Protein Molecular Mass to 1 Da by ^{13}C, ^{15}N Double-Depletion and FT-ICR Mass Spectrometry

Alan G. Marshall,*,† Michael W. Senko, Weiqun Li,† Ming Li,† Stephanie Dillon,† Shenheng Guan,† and Timothy M. Logan†

Center for Interdisciplinary Magnetic Resonance
National High Magnetic Field Laboratory
1800 E. Paul Dirac Drive
Florida State University, Tallahassee, Florida 32310

Received August 26, 1996

Accurate determination of protein molecular mass to within 1 Da would be a boon to protein characterization. It would then become possible to (a) count the number of disulfide bridges (−S−S− is 2 Da lighter than 2 −SH); (b) identify deamidation (−NH_2 is 1 Da lighter than −OH); (c) identify such post-translational modifications as phosphorylation and glycosylation; (d) resolve and identify adducts; (e) identify variant amino acid sequences; etc. Determination of the molecular mass of a neutral protein to within 1 Da from measurement of the mass of its gas-phase ion might appear easy. After all, electrospray ionization can now routinely generate abundant multiply-charged gas-phase unhydrated quasimolecular ions, $(M + nH)^{n+}$, for most proteins,[1,2] and Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry[3−7] can determine the ion mass to parts-per-million accuracy at typical electrosprayed protein multiply-charged ion mass-to-charge ratios, $500 \leq m/z \leq 2000$.[8] However, the monoisotopic mass (see below) of a protein inferred from the mass(es) of its corresponding ions may still be wrong by up to several Dalton!

There are three stages in the determination of the molecular weight of a neutral protein from its electrosprayed ion mass spectrum.[9] (Our electrospray FT-ICR mass spectra were obtained with a homebuilt instrument operating at 9.4 T, as described elsewhere.[10]) First, electrospray ionization produces protein ions with various numbers of attached protons and thus several charge states. The first stage in protein mass analysis is therefore to separate the individual charge states (e.g., $(M + zH)^{z+}$, $(M + (z+1)H)^{(z+1)+}$, etc.). Second, since mass spectrometry reports mass-to-charge *ratio*, it is necessary to determine the ion charge in order to determine its mass. The mass-to-charge ratio spectrum of a protein of a given charge state exhibits numerous "isotopic" peaks (see below) spaced ∼1 Da apart (Figure 1). High-resolution FT-ICR mass spectrometry can resolve those peaks for proteins of molecular mass up to more than 100 000 Da, so that the charge state, *z*, may be determined simply as the reciprocal of the separation between two adjacent isotopic peaks differing in mass by ∼1 Da.[11] However, protein mass measurement accuracy is presently limited by the third stage of mass analysis: namely, knowledge of the *isotopic composition* (i.e., the constituent chemical formula(s) composing each mass spectral peak). For organic molecules of less than ∼1000 Da, determination of molecular weight from the singly-charged molecular (M^+) or quasi-molecular (e.g., $(M + H)^+$) ion is relatively simple. Why then is it so much more difficult to determine the mass of a biological macromolecule? The problem is apparent from Figure 1 (top). The natural abundance of ^{13}C is 1.066−1.106% relative to ^{12}C as 100%.[8] However, for a molecule containing *n* carbons, the isotopic distribution is a binomial expansion $(0.9889 + 0.0111)^n$, and it is ∼*n*% as likely that a given molecule will contain one ^{13}C as that all of the carbons will be ^{12}C. For a protein with hundreds of carbons, hydrogens, nitrogens, and oxygens, the combined binomial distributions for its constituent elements produce a wide spread in natural-abundance isotopic relative abundances (Figure 1, top). Only one resolved peak has a unique isotopic composition, namely the "monoisotopic" species, $^{12}C_{527}{}^{1}H_{830}{}^{14}N_{146}{}^{16}O_{155}{}^{32}S_3{}^{+}$, in this case. Every other peak represents an unresolved superposition of several isotopic variants, e.g., the next highest nominal mass peak includes $^{13}C^{12}C_{526}{}^{1}H_{830}{}^{14}N_{146}{}^{16}O_{155}{}^{32}S_3$ and $^{12}C_{527}{}^{1}H_{830}{}^{15}N^{14}N_{145}{}^{16}O_{155}{}^{32}S_3$ (as well as other combinations). Thus, accurate unambiguous determination of protein molecular mass to within 1 Da reduces to correct identification of the monoisotopic mass.[12] For very small proteins (≤5−10 kDa), the "monoisotopic" species, $^{12}C_v{}^{1}H_w{}^{14}N_x{}^{16}O_y{}^{32}S_z$ may be detected directly. However, for proteins larger than ∼ 2 kDa, the most abundant peak is shifted upward in mass from the monoisotopic peak by ∼1 Da for every 1.5 kDa of molecular mass; at ≥ 15 kDa the "monoisotopic" ion abundance is below detectability (≪1%).[8,13] Moreover, variation in the natural relative abundance of $^{13}C^{14}$ can shift

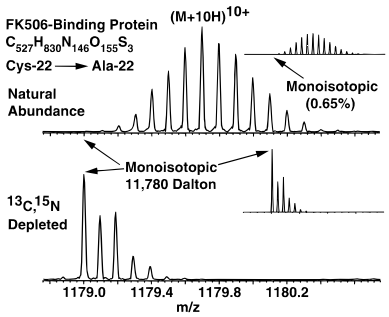

Figure 1. Electrospray ionization FT-ICR mass spectra (9.4 T) of a mutant (C22A) FK506-binding protein. (top) Natural-abundance isotopic distribution (∼98.89% ^{12}C; ∼99.63% ^{14}N. (bottom) Isotopic distribution for the same protein grown on a medium with 99.95% ^{12}C and 99.99% ^{14}N. Insets: Isotopic distributions calculated (same vertical scale) from the chemical formula for natural-abundance (top) and ^{13}C, ^{15}N doubly-depleted (bottom) FK506 binding protein. For an electrospray-ionized small protein, the "monoisotopic" peak (all carbons are ^{12}C; all hydrogens are ^{1}H; all nitrogens are ^{14}N; all oxygens are ^{16}O, and all sufurs are ^{32}S) is barely observable, and each other resolved "isobaric" peak consists of multiple isotopic combinations. The charge state, *z*, is determined as the reciprocal of the spacing between adjacent mass-to-charge ratio species differing in mass by 1 Da. The accuracy of determination of the neutral protein molecular weight depends on accurate matching of the relative isotopic abundances of the computed and experimental mass spectra, not simply on the mass accuracy of the individual mass spectral peaks (see text).

* To whom correspondence should be addressed.
† Department of Chemistry, Florida State University.

(1) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F. *Mass Spectrom. Rev.* **1990**, *9*, 37−70.
(2) Smith, R. D.; Loo, J. A.; Ogorzalek Loo, R. R.; Busman, M.; Udseth, H. R. *Mass Spectrom. Rev.* **1992**, *10*, 359−451.
(3) Marshall, A. G. *Acc. Chem. Res.* **1985**, *18*, 316−322.
(4) Buchanan, M. V.; Hettich, R. L. *Anal. Chem.* **1993**, *65*, 245A−259A.
(5) Wilkins, C. L. *Fourier Transform Mass Spectrometry*; Trends in Analytical Chemistry 13; Wilkins, C. L., Ed.; Elsevier: Amsterdam, 1994; pp 223−251.
(6) Holliman, C. L.; Rempel, D. L.; Gross, M. L. *Mass Spectrom. Rev.* **1994**, *13*, 105−132.
(7) Marshall, A. G. *Acc. Chem. Res.* **1996**, *9*, 307−316.
(8) McLafferty, F. W. *Acc. Chem. Res.* **1994**, *27*, 379−386.
(9) Loo, J. A.; Edmonds, C. G.; Smith, R. D. *Science* **1990**, *248*, 201−204.
(10) Senko, M. W.; Hendrickson, C. L.; Pasa-Tolic, L.; Marto, J. A.; White, F. M.; Guan, S.; Marshall, A. G. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1824−1828.
(11) Henry, K. D.; McLafferty, F. W. *Org. Mass Spectrom.* **1990**, *25*, 490−492.
(12) Yergey, J.; Heller, D.; Hansen, G.; Cotter, R. J.; Fenselau, C. *Anal. Chem.* **1983**, *55*, 353−356.
(13) Winger, B. E.; Hofstadler, S. A.; Bruce, J. E.; Udseth, H. R.; Smith, R. D. *J. Amer. Soc. Mass Spectrom.* **1993**, *4*, 566−577.
(14) Pomerantz, S. C.; McCloskey, J. A. *J. Org. Mass Spectrom.* **1987**, *22*, 251−253.

S0002-7863(96)03004-1 CCC: $14.00

the most abundant isotopic peak of carbonic anhydrase from 29 024.73 to 29 025.73 Da.[8]

In the absence of the monoisotopic peak, the isotopic composition of a protein may be inferred by either of two indirect means. First, for a known (or assumed average[15]) protein amino acid composition, adduction (if any), and derivatization (if any), the natural-abundance isotopic distribution can be calculated[16] and aligned with the observed distribution until the relative abundances match;[15] it is not even necessary to resolve the individual isobars. However, there must be enough ions (at least several thousand) to produce a statistically reliable experimental distribution. A second method is to break apart ions of a given isotopic composition, and then analyze the relative abundances of fragment ions of lower mass and lower charge state;[17] however, the resulting lower signal-to-noise ratio can more than compensate for the simpler isotopic distributions of the fragment ions.

Fortunately, because of the widespread production of proteins (in milligram quantity) isotopically *enriched* in ^{13}C and ^{15}N for multidimensional heteronuclear FT-NMR experiments, it is easy to produce any of the same proteins isotopically *depleted* in ^{13}C and ^{15}N by substitution of appropriate ^{13}C- and ^{15}N-depleted nutrients. Here, as a first example, we show that the mass shift and width of the natural-abundance isotopic distribution for FK506-binding protein (FKBP, ~11 800 Da) may be reduced substantially by isolating the protein from *E. coli* grown on 99.95% glucose-$^{12}C_6$ and 99.99% ammonium sulfate-$^{14}N_2$.[18,19]

FKBP is a small (107 amino acids) protein that exhibits peptidyl-prolyl cis−trans isomerase activity. The 3D structure of FKBP, both free[20] and complexed to ascomycin or rapamycin,[21−24] is known to high precision, and the equilibrium unfolding behavior in urea has been thoroughly characterized,[25] making it an attractive model for protein folding research. The structural properties of FKBP unfolded in concentrated urea and guanidine hydrochloride solutions are known,[26] as are kinetics of refolding from concentrated denaturant solutions.[27] We are interested in characterizing the structural properties of the unfolded state to determine its role in the folding and stability of FKBP, by H/D amide exchange to characterize the presence and structure of kinetic folding intermediates.

Figure 1 (bottom) shows the dramatic improvement in mass spectral quality afforded by double isotopic depletion of FKBP. The monoisotopic species, present at only 0.65% at natural abundance, becomes the "base" (largest) peak in the mass spectrum of the ^{13}C,^{15}N doubly-depleted (99.95% ^{12}C, 99.99% ^{14}N) protein! The molecular weight of the neutral protein is thus determined immediately and unambiguously as 11 780.01 (vs 11 780.07 computed from the amino acid sequence).

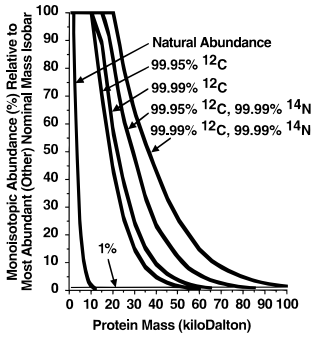

Figure 2. Monoisotopic relative abundance for a protein of average amino acid composition as a function of protein mass, for various possible levels of isotopic depletion of ^{13}C and ^{15}N.

Moreover, for a similarly doubly-depleted protein of 60 kDa molecular mass, the monoisotopic peak would still be 5% abundant, and thus easily identifiable. Figure 2 displays the monoisotopic abundance (relative to the largest other nominal-mass isobar) as a function of mass, for a protein of average[15] amino acid sequence. For example, the upper mass limit for 1% monoisotopic relative abundance increases from ~10 kDa (naturally abundant ^{12}C and ^{14}N) to ~53 kDa (99.95% ^{12}C), 81 kDa (99.95% ^{12}C and 99.99% ^{14}N), and ~100 kDa (99.99% ^{12}C and 99.99% ^{14}N). Depletion of rare isotopes increases mass spectral signal-to-noise ratio (because the same number of ions now exhibit fewer isotopic variants). Thus, mass spectral sensitivity and detection limit improve accordingly. Distortions (especially skewing of the shape of the detected isotopic distribution) resulting from space charge are reduced because a mass spectrum of a given peak-height-to-noise ratio requires fewer total number of ions. Tandem mass spectrometry or MS/MS experiments are improved because the narrow mass-to-charge ratio distribution makes it easier to isolate the desired parent ions and facilitates identification of fragments [e.g., 1 Da difference between loss of H_2O *vs* NH_3 or glutamic acid (129 Da) *vs* glutamine (128 Da)]. Depletion narrows all isotopic distributions, including any adducts (e.g., $(M + nH + mNa)^{(n+m)+}$) and thus dramatically increases the upper molecular weight limit before mass assignment is affected due to isotopic overlap of such impurities. The same advantage applies to detection of deamidation (1 Da difference between $-NH_2$ *vs* $-OH$), e.g., a recent attempt to identify deamidation for a 43 kDa protein was made difficult by 2 Da mass accuracy.[28] A narrower protein isotopic distribution makes it easier to observe and characterize non-covalent binding (protein:protein, protein:nucleic acid, enzyme:inhibitor, etc.). Identification of surface-accessible residues by H/D exchange is simpler because of simpler deconvolution to yield the deuterium number distribution. Isotopically-depeleted proteins provide a good mass calibrant, whose own isotopic distribution is narrower than other natural-abundance proteins of similar molecular weight.

The double-depletion method is general and can be applied to dozens of proteins (and RNA's and DNA's) for which ^{13}C and ^{15}N *enrichment* is already available for FT-NMR applications and to hundreds more for which DNA has been cloned. Although demonstrated here with Fourier transform mass spectrometry, our method offers proportionate benefits to other types of mass analyzers as well. In the same way that ^{13}C and ^{15}N *enrichment* simplifies and extends the applicability of NMR, ^{13}C and ^{15}N *depletion* simplifies and extends the applicability of mass spectrometry for biological macromolecules.

Acknowledgment. This work was supported by the NSF (CHE-94-13008), NIH (GM-31683), Florida State University, and National High Magnetic Field Lab in Tallahassee, FL.

JA9630046

(15) Senko, M. W.; Beu, S. C.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 229−233.
(16) Yergey, J. *Int. J. Mass Spectrom. Ion Processes* **1983**, *52*, 337−349.
(17) O'Connor, P. B.; Little, D. P.; McLafferty, F. W. *Anal. Chem.* **1996**, *68*, 542−545.
(18) Barettino, D.; Feigenbutz, M.; Valcarcel, R.; Stunnenberg, H. G. *Nucl. Acids Res.* **1994**, *22*, 541−542.
(19) Holzman, T. L.; Egan, D. A.; Edalji, R.; Simmer, R. L.; Helfrich, R.; Taylor, A.; Burres, N. S. *J. Biol. Chem.* **1991**, *226*, 2474−2479.
(20) Michnik, S.; Rosen, M. K.; Wandless, T. J.; Karplus, M.; Schreiber, S. L. *Science* **1991**, *252*, 836.
(21) van Duyne, G. D.; Standaert, R. F.; Karplus, P. A.; Schreiber, S. L.; Clardy, J. *Science* **1991**, *252*, 839−842.
(22) van Duyne, G. D.; Standaert, R. F.; Schreiber, S. L.; Clardy, J. *J. Am. Chem. Soc.* **1991**, *113*, 7433−7434.
(23) van Duyne, G. D.; Standaert, R. F.; Karplus, P. A.; Schreiber, S. L.; Clardy, J. *J. Mol. Biol.* **1993**, *229*, 105−124.
(24) Meadows, R. P.; Nettesheim, D. G.; Xu, R. X.; Olejniczak, E. T.; Petros, A. M.; Holzman, T. F.; Severin, J.; Gubbins, E.; Smith, H.; Fesik, S. W. *Biochemistry* **1993**, *32*, 754−765.
(25) Egan, D. A.; Logan, T. M.; Liang, H.; Matayoshi, E.; Fesik, S. W.; Holzman, T. F. *Biochemistry* **1993**, *32*, 1920−1927.
(26) Logan, T. M.; Theriault, Y.; Fesik, S. W. *J. Mol. Biol.* **1994**, *236*, 637−648.
(27) Veeraraghavan, S.; Holzman, T. F.; Nall, B. T. *Biochemistry* **1996**, *35*, 10601−10607.
(28) Wood, T. D.; H., C. L.; Kelleher, N. L.; Little, D. P.; Kenyon, G. L.; McLafferty, F. W. *Biochemistry* **1995**, *34*, 16251−16254.