J|A|C|S
COMMUNICATIONS

Published on Web 10/19/2004

Mass Spectrometry Assisted Assignment of NMR Resonances in ^{15}N Labeled Proteins

Lianmei Feng, Ron Orlando, and James H. Prestegard*

Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602

Received July 14, 2004; E-mail: jpresteg@ccrc.uga.edu

Nuclear magnetic resonance (NMR) has proven a useful tool for the structural characterization of biomolecules, particularly when those molecules fail to form crystals suitable for diffraction studies.[1] However, until recently, characterization has been limited to rather small molecules because of the near linear increase of NMR line widths with molecular weight. Transverse relaxation optimized spectroscopy (TROSY) methods that capitalize on interference between dipolar interactions between ^{1}H−^{15}N spin pairs in amide bonds of proteins and the inherent chemical shift anisotropy of the ^{1}H and ^{15}N sites have changed this. Dramatic improvements in line widths for ^{1}H−^{15}N cross-peaks in TROSY versions of heteronuclear single quantum coherence (HSQC) spectra have now been demonstrated for a large number of proteins.[2] It is also true that one requires only ^{15}N isotopic labeling in this basic experiment, something that can be advantageous in studying proteins that are more difficult to express in isotopically labeled forms. There is, therefore, good reason to think about structure determination strategies that rely more heavily on the basic ^{1}H−^{15}N HSQC experiment. One problem that must be overcome in implementing new strategies is assignment of cross-peaks without the aid of the triple resonance experiments normally used for resonance assignment. Here we develop a novel approach for protein resonance assignment that relies on integration of NMR and mass spectrometry (MS) methods. The approach relies on the fact that both NMR[3−6] and MS[7−9] are able to monitor rates of exchange of amide protons for water deuterons.[10] Correlating the rates can connect cross-peak positions from NMR data with fragment masses from MS data to support sequential assignment. The example provided here is for a small model protein, ubiquitin, but the potential for application to large, more difficult to express proteins is clearly demonstrated.

Two-dimensional HSQC spectra show a cross-peak at the intersection of ^{15}N and ^{1}H chemical shifts for each amino acid backbone site (except proline). These cross-peaks can be used to extract amide exchange rates by monitoring loss of individual cross-peak intensities as a function of time after dissolving a protein in a deuterated buffer. Rates of exchange in typical proteins, at normal pH's, follow an EX2 mechanism in which base-catalyzed exchange of amide protons occurs occasionally from the open, solvent-exposed forms that exist in equilibrium with normally well-folded forms of backbone segments.[11] The fraction of open form varies widely with local stability, causing rates of amide proton exchange to vary by more than 6 orders of magnitude. This range provides a broad frequency range for additional resolution of HSQC peaks.

Measurement of rates of amide proton exchange at the rapid end of the range is normally limited by the length of time required to collect an HSQC spectrum. Recently, we introduced some methodology based on Hadamard transform (HT) encoded NMR spectroscopy[12,13] that reduces acquisition time to approximately 40 s for 0.5 mM samples of small proteins. Here we have extended our original application to human ubiquitin (1D3Z, 8547 Da) to provide a nearly complete assessment of amide proton exchange rates for this well-studied protein. These rates provide a basis for comparison of selective rates measured by MS methodology.

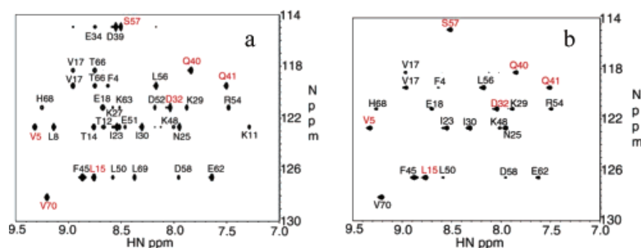

Figure 1. Reconstructed Hadamard [^{1}H,^{15}N]-HSQC spectra for ubiquitin. (a) Data in H_2O collected with 64 t_1 increments in 10 min. The sample was then lyophilized overnight and brought back to its initial volume with 99.9% D_2O at pH 6.0 and immediately returned to the spectrometer for rapid collection of a series of Hadamard spectra. The positions of excited ^{15}N frequencies are shown in red. (Residue I36 at 6.1 ppm, 123.4 ppm is not included in the spectra.) (b) First point after 1 min 22 s in D_2O collected with four scans in 42 s.

^{14}N unlabeled bovine ubiquitin was purchased from Sigma (St. Louis, MO) for the MS studies, and the ^{15}N labeled human ubiquitin was purchased from Cambridge Isotope Laboratories (Andover, MA) for the NMR studies. While the species for these two ubiquitins is different, the sequence is actually identical.

For the NMR studies a 0.5 mM human ubiquitin sample with 90% ^{15}N labeling was prepared in phosphate buffer at pH 6.0 and observed at 25 °C using an 800 MHz spectrometer. Initially the sample was prepared in H_2O, an ^{1}H−^{15}N HSQC reference spectrum was run, and the sample was lyophilized. At time zero the sample was dissolved in D_2O and transferred to the spectrometer for acquisition. Dissolution and transfer was done manually, limiting the first observation point to approximately 1 min. The spectra were acquired at geometrically increasing time points from 1 min to 48 h. As in our previous work,[14] seven ^{15}N frequencies were selectively excited. However, these frequencies were chosen to complement the original set, providing redundant data on at least two cross-peaks for the purpose of assessing reproducibility and providing many new pieces of information.

Figure 1a shows the HT NMR spectrum of ubiquitin in H_2O collected with 64 scans for each of the eight increments in the Hadamard encoding matrix. Figure 1b shows an equivalent spectrum taken 1.37 min after the addition of D_2O to the lyophilized protonated sample using four scans at each frequency. The latter spectrum required 42 s. Data processing, including the Hadamard transform, was achieved using nmrPipe.[15] Twelve out of the 35 peaks seen in Figure 1a have disappeared even at the first time point. However, reductions in intensities of the other peaks are easily quantified. Executing a third set and combining all data, the exchange rates of 68 out of 76 potential amide cross-peaks are obtained. Among the eight amino acids missing, G47 and G75 are not covered by ^{15}N frequency selection; M1, E24, and G53 are not present in the HSQC spectra, probably due to broadening from

10.1021/ja0457664 CCC: $27.50 © 2004 American Chemical Society

intermediate chemical exchange effects, and P19, P37, and P38 are prolines, which do not have amide protons. A complete table of exchange rates is supplied as Supporting Information.

MS measurements of H/D exchange proceeded in a similar fashion in that ubiquitin (nonlabeled bovine in this instance) was dissolved in deuterated buffer at time zero and analysis was conducted at roughly geometric intervals from 1 min to 44 h. However, in this case, small aliquots were extracted and subjected to a more complex analysis that included quenching of exchange by lowering pH and lowering temperature, digesting the protein with pepsin and analyzing fragment masses by matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) MS. To begin the experiment, 10 μL of 20 mg/mL ubiquitin in 50 mM phosphate buffer at pH 6.1 was dried in an Eppendorf tube, and then at time zero 100 μL of D_2O solution was added; the final pH was determined to be approximately 5.9. At each time point, 5 μL was taken out and put in the upper part of a 20 μL aerosol tip, which held 25 μL of a pepsin resin slurry from Pierce Chemicals (Rockford, IL) in 0.1% TFA (ubiquitin/pepsin = 1:3). The sample was immediately quenched by the addition of 45 uL of 0 °C 0.1% TFA, pH 2.0. After occasionally mixing for 40 s at room temperature, the sample was ejected though the filter of the tip into an Eppendorf tube. Each sample was quickly spotted on a chilled MALDI target, mixed with the matrix prepared as 5 mg/mL α-cyano-4-hydroxycinnamic acid in a solution containing 1:1:1 acetonitrile, ethanol, and 0.1% TFA (pH 2.0). The plate was immediately placed in a desiccator under a moderate vacuum such that the spots would dry in 1−2 min. Masses were then analyzed by MALDI TOF/TOF MS (Applied Biosystems 4700 Proteomics Analyzer). The H/D exchange experiments were repeated three times; in addition, samples at each time point were spotted and measured in triplicate. All results are averaged and reported along with the range of measurements as an estimate of error.

A reference mass spectrum of ubiquitin in H_2O was also analyzed using the same procedure to allow a more automated peptide fragment identification. Accurate mass measurements in combination with sequence analysis were used to identify the individual peptides from ubiquitin generated by pepsin cleavage. The online databases used for peptic peptide identification were MS-Digest http://prospector.ucsf.edu/ucsfhtml4.0/msdigest.htm and Peptide-Mass http://au.expasy.org/tools/peptide-mass.html. In addition, MS/MS ions produced by MALDI-TOF/TOF and the results of a Mascot search http://www.matrix-science.com/ were used to verify fragment identification. The 20 identified peptic peptides (500−2000 Da) marginally cover 100% of the entire ubiquitin sequence. The sequences covered represent buried and surface segments, as well as every type of secondary structure. However, it is clear that coverage could be improved by more complete digestion and an ability to monitor shorter peptides masked by matrix peaks. The latter is not a problem with electrospray ionization.

We focus here on five representative isotopic peptide peak clusters. They have calculated centroid masses of 1021.2, 1096.0, 1175.5, 1346.7, and 1390.4 Da in H_2O. The centroid masses of the peptide fragments incubated in D_2O begin near these reference values and shift with time as indicated in Figure 2 for the fragment of 1390.4 Da. The raw number of deuterons incorporated at each time point was determined by taking the difference between the centroid of the isotopic peak cluster for the deuterated sample and the centroid of the undeuterated control. These raw numbers must, however, be corrected for back-exchange that occurs during the quench, digestion, and MALDI steps in the procedure. These levels can be estimated from the proton content observed in long time points, but more accurate levels can be obtained by using a single scale factor as an adjustable parameter in fitting time courses predicted from NMR data. The percentage of back-exchange determined in this way ranged from 25 to 50%, after excluding that part from rapidly exchanging side-chain protons and correcting for the 10% deuterium in the final quench/matrix mixture. These levels of back-exchange compare favorably with levels reported in the literature.[7,16]

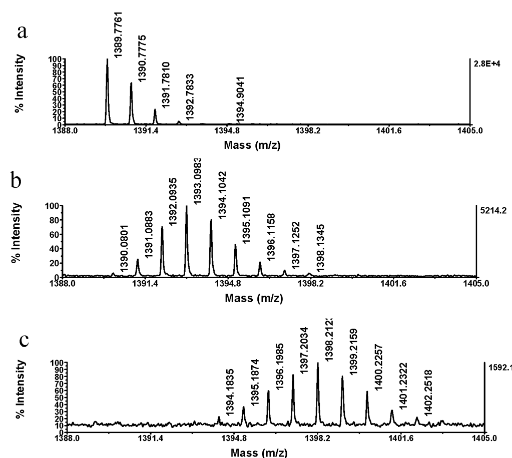

Figure 2. Time course of mass increase on deuterium incorporation. Lyophilized ubiquitin was dissolved in D_2O buffered with Na_2HPO_4 at pH 6.1 and incubated for varying lengths of time at room temperature before quenching and digesting the sample. The mass spectrum shows the region around the peptide of average mass 1390.4 Da (Res. #46−58: AGKQLEDGRTLSD). The undeuterated spectrum is shown in panel a as a reference. Panels b and c are for exchange times of 1 min and 4 h.

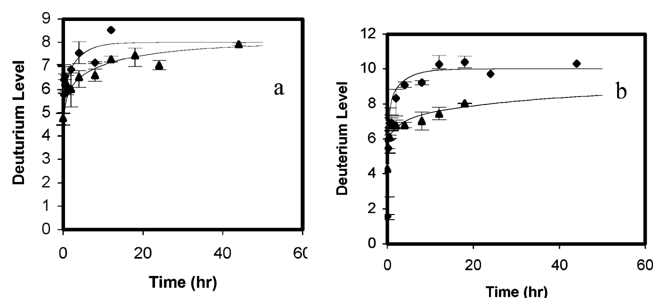

Figure 3. Correlation of deuterium incorporation from MS data (corrected for back-exchange) with levels predicted from NMR rate constants for two pairs of peptides from ubiquitin. (a) Compares data for the 1021 Da peptide (▲) and the 1096 Da peptide (◆). (b) Compares the 1347 Da peptide (◆) and the 1176 Da peptide (▲). The NMR model is calculated using the equation $D(t) = N - \sum_i^N \exp(-k_i t)$ at pH 5.85.

The time courses of exchange predicted from NMR data were calculated by summing the contributions expected for each amino acid given the NMR determined rate constants corrected for differences in pH. In practice, the deuteron contribution was set to 1 if the half-life of the amide H/D exchange was shorter than 1 min and to 0 if the half-life of the amide H/D exchange was longer than 1 week (>104 min). Otherwise, the exchange rate of each amino acid was used to calculate the contribution as $1 - e^{-kt}$, where k is the amide proton exchange rate and t is the time interval for exchange. Exchange rates where data are missing (4 of 72 possible measurements) are estimated from literature values or are taken to be averages of rates for preceding and succeeding residues.

Figure 3 shows predicted time courses for deuterium incorporation from NMR data (solid lines) and superimposed experimental points from the MS data. The peptides have been paired in two panels based on similarity of fragment masses. In Figure 3a, data are shown for peptides of mass 1021.2 Da (Res. #68−76: HLVLRLRGG) and mass 1096.0 Da (Res. #59−67: YNIQKESTL);

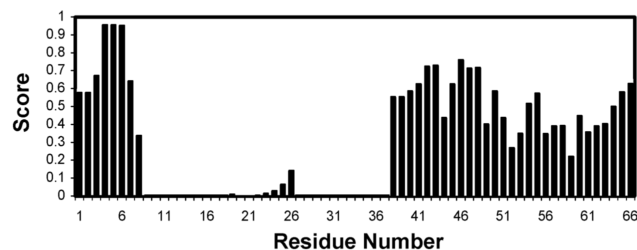

Figure 4. Sequential assignment scores for peptide 1175.5 of ubiquitin (Res. #5−15). The highest scores are seen near the proper placement position (4−6).

both have nine residues. The deuteron incorporation determined by MS in both cases agrees reasonably well with the NMR model. In both fragments, the number of deuterons plateaus at 8, the number of amide sites expected if the terminal NH_3^+ and side-chain amides are excluded. However, the different rates of mass increase correlate with the different environments these peptides see in natively folded ubiquitin. About half of the peptide of mass 1021.2 Da, is predicted to be a β strand, while the peptide of mass 1096.6 Da is predicted to have little regular secondary structure, hence its faster exchange.

In Figure 3b, data are shown for peptides of mass 1346.7 Da (Res. #59−69: YNIQKESTLHL) and mass 1175.5 Da (Res. #5−15: VKTLTGKTITL); both have 11 residues. Again, the agreement between MS experimental data and the NMR calculated model is good. Since the peptide of mass 1175.5 Da has a very slowly exchanging amino acid, L15 ($T_{1/2} > 1.91 \times 10^4$ min), the apparent plateau point for these two peptides is 10 and 9, respectively. The plot also shows different exchange rates for these two peptic fragments, again reflecting their different chemical environment and secondary structure in the native protein.

Using exchange information such as that presented in Figure 3 in an assignment strategy would require comparing MS exchange data with exchange predictions calculated from NMR data on all strings of connected HSQC cross-peaks of appropriate length. In Figure 4, we illustrate how a comparison can be done by using prior NMR assignments to predict exchange data for all 11 residue segments in the protein sequence and comparing those data to MS data for the 1175.5 Da peptide (Res. #5−15). The comparison is done using the equation $\exp(-\sum(D_{\rm expt} - D_{\rm calc})^2/(N^*\sigma^2))$, where N stands for the number of data points, σ is the average estimated error in measurement, and $D_{\rm expt}$ and $D_{\rm calc}$ are deuterium levels from experiment and calculation. The equation yields a score for each position. Sections that appear to lack scores either have very low scores or correspond to sequences that would be interrupted by prolines. The scores at positions 4−6 are highest, producing an assignment in agreement with expectation.

While in practice short segments of peptide can be connected in HSQC based experiments through nuclear Overhauser effects (NOEs), these connections are not always unambiguous and seldom run for more than a few residues. Clearly it would be better to have MS data on very short peptides or at least overlapping pairs of peptides so that differences in exchange rates could be associated with single sites or segments of two to three residues. We do have a pair of overlapping fragments in peptides of mass 1346.7 and 1096.0 Da. These differ in that the peptide of mass 1346.7 Da has two more residues, H and L, at the end. Differences in levels of deuterium as a function of time for these two peptides give the rates of deuterium incorporation for this terminal pair. The results show the presence of two rapidly exchanging amides with half-lives of approximately 30 min. Examining all of the NMR data on ubiquitin we find that approximately 14 out of a possible 72 pairs of connected HSQC cross-peaks can fit this pattern, including the correct pair, $H_{68}L_{69.}$

In summary, we have been able to demonstrate that correlation of amide exchange rate data from NMR and MS experiments can provide novel additional data to aid in assignment of cross-peaks in HSQC style spectra. The examples given here are limited in number and pertain only to a small protein. However, experiments can clearly be improved. Use of additional proteases can improve fragment coverage and provide smaller fragments.[8] In addition, new instrument fragmentation techniques such as electron capture dissociation and electron transfer dissociation might help to isolate deuterium incorporation sites. Also, NMR and MS data can be combined with procedures such as amino acid specific labeling[17] to reduce the number of HSQC peaks in spectra of larger proteins to the point where exchange rates can make useful distinctions. The most important point is that a strategy based on these ideas can be applied using high-resolution TROSY versions of HSQC spectra and can be applied to proteins labeled only with ^{15}N.

Acknowledgment. This work was supported in part by the Research Resource for Integrated Glycotechnology in the Complex Carbohydrate Research Center of the University of Georgia under Award Number RR05351 by the National Institutes of Health.

Supporting Information Available: The complete table of amide proton exchange rates of ubiquitin measured by Hadamard transform NMR spectroscopy. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Wüthrich, K. *Nat. Struct. Biol.* **2001**, *8*, 923−925.
(2) Galvao-Botton, L. M. P.; Katsuyama, A. M.; Guzzo, C. R.; Almeida, F. C. L.; Farah, C. S.; Valente, A. P. *FEBS Lett.* **2003**, *552*, 207−213.
(3) Andrec, M.; Hill, R. B.; Prestegard, J. H. *Protein Sci.* **1995**, *4*, 983−993.
(4) Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 75−86.
(5) Huyghues-Despointes, B. M. P.; Pace, C. N.; Englander, S. W.; Scholtz, J. M. *Methods Mol. Biol.* **2001**, *168*, 69−92.
(6) Dempsey, C. E. *Prog. Nucl. Magn. Reson. Spectrosc.* **2001**, *39*, 135−170.
(7) Mandell, J. G.; Falick, A. M.; Komives, E. A. *Anal. Chem.* **1998**, *70*, 3987−3995.
(8) Englander J. J.; Mar, C. D.; Li, W.; Englander, W. S.; Kim, J. S.; Stranz, D. D.; Hamuro, Y.; Woods, V. L., Jr. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7057−7062.
(9) Zhang, Z.; Smith, D. L. *Protein Sci.* **1993**, *2*, 522−531.
(10) Zhang, Z.; Li, W.; Logan, T. M.; Li, M.; Marshall, A. G. *Protein Sci.* **1997**, *6*, 2203−2217.
(11) Engen, J. R.; Smith, D. L. *Anal. Chem.* **2001**, 257A−265A.
(12) Kupce, E.; Freeman, R. *J. Magn. Reson.* **2003**, *163*, 56−63.
(13) Kupce, E.; Freeman, R. *J. Biomol. NMR* **2003**, *25*, 349−354.
(14) Bougault, C.; Feng, L. M.; Glushka, J.; Kupce, E.; Prestegard, J. H. *J. Biomol. NMR* **2004**, *28*, 385−390.
(15) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277−293.
(16) Ghaemmaghami, S.; Fitzgerald M. C.; Oas, T. G. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8296−8301.
(17) Weigelt, J.; Wikstrom, M.; Schultz, J.; Van D., Maria J. P. *Comb. Chem. High Throughput Screening* **2002**, *5*, 623−630.

JA0457664