SA, SB, and SC that are not resolved is as large as 0.3 (for $L = 10$), as may be seen by a comparison of Figs. 7a and 4c. Larger fractions of unresolved signals or longer peptides generally lead to numerous consistent assignments.

The correspondence between the resolution parameter R and the NMR line widths depends somewhat on the line shapes and, in practice, on the signal-to-noise ratio. A 1.0 ppm splitting between NMR lines of equal intensity in a 1D spectrum cannot be resolved with 1.5 ppm Gaussian line widths or with 2.0 ppm Lorentzian line widths.

Experimentally achievable line widths depend on a number of factors (VanderHart et al., 1981). ^{13}C line widths less than 10 Hz (0.1 ppm in a 9.4 T magnetic field) have recently been observed in polycrystalline samples of small molecules without large magnetic susceptibility anisotropies (Griffin, 1996; Zilm, 1996). These surprisingly narrow lines result from a combination of MAS, high proton decoupling fields, phase-modulated decoupling techniques (Bennett et al., 1995), and a high degree of structural order and rigidity. Similarly, narrow lines may well be achievable in natural-abundance ^{13}C and ^{15}N MAS spectra of much larger compounds in polycrystalline form. In uniformly labeled compounds, ^{13}C-^{13}C and ^{13}C-^{15}N scalar couplings contribute approximately 30–50 Hz to the line widths. The effects of ^{13}C-^{15}N scalar couplings may be removed by heteronuclear decoupling techniques, using continuous rf irradiation or rf pulse sequences. In principle, selective irradiations or pulse sequences could be designed to remove the effects of ^{13}C-^{13}C scalar couplings as well. Residual homonuclear dipole–dipole couplings may also make significant contributions to the line widths in uniformly labeled samples at moderate MAS speeds. Very high speed MAS would reduce these contributions. Thus, it is possible that complete, unique backbone resonance assignments can be obtained for uniformly labeled polycrystalline (or otherwise highly ordered) peptides with $L \gg 30$ using 3D solid-state NMR measurements of the type assumed in the simulations.

In the case of noncrystalline samples of specifically labeled peptides and proteins, in the form of frozen solutions, in noncrystalline solids, or in association with biological membranes, the line widths observed to date in MAS NMR spectra have been consistently 1 ppm or more. These line widths are typically dominated by inhomogeneous broadening that results from structural disorder. Cole et al. (1988) reported approximately 0.4 ppm ^{13}C line widths and 1.3 ppm ^{15}N line widths in MAS spectra of specifically labeled staphylococcal nuclease in polycrystalline form, and line widths at least several times greater in lyophilized form. Hing et al. (1994) reported approximately 6 ppm line widths in ^{15}N MAS spectra of uniformly ^{15}N-labeled glutamine-binding protein in lyophilized form. Hing and Schaefer (1993) observed approximately 1.0 ppm line widths in ^{13}C MAS spectra of ^{13}C-labeled gramicidin A in dimyristoylphosphatidylcholine (DMPC) multilamellar dispersions at 44 °C, where there is substantial motional narrowing. Christensen and Schaefer (1993) reported approximately 1.2 ppm line widths in ^{31}P MAS spectra of the complex of shikimate 3-phosphate and phosphoenolpyruvate with 5-enolpyruvylshikimate-3-phosphate synthase in samples that were lyophilized after relatively rapid freezing, and substantially larger line widths in spectra of samples that were lyophilized after slow freezing. Smith et al. (1996) reported approximately 1.0 ppm line widths in ^{13}C MAS spectra of a ^{13}C-labeled peptide derived from the neu/erbB-2 receptor incorporated into DMPC:dimyristoylphosphatidylserine multilamellar dispersions at –25 °C. Lazo et al. (1993) found approximately 2.5 ppm line widths in ^{15}N MAS spectra of specifically ^{15}N-labeled gramicidin A in DMPC in lyophilized form after rapid freezing. Griffiths et al. (1994) reported lines as narrow as 1.4 ppm in ^{13}C MAS spectra of specifically labeled bacteriorhodopsin in purple membrane samples at –60 °C. Tomita et al. (1994) observed approximately 2 ppm line widths in ^{13}C MAS spectra of $^{13}C_2$-phosphoglycolic acid bound to triose phosphate isomerase in samples prepared by precipitation with polyethylene glycol, and 4–6 ppm line widths in samples prepared by lyophilization. Zysmilich and McDermott (1994) reported line widths between 1 and 3 ppm in optically pumped ^{15}N MAS spectra of uniformly ^{15}N-labeled photosynthetic reaction centers at –45 °C.

Structural disorder contributes to MAS NMR line widths through the dependence of the isotropic chemical shifts on local structure. Two components of structural disorder can be distinguished. The first component is disorder in intermolecular packing, e.g., solvent disorder in frozen solutions or disorder in intermolecular contacts in lyophilized or precipitated proteins. The effect of this component is difficult to estimate. The second component of structural disorder is disorder in the conformation of

the peptide or protein itself. This contribution can be estimated by ab initio calculations of conformation-dependent chemical shifts (Ando et al., 1984; de Dios et al., 1993; Jiao et al., 1993; de Dios and Oldfield, 1994; Le et al., 1995). Calculations of the dependence of carbonyl and C^{α} ^{13}C chemical shifts on the peptide backbone dihedral angles ϕ and ψ show that the angular derivatives of the chemical shifts typically range from 0 to 0.1 ppm per degree. The amplitude of the static variations of backbone dihedral angles in noncrystalline solid peptides and proteins is uncertain, but a crude estimate can be obtained by assuming that these static angular variations are comparable to the dynamic angular fluctuations that occur in isotropic solutions. Analyses of spin relaxation measurements on proteins in solution within the framework of the 'model-free' formalism commonly lead to generalized order parameters $S^2 \approx 0.9$ for N-H vectors in well-structured regions of the protein backbone (Torchia et al., 1993). Order parameters of 0.9 imply root-mean-squared (rms) variations of approximately 10° in the N-H bond directions. If the variations in a particular N-H bond direction result from uncorrelated fluctuations in the ϕ and ψ angles of the corresponding amino acid residue, then these dihedral angle fluctuations are approximately 8° in rms amplitude. Dihedral angle fluctuations in the range from 10° to 20° in rms amplitude are typically observed in molecular dynamics simulations on hydrated proteins at 300 K (D.J. Tobias, 1996, personal communication). If these dihedral angle fluctuations are simply frozen in place in noncrystalline solids, then inhomogeneous broadening of the NMR lines on the order of 1 ppm or more may be expected.

A certain degree of structural and conformational disorder is inevitable in systems that are not crystalline. In the light of the simulated assignment results described above, the importance of minimizing disorder-induced inhomogeneous broadening in the preparation of uniformly labeled samples for multidimensional solid-state NMR measurements is evident.

Supplemental assignment strategies

As should be obvious, not all available spectroscopic information is utilized in the assignment algorithm described above. This algorithm was designed to rely on experimental measurements that currently appear to be feasible, although not necessarily simple. The assignment algorithm was also chosen so that it would be applicable to any peptide in a clearly defined, automated fashion, irrespective of the amino acid sequence. Additional sources of constraints on the assignments and additional experimental measurements that may permit the assignment of backbone resonances in larger peptides and proteins are briefly discussed in this section.

The simulations described above do not make use of residue-specific information that may be useful in the analysis of real experimental data. The fact that different amino acids exhibit different chemical shift distributions (Wishart et al., 1991) is not exploited. For example, glycine residues exhibit C^{α} shifts near 45 ppm, generally smaller than (i.e., upfield of) those of other amino acids (Fig. 2a). Glycine amide N shifts also tend to be relatively small. Proline N and C^{α} shifts tend to be relatively large. These effects are likely to be very useful as additional constraints on assignments. As a quantitative indication of the impact of constraints based on these effects, Fig. 8 shows the results of assignment simulations in which the condition that C^{α} chemical shifts assigned to glycine residues be less than 50 ppm is used as an additional constraint, along with 3D data sets SA through SG. The results shown in Fig. 8 are for a peptide with L = 20,

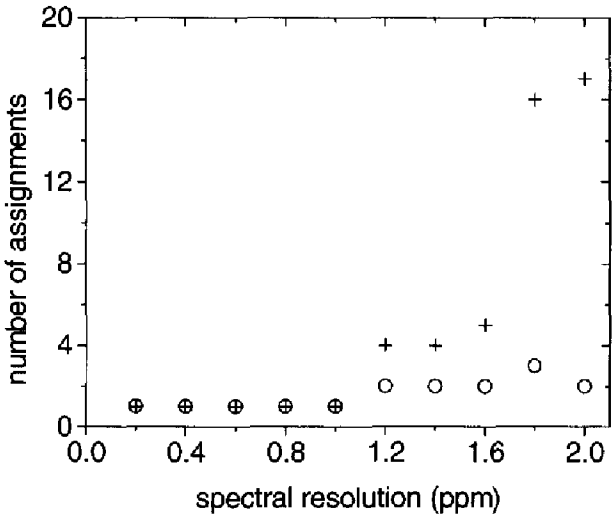

Fig. 8. Number of assignments consistent with 3D data sets SA, SB, SC, SD, SE, SF, and SG for an artificial peptide derived from residues 30–49 of the ubiquitin sequence as a function of the spectral resolution. Results of simulations with (open circles) and without (crosses) the additional constraint that the α-carbons of glycine residues at positions 6 and 18 of the peptide have chemical shifts less than 50 ppm.

derived from residues 30–49 of ubiquitin. This peptide contains two glycines, at positions 6 and 18. The use of this additional residue-specific chemical shift constraint significantly reduces the number of consistent assignments for R values above 1.0 ppm.

The NMR properties of amino acid side chains may also be useful. For example, since only glycine C^{α} carbons have two directly bonded protons and only proline nitrogens have no directly bonded protons, spectral editing techniques (Wu et al., 1994) may permit the selection and assignment of glycine or proline signals. The inclusion of chemical shift correlations with side-chain ^{13}C signals may facilitate the assignment of backbone signals. For example, the C^{β} carbons of alanine residues have smaller shifts (near 17 ppm) and the C^{β} carbons of serine and threonine residues have larger shifts (> 60 ppm) than those of other amino acids (Fig. 2b). Three-dimensional data sets that correlate $C^{\beta}(i)$, $C^{\alpha}(i)$, and N(i) shifts or $C^{\beta}(i)$, $C^{\alpha}(i)$, and CO(i) shifts would therefore be particularly useful in the assignment of alanine, serine, and threonine residues. These data sets could be obtained from measurements of the type shown in Fig. 1a or from 3D ^{13}C-^{13}C-^{13}C chemical shift correlation measurements. Further simulations would be required to assess the impact of these data sets on the assignment problem.

Higher dimensional (e.g., 4D) measurements may also permit the complete, unique assignment of larger peptides. The feasibility of higher dimensional measurements depends primarily on considerations of sensitivity. These considerations are likely to be very important in real experiments. The sensitivity of higher dimensional measurements would be substantially lower than that of 3D measurements, because the need to maximize the spectral resolution in each dimension implies that the loss of signal amplitude due to transverse relaxation in each dimension would necessarily be appreciable. If polarization transfers between evolution periods were mediated by dipole–dipole couplings and were limited to one-bond transfers by the use of relatively short polarization transfer periods, then the efficiency of the polarization transfers would necessarily be less than optimal, thus reducing the sensitivity. The total measurement time for an n-dimensional experiment can be estimated by assuming that (i) the sample size is such that only one FID is required to produce adequate signal-to-noise for each isotopically labeled site in a 1D spectrum (corresponding to 10 mg of a 6 kDa protein with 1 ppm ^{13}C line widths in a 9.4 T magnet, roughly speaking); (ii) the spin-lattice relaxation times permit signal averaging with a 1 s recycle delay; (iii) the maximum value of each evolution period is such that, on average, signals are reduced by a factor of 2 by decay in each evolution period; and (iv) the efficiency of each polarization transfer is a constant value E. Under these conditions, the measurement time required for adequate signal-to-noise in an n-dimensional spectrum becomes approximately $(2\sqrt{2}/E)^{2n-2}$ s. For a 3D experiment with E = 20%, this is about 11 h. For a 4D experiment with E = 20%, this is about 93 days.

Finally, it may be possible to make use of proton chemical shifts. At present, relatively little is known about the achievable line widths in proton MAS spectra of peptides and proteins. Very rapid MAS, very high magnetic fields, multiple-pulse homonuclear decoupling sequences, and selective or random fractional deuteration may be used to enhance the resolution in the proton spectra. If some combination of these methods resulted in line widths that were significantly less than 4 ppm for N-H protons and 2 ppm for C^{α}-H protons (the respective chemical shift ranges), and if polarization transfers between protons and ^{13}C and ^{15}N nuclei could be restricted to one-bond or two-bond transfers, then information contained in proton chemical shifts would become valuable in the assignment process.

Conclusions

The assignment simulations presented in this paper demonstrate that spectral resolution is a critical factor in assessing the feasibility of complete backbone chemical shift assignments from multidimensional chemical shift correlation measurements on unoriented, uniformly ^{13}C- and ^{15}N-labeled peptides and proteins in solid-state NMR. In the absence of residue-specific chemical shift information, information from side-chain NMR signals, and information from proton chemical shifts, the largest peptide for which complete, unique backbone assignments can be made from 3D chemical shift correlation spectra is likely to be in the range of 20–30 amino acid residues in length if the resolution in the spectra is roughly 1 ppm. In such a case, complete assignment of backbone chemical shifts requires the use of chemical shift correlation data derived from both one-bond and two-bond polarization transfers. Significant increases in the upper limit on peptide or protein size appear to be possible if residue-specific chemical shift information and side-chain signals can be used as additional assignment constraints. Additional simulations and experiments are required before a definite upper limit can be established. It is possible that complete backbone chemical shift assignments will be attainable experimentally for proteins of 50–100 residues in noncrystalline, unoriented samples if additional spectroscopic information is available.

Acknowledgements

I thank Henry W. Long, Andrew E. Bennett, and David P. Weliky for helpful comments on a preliminary presentation of this work and Douglas J. Tobias for providing unpublished data on dihedral angle fluctuations in protein dynamics simulations.

References

Ando, I., Saito, H., Tabeta, R., Shoji, A. and Ozaki, T. (1984) *Macromolecules*, **17**, 457–461.

Baldus, M., Tomaselli, M., Meier, B.H. and Ernst, R.R. (1994) *Chem. Phys. Lett.*, **230**, 329–336.

Barrett, S.E., Tycko, R., Pfeiffer, L.N. and West, K.W. (1994) *Phys. Rev. Lett.*, **72**, 1368–1371.

Bax, A. and Grzesiek, S. (1993) In *NMR of Proteins* (Eds, Clore, G.M. and Gronenborn, A.), CRC Press, Boca Raton, FL, U.S.A., pp. 33–52.

Bennett, A.E., Ok, J.H., Griffin, R.G. and Vega, S. (1992) *J. Chem. Phys.*, **96**, 8624–8627.

Bennett, A.E., Becerra, L.R. and Griffin, R.G. (1994) *J. Chem. Phys.*, **100**, 812–814.

Bennett, A.E., Rienstra, C.M., Auger, M., Lakshmi, K.V. and Griffin, R.G. (1995) *J. Chem. Phys.*, **103**, 6951–6958.

Boender, G.J., Raap, J., Prytulla, S., Oschkinat, H. and De Groot, H.J.M. (1995) *Chem. Phys. Lett.*, **237**, 502–508.

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

Cole, H.B.R., Sparks, S.W. and Torchia, D.A. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 6362–6365.

de Dios, A.C., Pearson, J.G. and Oldfield, E. (1993) *J. Am. Chem. Soc.*, **115**, 9768–9773.

de Dios, A.C. and Oldfield, E. (1994) *J. Am. Chem. Soc.*, **116**, 5307–5314.

Ernst, R.R., Bodenhausen, G. and Wokaun, A. (1987) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Oxford University Press, Oxford, U.K.

Fujiwara, T., Sugase, K., Kainosho, M., Ono, A. and Akutsu, H. (1995) *J. Am. Chem. Soc.*, **117**, 11351–11352.

Gerfen, G.J., Becerra, L.R., Hall, D.A., Griffin, R.G., Temkin, R.J. and Singel, D.J. (1995) *J. Chem. Phys.*, **102**, 9494–9497.

Gregory, D.M., Mitchell, D.J., Stringer, J.A., Kiihne, S., Shiels, J.C., Callahan, J., Mehta, M.A. and Drobny, G.P. (1995) *Chem. Phys. Lett.*, **246**, 654–663.

Griffin, R.G. (1996) Presentation at the 37th Experimental Nuclear Magnetic Resonance Conference, Pacific Grove, CA, U.S.A., March 17–22, 1996.

Griffiths, J.M., Lakshmi, K.V., Bennett, A.E., Raap, J., Van der Wielen, C.M., Lugtenburg, J., Herzfeld, J. and Griffin, R.G. (1994) *J. Am. Chem. Soc.*, **116**, 10178–10181.

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

Hing, A.W. and Schaefer, J. (1993) *Biochemistry*, **32**, 7593–7604.

Hing, A.W., Vega, S. and Schaefer, J. (1993) *J. Magn. Reson.*, **A103**, 151–162.

Hing, A.W., Tjandra, N., Cottam, P.F., Schaefer, J. and Ho, C. (1994) *Biochemistry*, **33**, 8651–8661.

Jelinek, R., Ramamoorthy, A. and Opella, S.J. (1995) *J. Am. Chem. Soc.*, **117**, 12348–12349.

Jiao, D., Barfield, M. and Hruby, V.J. (1993) *J. Am. Chem. Soc.*, **115**, 10883–10887.

Kolbert, A.C., Grandinetti, P.J., Baldwin, M., Prusiner, S.B. and Pines, A. (1994) *J. Am. Chem. Soc.*, **98**, 7936–7938.

Lazo, N.D., Hu, W., Lee, K.-C. and Cross, T.A. (1993) *Biochem. Biophys. Res. Commun.*, **197**, 904–909.

Le, H.-B., Pearson, J.G., de Dios, A.C. and Oldfield, E. (1995) *J. Am. Chem. Soc.*, **117**, 3800–3807.

Menger, E.M., Vega, S. and Griffin, R.G. (1986) *J. Am. Chem. Soc.*, **108**, 2215–2218.

Ok, J.H., Spencer, R.G.S., Bennett, A.E. and Griffin, R.G. (1992) *Chem. Phys. Lett.*, **197**, 389–395.

Raleigh, D.P., Harbison, G.S., Neiss, T.G., Roberts, J.E. and Griffin, R.G. (1987) *Chem. Phys. Lett.*, **138**, 285–290.

Ramamoorthy, A., Wu, C.H. and Opella, S.J. (1995) *J. Magn. Reson.*, **B107**, 88–90.

Smith, S.O., Smith, C.S. and Bormann, B.J. (1996) *Nat. Struct. Biol.*, **3**, 252–258.

Sun, B.-Q., Costa, P.R., Kocisko, D., Lansbury, P.T. and Griffin, R.G. (1995) *J. Chem. Phys.*, **102**, 702–707.

Tomaselli, M., Meier, B.H., Baldus, M., Eisenegger, J. and Ernst, R.R. (1994) *Chem. Phys. Lett.*, **225**, 131–139.

Tomita, Y., O'Connor, E.J. and McDermott, A. (1994) *J. Am. Chem. Soc.*, **116**, 8766–8771.

Torchia, D.A., Nicholson, L.K., Cole, H.B.R. and Kay, L.E. (1993) In *NMR of Proteins* (Eds, Clore, G.M. and Gronenborn, A.), CRC Press, Boca Raton, FL, U.S.A., pp. 190–219.

Tycko, R. and Dabbagh, G. (1990) *Chem. Phys. Lett.*, **173**, 461–465.

Tycko, R. and Dabbagh, G. (1991) *J. Am. Chem. Soc.*, **113**, 9444–9448.

Tycko, R. and Smith, S.O. (1993) *J. Am. Chem. Soc.*, **98**, 932–943.

Tycko, R. (1994) *J. Am. Chem. Soc.*, **116**, 2217–2218.

Tycko, R., Barrett, S.E., Dabbagh, G., Pfeiffer, L.N. and West, K.W. (1995) *Science*, **268**, 1460–1463.

VanderHart, D.L., Earl, W.L. and Garroway, A.N. (1981) *J. Magn. Reson.*, **44**, 361–401.

Vijay-Kumar, S., Bugg, C.E. and Cook, W.J. (1987) *J. Mol. Biol.*, **194**, 531–544.

Wang, A.C., Grzesiek, S., Tschudin, R., Lodi, P.J. and Bax, A. (1995) *J. Biomol. NMR*, **5**, 376–382.

Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) *J. Mol. Biol.*, **222**, 311–333.

Wu, X., Burns, S.T. and Zilm, K.W. (1994) *J. Magn. Reson.*, **A111**, 29–36.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York, NY, U.S.A., pp. 130–161.

Zilm, K.W. (1996) Presentation at the 37th Experimental Nuclear Magnetic Resonance Conference, Pacific Grove, CA, U.S.A., March 17–22, 1996.

Zysmilich, M.G. and McDermott, A. (1994) *J. Am. Chem. Soc.*, **116**, 8362–8363.