Mass Spectrometry Assisted Assignment of NMR Resonances in Reductively ^{13}C-Methylated Proteins

Megan A. Macnaughtan, **Austin M. Kane**[†], and **James H. Prestegard**[*]
Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602

Nuclear magnetic resonance (NMR) spectroscopy is a powerful tool for studying protein structure and dynamics, but in most cases, it relies on extensive incorporation of stable magnetic isotopes to aid in resolution and assignment. Typically, overexpression in a bacterial host is used to uniformly incorporate ^{13}C, ^{15}N, and ^{2}H into a protein. For eukaryotic proteins that are less amenable to bacterial host expression (membrane proteins or glycosylated proteins), labeling through over-expression in eukaryotic hosts can be difficult and expensive. An alternate approach is to isotopically label proteins post-expression via chemical methods. Chemical modification of proteins is an established field of science with highly selective reactions for certain amino acids.[1] In many cases the products also mimic those produced by natural post-translational modifications. One protein modification that has already been used successfully with NMR and x-ray crystallography is reductive methylation of lysines. In principle, for NMR, the signals from ^{13}C methyls introduced in this way can be sensitive probes of protein-protein interactions and easily observable sites for collection of distance restraints from paramagnetic perturbations.[2] However, assignment of resonances is a prerequisite for these applications and this has always been a problem. In this communication we offer a solution to that problem.

Reductive methylation of proteins was introduced by Means and Feeney in 1968.[3] Under mild reaction conditions, the lysine ε-NH$_2$ and N-terminal α-NH$_2$ can be selectively methylated using ^{13}C labeled formaldehyde and a reducing agent such as dimethylamine borane to produce the dimethylated amine. Since its introduction it has been used a number of times to introduce probes of both dynamic and structural properties.[2,4,5] Assignment approaches have varied, but for the most part these have relied on small numbers of sites, known structural properties of proteins, or extensive genetic modification. For example, Zhang and Vogel studied reductively ^{13}C-methylated calmodulin and demonstrated the complete assignment of the dimethylamine signals through site-directed mutagenesis.[6] In 2000, Fisher *et al.* used MALDI-TOF to identify partially methylated lysines of trypsin-digested human MIP-1α by correlating disappearance of unmodified peptide masses in MS with the appearance of mono- and dimethylamine signals observed in a 2D ^{1}H-^{13}C NMR spectrum.[7] Here we present a general method for assigning dimethyllysine and dimethylamine resonances in 2D ^{1}H-^{13}C NMR spectra without prior knowledge of the protein's structure or properties other than the amino acid sequence. This method builds on the previous use of mass spectrometry, but introduces isotope ratio measurements that make quantitation and more extensive assignments possible.

Reductive methylation of hen egg white lysozyme with ^{13}C-formaldehyde was carried out following the procedure described by Rayment.[8] Dimethylamine borane complex was used as the reducing agent and various amounts of ^{13}C-formaldehyde were used to reductively methylate aliquots of lysozyme at pH 7.5 in phosphate buffer at 4 °C. Four partially ^{13}C-labeled samples were prepared with molar ratios of ^{13}C-formaldehyde to primary amine of 0.5:1, 1:1,

[†]Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556
E-mail: jpresteg@ccrc.uga.edu

2:1, and 5:1. These samples were subsequently reacted with excess natural abundance formaldehyde to complete the reductive methylation and produce homogeneously modified protein samples. An additional sample was prepared using excess ^{13}C-formaldehyde and was used as a normalization reference for the NMR and MS data.

2D [1]H-^{13}C heteronuclear single quantum correlation spectroscopy (HSQC) NMR and matrix assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry (MS) were used to analyze all reductively methylated lysozyme samples. An NMR spectrum of the fully ^{13}C-labeled lysozyme sample is shown in Figure 1. The 7 expected dimethylamine resonances are observed with high sensitivity and good resolution in the 41-43 ppm ^{13}C chemical shift range. As expected, the monomethylamine region around 32 ppm (not shown) did not show any resonances. Complete reductive methylation of the protein is essential for both quality NMR spectra and quantitative analysis of spectra. Incomplete methylation gives both dimethylamine and monomethylamine resonances. These are resolved in themselves; however, effects of differential methylation of proximate sites perturbs chemical shifts leading to poor quality spectra.[5,9]

Assignment of the dimethylamine NMR peaks was accomplished by correlating the amount of ^{13}C incorporation into each dimethylamine group for the four partially ^{13}C-labeled samples. In the NMR spectra, each peak was integrated and normalized to the reference ^{13}C-dimethylamine lysozyme sample. Table 1 lists the ^{13}C incorporation percentage for each peak and each sample. As expected, the amount of ^{13}C incorporated at each site (lysine or N-terminal amine) increased with increasing moles of ^{13}C-formaldehyde used in the reaction mixture. More importantly, for most ratios, the extent of ^{13}C incorporation at each site differed indicating that the individual chemical environments play a role in the reaction and the dimethylamine groups can be distinguished on this basis.

MALDI-TOF MS using a α-cyano-4-hydroxycinnamic acid/acetonitrile matrix was then used to measure levels of ^{13}C incorporation in specifically identified peptide sequences. A small aliquot from each sample was proteolyzed with trypsin (trypsin does not cleave efficiently at dimethyllysine, so primarily arginine sites were cleaved). Figure 2 shows the MALDI-TOF data from the peptide containing lysine 13 for every sample. The average mass in the profile shifts by 2 mass units as the amount of ^{13}C-formaldehyde used is increased; this limit corresponds to the final addition of 2 [13]C's in the methyl groups. For each peptide containing a dimethylamine group, the average mass was calculated at each reaction ratio. The reference 13C-dimethylamine lysozyme sample was used to correct for natural abundance contributions at non-methyl sites and to normalize the data to give the percentages of ^{13}C incorporation listed in Table 2.

Analysis of differential levels of isotope incorporation is superior to simply monitoring the disappearance of unmodified or monomethylated peptides. The intensities of signals from different chemical species in MS are not directly related to concentration, but related to ionization efficiencies and other factors, and the intensities of peaks from nonmethylated lysine, monomethylated lysine, and dimethylated lysine, could not easily be used to calculate the percent of modification. Completing the reductive methylation on partially ^{13}C-labeled samples with natural abundance formaldehyde also produces homogeneous protein samples that facilitate analysis of both NMR and MS data.

Assignment of the NMR peaks can be made by comparing the NMR and MS percentages of ^{13}C incorporation in Tables 1 and 2. While all reaction ratios can contribute to assignment, we illustrate this with data from the 5:1 reaction. The four peptides containing single lysines, including K13, K33, K97, and K116, are simplest to analyze. The fastest methylating lysine is K97 (81.8% [13]C incorporation) and can be assigned to NMR peak 7 with 78.3%

incorporation. Peaks 1 and 3 have very similar incorporation values of 71.4% and 70.3%, respectively, and either can be assigned to K13 (71.5%). Likewise, the values for peaks 2 and 5 are comparable (61.1% and 60.0%) and can be assigned to either K33 (64.5%) or K116 (56.9%).

MS data for K96 occurs only in combination with K97. Since trypsin is not efficient at cleaving at dimethyllysine, these lysines remain together in the dominant tryptic peptide and their MS data are an average of the two sites. As it turns out, a small percentage of unmethylated and mono-methylated sites do exist in the protein and a small amount of cleaved peptide is actually present giving data separately for K97. The value for K96 can, therefore, be back-calculated using MS data for K97 and the K96-K97 pair. K96 turns out to be the slowest methylating site with 23.7% ^{13}C incorporation, and can be assigned to peak 6 (30.2%). By process of elimination, the N-terminal amine and K1 must be assigned to peak 4 and one of the peaks 1 or 3. The N-terminal dimethylamine peak can be distinguished from the other peaks by the presence of a strong NOE to the Hα of the N-terminal amino acid (data not shown). On the basis of the appearance of such an NOE, peak 4 can be assigned to the N-terminal dimethylamine and peaks 1 or 3 to K1. Hence, we have been able to make definitive assignments for three of the seven peaks and reduce ambiguities to a pairwise assignment for the remaining four. It is possible that running the methylation reaction at a different pH could remove this final ambiguity.

ACKNOWLEDGMENT

References

(1). Lundblad, RL. *Chemical reagents for protein modification*;. CRC Press; Boca Raton: 1991.

(2). (a) Moore GR, Cox MC, Crowe D, Osborne MJ, Rosell FI, Bujons J, Barker PD, Mauk MR, Mauk AG. Biochem. J 1998;332:439–449. [PubMed: 9601073] (b) Dick LR, Geraldes CFGC, Sherry AD, Gray CW, Gray DM. Biochem 1989;28:7896–7904. [PubMed: 2514796]

(3). Means GE, Feeney RE. Biochem 1968;7:2192–2201. [PubMed: 5690712]

(4). (a) Brown LR, Bradbury JH. Eur. J. Biochem 1975;54:219–227. [PubMed: 238843] (b) Jentoft JE, Gerken TA, Jentoft N, Dearborn DG. J. Biol. Chem 1981;256:231–236. [PubMed: 6256347] (c) Goux WJ, Teherani J, Sherry AD. Biophys. Chem 1984;19:363–373. [PubMed: 6430360] (d) Dick LR, Sherry AD, Newkirk MM, Gray DM. J. Biol. Chem 1988;263:18864–188872. [PubMed: 3143718] (e) Jentoft JE. Methods in Enzymol 1991;203:261–274. [PubMed: 1762560] (f) Sparks DL, Phillips MC, Lund-Katz S. J. Biol. Chem 1992;267:25830–25838. [PubMed: 1464597]

(5). Gerken TA, Jentoft JE, Jentoft N, Dearborn DG. J. Biol. Chem 1982;257:2894–2900. [PubMed: 7061454]

(6). Zhang M, Vogel HJ. J. Biol. Chem 1993;268:22420–22428. [PubMed: 8226750]

(7). Ashfield JT, Meyers T, Lowne D, Varley PG, Arnold JRP, Tan P, Yang J-C, Czaplewski LG, Dudgeon T, Fisher J. Protein Science 2000;9:2047–2053. [PubMed: 11106181]

(8). Rayment I. Methods in Enzymol 1997;276:171–179. [PubMed: 9048376]

(9). Chemical shift changes with the extent of methylation were observed for peaks 6 and 7 for partially reductively ^{13}C-methylated lysozyme samples not described in this paper
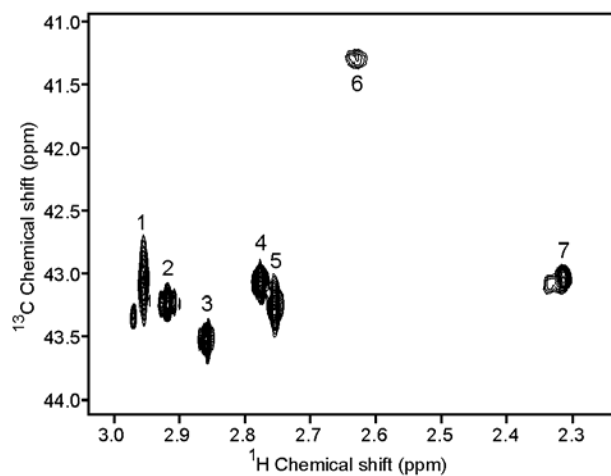
Figure 1.
^1H-^{13}C HSQC of ^{13}C-dimethylated lysozyme (120 μM) at pH 8.1 acquired on a 600 MHz spectrometer in approximately 5 minutes.
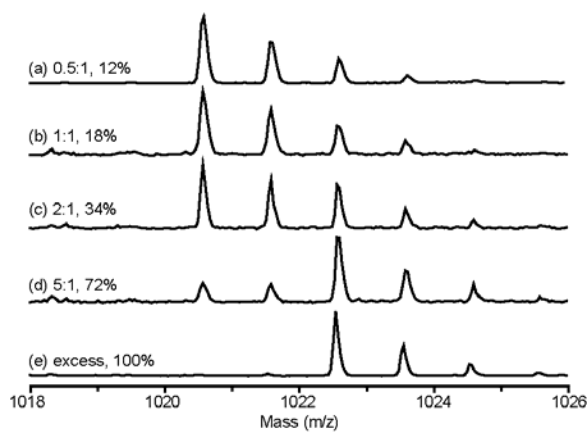
Figure 2.
MS isotope profiles for the tryptic peptide containing a single dimethyllysine (Res. #6-14 CELAAAMKR) at various reaction stoichiometries. The reaction conditions in moles of ^{13}C formaldehyde relative to the moles of reactive amines are listed along with the normalized percentages of ^{13}C incorporation calculated from the profiles.

Table 1

Normalized peak volumes from ^{1}H-^{13}C HSQC spectra of the four partially ^{13}C-labeled samples.

NMR peak	0.5:1 sample	1:1 sample	2:1 sample	5:1 sample
1	9.9	18.0	33.7	71.4
2	8.8	12.8	25.3	61.1
3	11.6	17.3	32.0	70.3
4	5.2	7.5	15.4	38.6
5	8.8	11.9	26.1	60.0
6	---	---	10.7	30.2
7	19.9	21.5	34.7	78.3

Table 2

The normalized percentages of ^{13}C incorporation for each peptide containing either one or two dimethylamines.

Lysine/amine in peptide	0.5:1 sample	1:1 sample	2:1 sample	5:1 sample
N-terminal amine, K1	15.4	14.1	17.9	44.8
K13	12.1	17.7	33.6	71.5
K33	10.5	11.5	27.6	64.5
K96, K97	---	13.2	23.0	52.1
K97	15.5	19.8	36.0	81.8
K116	5.5	11.0	24.2	56.9
K96 (back-calculated)	---	4.7	11.2	23.7