Top Down versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry

Neil L. Kelleher, Hong Y. Lin, Gary A. Valaskovic, David J. Aaserud, Einar K. Fridriksson, and Fred W. McLafferty*

Contribution from the Department of Chemistry and Chemical Biology, Baker Laboratory, Cornell University, Ithaca, New York 14853-1301

Received October 21, 1997

Abstract: Characterization of larger proteins by mass spectrometry (MS) is especially promising because the information complements that of classical techniques and can be obtained on as little as 10^{-17} mol of protein. Using MS to localize errors in the DNA-derived sequence or modifications (posttranslational, derivatized active sites, etc.) usually involves extensive proteolysis to yield peptides of <3 kDa, with separation and MS/MS to compare their sequences to those expected (the "bottom up" approach). In contrast, an alternative "top down" approach limits the dissociation (proteolysis or MS/MS) to yield larger products from which a small set of complementary peptides can be found whose masses sum to those of the molecule. Thus a disagreement with the predicted molecular mass can be localized to a fragment(s) without examining all others, with further dissociation of the fragments in the same way providing further localization. Using carbonic anhydrase (29 kDa) as an example, Fourier transform mass spectrometry is unusually effective for the bottom up approach, in that a single spectrum of an extensive chymotryptic digest identifies 64 expected peptides, but these only cover 95% of the sequence; 20 fragment masses are unassigned so that any set whose masses sum to that of the molecule would be misleading. Extensive Lys-C dissociation yields 17 peptides, 23 unassigned masses, and 96% coverage. In the contrasting "top down" approach, less extensive initial dissociation by Lys-C, MS/MS, or CNBr in each case provides 100% coverage, so that modified protein fragment(s) could easily be recognized among the complementary sets. MS/MS of such a fragment or more extensive proteolysis provide further localization of the modification. The combined methods cleaved 137 of the 258 amide bonds between residues.

Introduction

The recent development of matrix-assisted laser desorption ionization[1] and electrospray ionization (ESI)[2] has revolutionized the applicability of tandem mass spectrometry (MS/MS) to biological and medical problems.[3] Combinations of the characteristic masses of the common components of linear biomolecules (20 amino acids for proteins, 4 bases for DNA or RNA) provide direct sequence information from as little as 10^{-17} mol of protein.[4] For newly isolated proteins, such MS data are so characteristic that they can be identified by matching a few fragment masses against a large database of DNA-derived sequences of previously identified proteins.[5] More challenging is the interpretation of MS fragment data for identifying errors in DNA-derived sequences[6] and for locating posttranslational and other covalent modifications,[7] such as specific derivatization at the active site of an enzyme. Basically, one or more errors or modifications are indicated if the relative molecular weight (M_r) differs from that predicted from the protein's DNA- derived sequence. Most MS studies have used extensive proteolysis to prepare small peptides (<3 kDa) whose M_r values by MS are matched against those expected from the DNA-predicted sequence. Matches indicate an unmodified region of the protein, but mass accuracy and redundancy limitations can require identity verification, such as by MS/MS. However, localization of the modification to a specific fragment by its difference in mass from that predicted is more difficult, especially if the fragments do not provide full sequence coverage, if more than one is modified, or if many fragment masses cannot be assigned. Confident identification of the modified peptide(s) requires dissociation of its ionized molecular species (MS/MS) to provide verification of its predicted sequence position and localize further the error or modification.

[1] Karas, M.; Hillenkamp, F. *Anal. Chem.* **1988**, *60*, 2301−2303.

[2] Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Mass Spectrom. Rev.* **1990**, 9, 37−70.

[3] Biemann, K.; Papayannopoulos, I. A. *Acc. Chem. Res.* **1994**, *27*, 370−378. Loo, J. A. *Bioconj. Chem.* **1995**, *6*, 644−665. Shackleton, C. H. L.; Witkowska, H. E. *Anal. Chem.* **1996**, 68, 29A-32A. Hao, Y.; Muir, T. W.; Kent, S. B. H.; Tische, E.; Scardina, J. M.; Chait, B. T. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 4020−4024. Hunt, D. F.; Henderson, R. A.; Shabanowitz, J.; Sakaguchi, K.; Michel, H.; Sevilir, N.; Cox, A. L.; Appella, E.; Engelhard, V. H. *Science* **1992**, *255*, 1261−1266.

[4] Valaskovic, G. A.; Kelleher, N. L.; McLafferty, F. W.*Science* **1996**, *273*, 1199−1202.

[5] (a) Shevchenko, A.; Jensen, O. N.; Podtelejnikov, A. V.; Sagliocco, F.; Wilm, M.; Vorm, O.; Mortensen, P.; Shevchenko, A.; Boucherie, H.; Mann, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 14440−14445. (b) Figeys, D.; Ducret, A.; Yates, J. R., III; Aebersold, R. *Nature Biotech.* **1996**, *14*, 1579−1583. (c) Mørtz, E.; O'Connor, P. B.; Roepstorff, P.; Kelleher, N. L.; Wood, T. D.; McLafferty, F. W.; Mann, M. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 8264−8267. (d) Jensen, O. N.; Podtelejnikov, A. V.; Mann, M. *Anal. Chem.* **1997**, *69*, 4741−4750.

[6] Mewes, H. W.; George, D. G. In *Microcharacterization of Proteins*; Kellner, R., Lottspeich, F., Meyer, H. E., Eds.; Verlagsgesellschaft mbH: Weinheim, Germany, 1994; p 220.

[7] (a) Wang, R.; Chait, B. T. In *Methods in Molecular Biology: Protein and Peptide Analysis by Mass Spectrometry*; Chapman, J. R., Ed.; Humana Press: Totowa, NJ, **1996**; Vol. 61, pp 161−170. (b) Wilm, M.; Neubauer, G.; Mann, M. *Anal. Chem.* **1996**, *68*, 527−533. (c) Blackburn, R. K.; Anderegg, R. J. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 483−494.

10.1021/ja973655h CCC: $18.00
Published on Web 01/16/1999

Such a “bottom up” approach is now common for important problems,[3,6,7] but this becomes increasingly difficult and tedious with increasing protein size. An increasing number of fragment masses must be assigned to the proposed sequences, and multiple mass modifications become more probable. The identified peptides from a single proteolysis usually represent only 50−90% of the protein sequence, frustrating the identification of mass modifications in the remainder. Even more seriously, spurious mass values are common, such as peptides formed by self-proteolysis and those from enzyme and protein impurities (even dandruff!); their masses can be mistaken for modified values of predicted peptides. In an impressive example, Mann and co-workers using MS/MS identified 14 of 22 tryptic peptides (94% sequence coverage) directly in the mixture from proteolysis of human carbonic anhydrase (29 kDa); a nearly equal number of additional mass values were not identified.[7b] For 1ck kinase (51 kDa), masses corresponding to 20 of 44 tryptic peptides were identified, but a greater number of unidentified masses were present.[7c]

Here we describe a “top down” approach in which limited dissociation of the ionized protein gives fragments sufficiently large so that the masses of one or more complementary sets of these sum to the value expected for the protein. For localization of sequence errors in five enzymes (7, 23, 27, 27, and 46 kDa) of the thiamin biosynthetic pathway,[8] and errors and derivatized active sites in thiaminase I (42 kDa)[9] and creatine kinase (43 kDa),[10] it was found far more informative and expeditious to use ESI data from order-of-magnitude larger (5−36 kDa) fragments first. Mass measurement of such large peptide and ionic fragments by Fourier transform (FT) MS with its unusually high resolving power ($>10^5$) makes possible accurate assignments of ESI charge state and mass, even for MS/MS.[11] To compare the general applicability of the “bottom up” and “top down” approaches, these are applied here to the known protein carbonic anhydrase (29 kDa), measuring far more extensive MS/MS and proteolysis data to illustrate the optimization of strategies to identify mass modifications at one or more of its 259 amino acids.

Experimental Section

Bovine carbonic anhydrase B (CA-B) from CalBiochem (Lot No. 778093); endoproteinase Lys-C, α-chymotrypsin, phenylmethylsulfonyl fluoride (PMSF), trifluoroacetic acid (TFA), urea, cyanogen bromide, and HPLC grade solvents from Sigma (St. Louis, MO); and formic acid from Acros were used without further purification. The peptide DFPIANGERQSPVNIDTK was synthesized at the Cornell Analytical and DNA/Peptide Synthesis Facility. α-Chymotrypsin: to 100 μL of 10 μM CA-B (1 nmol) in 50 mM Tris (pH 8.2) and 2 M guanidine HC1 was added 2.5 μL of 4 μM α-chymotrypsin (10 pmol) in 50 mM Tris (pH 7.5); after 10 min at 25 °C the reaction was quenched with 1 μL of 100 mM PMSF (100 nmole) in ethanol. Lys-C: to 75 μL of 13 μM CA-B (1 nmol) in 50 mM Tris (pH 8.2) and 2 M urea was added 20 μL of 1.5 μM Lys-C (30 pmol, 1 μg) in 25 mM Tricine, 5 mM EDTA (pH 8.0); after 20 min the reaction was quenched with 2 μL of neat acetic acid and freezing (−80 °C); this was repeated with 500 μL of 10 μM CA-B (5 nmol), without urea, quenching after 30 min or 8 h. The crude digests were desalted by loading onto reversed-phase peptide traps (Michrom Bioresources, Auburn, CA), washed with 1 mL of 1:98:1 MeOH/H_2O/HOAc, and step-eluted with ∼20−30 μL of 70:28:2 MeOH/H_2O/HOAc. Peptide solutions were diluted with 25 μL of H_2O (or with dilute NH_4OH for negative ions, final pH ∼ 9). Cyanogen bromide (CNBr): to CA-B (1 nmol) in 100 μL of fresh 60:40 H_2O/formic acid was added 10 μL of 10 mM CNBr (100 nmol) in ethanol; after 2 h in the dark at 25 °C, the sample was diluted with 400 μL of H_2O. After 4 h, the solvent was removed and the residue partially redissolved in 80:18:2 MeOH/H_2O/HOAc. CA-human: human blood (5 μL) was extracted with 100 μL of 1:1:1 MeOH/H_2O/$CC1_4$, and the aqueous layer was desalted twice by ultrafiltration (10 kDa cutoff) and acidified with 400 μL of 10 mM HOAc;[4] a spectrum was acquired after 48 h.

Five microliters of the desalted solutions were loaded into either a Nanospray[12] or PicoTip[13] (New Objective, Inc., Cambridge, MA) ESI emitter with a 1−3 μm i.d. tip; an ESI voltage of 0.6−1.5 kV gave ∼1−50 nL/min flow rates. The resulting ions were guided through a heated metal capillary (110 °C), skimmer, and 3 rf only quadrupoles into the ion cell (10^{-9} Torr) of a 6 T modified Finnigan FT/MS 2000, described earlier.[14] Fragmentation of ions entering the FTMS employed nozzle-skimmer (NS) dissociation[15] with 100−200 V potential difference. Desired ions were isolated by stored waveform inverse Fourier transform (SWIFT)[16] and collisionally dissociated by using sustained off-resonance irradiation (SORI)[17] for 1.5 s at $\sim 10^{-6}$ Torr N_2, 1.2−1.6 kHz off resonance from the precursor ion. Transients were stored with an Odyssey Data Station as 128, 256, or 512 K data sets. Peptides were assigned with use of the Protein Analysis Worksheet (PAWS) created by R. Beavis.[18] Spectra were mass calibrated internally by using identified peptides unless otherwise noted. Theoretical isotopic distributions were generated by using Isopro v2.0 and fit to experimental data by least squares to assign the most abundant isotopic peak.[19] The reported mass value is that of the most abundant isotopic peak; the value of the mass difference (in units of 1.0034 Da) between the most abundant isotopic peak and the monoisotopic peak follows this mass value in italics. For example, the theoretical M_r value of CA-B is 29024.7-*17* for the most abundant isotopic peak, while the average molecular weight using the natural isotopic abundances of the elements is 29024. 85.[11a]

Results and Discussion

Top Down Sequencing Strategy. A well-known puzzle is that of detecting false coins by a minimum number of weighings. For 100 visually identical gold coins that should weigh 10 g each, finding that their total weight is less than 1000 g indicates that one or more have short weight. To find those modified, each coin could be weighed separately; obviously it is far more

[8] Kelleher, N. L.; Taylor, S. V.; Grannis, D.; Kinsland, C.; Chiu, H.-J.; Begley, T. P.; McLafferty, F. W. *Protein Sci*. **1998**, *8*, 1796−1801.

[9] (a) Kelleher, N. L.; Costello, C. A.; Begley, T. P.; McLafferty, F. W. L. *Am. Soc. Mass Spectrom.* **1995,** *6*, 981−984. (b) Costello, C. A.; Kelleher, N. L.; Abe, M.; McLafferty, F. W.; Begley, T. P. *J. Biol. Chem.* **1996**, 271, 3445−3452. (c) Kelleher, N. L.; Nicowonger, N. B.; Begley, T. P.; McLafferty, F. W. *J. Biol. Chem.* **1997,** *272*, 32215−32220.

[10] (a) Wood, T. D.; Chen, L. H.; Kelleher, N. L.; Little, D. P.; Kenyon, G. L.; McLafferty, F. W. *Biochemistry* **1995**, *34*, 16251−16254. (b) Wood, T. D.; Chen, L. H.; White, C. B.; Babbitt, P. C.; Kenyon, G. L.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 11451−11455. (c) Wood, T. D.; Guan, Z.; Borders, C. L., Jr.; Chen, L. C.; Kenyon, G. L.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 3562−3365.

[11] (a) McLafferty, F. W. *Acc. Chem. Res.* **1994**, *27*, 379−386. (b) O’Connor, P. B.; Speir, J. P.; Senko, M. W.; Little, D. P.; McLafferty, F. W. *J. Mass Spectrom.* **1995**, *30*, 88−93. (c) Speir, J. P.; Senko, M. W.; Little, D. P.; Loo, J. A.; McLafferty, F. W. *J. Mass Spectrom.* **1995**, *30*, 39−42. (d) Kelleher, N. L.; Senko, M. W.; Siegel, M. M.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom* **1997**, *8*, 380−383. (e) Williams, E. R. *Anal. Chem.* **1998**, *70*, 179A−185A.

[12] Wilm, M. S.; Mann, M. *Int. J. Mass Spectrom. Ion Processes* **1994**, *136*, 167−180.

[13] Valaskovic, G. A.; Kelleher, N. L.; Little, D. P.; Aaseruud, D. J.; McLafferty, F. W. *Anal. Chem.* **1995**, *67*, 3802−3805.

[14] Beu, S. C.; Senko, M. W.; Quinn, J. P.; Wampler, F. M., III; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 557−565.

[15] Loo, J. A.; Udseth, H. R.; Smith R. D. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 207−210.

[16] Marshall, A. G.; Wang, T. C. L.; Ricca, T. L. *J. Am. Chem. Soc.* **1985**, *107*, 7893−7897.

[17] Gauthier, J. W.; Trautman, T. R.; Jacobsen, D. B. *Anal. Chim. Acta* **1991**, *246*, 211−225. Senko, M. W.; Speir, J. P.; McLafferty, F. W. *Anal. Chem.* **1994**, *66*, 2801−2808.

[18] http://mcphar04.med.nyu.edu/software/contents.htm. Fenyö, D.; Zhang, W.; Chait, B. T.; Beavis, R. C. *Anal. Chem.* **1996**, *68*, A721−A726.

[19] Senko, M. W.; Beu, S. C.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 229−233.

AcS	H	H	W	G	Y	G	***K***	H	D	G	**P**	E	H	W	15
H	***K***	D	F	**P**	I	A	N	G	E	R	Q	S	**P**	V	30
N	I	D	T	***K***	A	V	V	Q	D	**P**	A	L	***K***	**P**	45
L	A	L	V	Y	G	E	A	T	S	R	R	**Ⓜ**	V	N	60
N	G	H	S	F	N	V	E	Y	D	D	S	Q	D	***K***	75
A	V	L	***K***	D	G	**P**	L	T	G	T	Y	R	L	V	90
Q	F	H	F	H	W	G	S	S	D	D	Q	G	S	E	105
H	T	V	D	R	***K***	***K***	Y	A	A	E	L	H	L	V	120
H	W	N	T	***K***	Y	G	D	F	G	T	A	A	Q	Q	135
P	D	G	L	A	V	V	G	V	F	L	***K***	V	G	D	150
A	N	**P**	A	L	Q	***K***	V	L	D	A	L	D	S	I	165
K	T	***K***	G	***K***	S	T	D	F	**P**	N	F	D	**P**	G	180
S	L	L	**P**	N	V	L	D	Y	W	T	Y	**P**	G	S	195
L	T	T	**P**	**P**	L	L	E	S	V	T	W	I	V	L	210
K	E	**P**	I	S	V	S	S	Q	Q	**Ⓜ**	L	***K***	F	R	225
T	L	N	F	N	A	E	G	E	**P**	E	L	L	**Ⓜ**	L	240
A	N	W	R	**P**	A	Q	**P**	L	***K***	N	R	Q	V	R	255
G	F	**P**	***K***												259

Figure 1. The DNA-derived sequence of bovine carbonic anhydrase B. Of the residues designated in bold, ***K***, **M**, and **P** indicate expected sites of Lys-C, CNBr, and CAD (MS/MS) cleavages, respectively.

efficient to weigh two groups of 50 first (a "complementary pair", vide infra), as a 500 g value would indicate a group of genuine coins. This halving process would be repeated for the low weight mass groups; if only one coin were modified, a maximum of six differential weighings would be required.[20] Analogously, the M_r value of a protein can indicate by what mass any errors or modifications have changed the DNA-derived protein M_r value. For large protein fragments (MS/MS or proteolysis) "weighed" by MS, a complementary pair (or larger set) whose masses sum to M_r then map the entire molecule and identify the region(s) that contains the mass deviation(s) (those values that differ from the sequence prediction), suggesting the "top down" terminology. Further localization of the error or modification can employ dissociation of the modified fragment; when the region of modification is reduced to <3 kDa, the localization procedure is essentially that of the conventional ("bottom up") MS/MS approach to identify the type as well as localization of the modification. A further advantage of first examining large ionic fragments or peptides is that they should be formed by cleaving the most labile protein bonds ("hot spots"), whose identity can be used for a smaller fragment to assign the most probable of several possible sequences.

These strategies will be compared by using bovine carbonic anhydrase B (CA-B), whose measured M_r value of 29024.3-*17* for the most abundant isotopic peak[11b] compared well with the 29024.7-*17* expected for the DNA-derived sequence (Figure 1; the -*17* indicates that the main component ion of this peak contains 17 ^{13}C atoms). For FTMS of such ~20−45 kDa proteins, the M_r error is usually no more than 1 Da,[8−11] an accuracy sufficient to distinguish disulfide bonds (internal 2 SH $\rightarrow$ S−S + H_2 reduces M_r by 2.0 Da). A hypothetical case will also be considered in which an incorrect M_r value of 29010.7-*17* is predicted because of an error in the DNA-derived sequence; the reader is invited to test other hypothetical errors with the table and figure data. First we will consider the extent to which the higher quantity and quality of FTMS data can improve the bottom up approach.

Chymotryptic Bottom Up Data. In a definitive study,[7b] the predicted sequence of human CA (29 kDa) was characterized by using extensive digestion; with parent ion MS/MS characterization of their molecular ion species, 14 of 22 tryptic peptides (445−2796 Da and one of 4542 Da, 94% coverage) were identified without LC, representing an average of 16.5 residues per peptide. In this FTMS study with the less specific chymotrypsin, the ESI spectrum of the unseparated mixture from proteolysis of denatured CA-B yielded 135 discernible isotopic distributions (Figure 2) representing 69 distinct mass values, with 49 corresponding to the M_r value of an expected chymotryptic peptide. ESI spectra of the same sample at other trapping potentials uncovered 15 additional M_r values, all identifiable (Table 1). Standard deviation (σ) for the measured masses (590 to 6746 Da) of the 64 identified is ±0.037 Da, with $\sigma = 0.028$ Da for the 50 most abundant. These products correlate only qualitatively with the known activity of α-chymotrypsin to produce peptides with specific C-terminal residues; the percentage values of these initiating bond cleavages (versus those found previously)[21] are W, 21% (73%); Y, 50% (69%); F, 27% (63%); L, 46% (38%); M, 33% (28%); H, 9% (17%); N, 8% (7%); and Q, 42% (7%). These proteolysis "hot spots" are useful to indicate the most probable of multiple assignment possibilities. For example, the 1670.93-*0* peptide fits either K^{17}−N^{31} (error 0.09 Da) or I^{208}−L^{222} (0.00 Da); supporting the latter, the peptides I^{208}−P,^{224} I^{208}−L,^{238} and I^{208}−L^{238} are formed by cleavage of the W^{207}/I^{208} bond and K^{211}−L^{222} and L^{202}−L^{222} by cleavage of L^{222}/K.^{223} For K^{17}−N,^{31} the bond cleavages H^{16}/K^{17} and N^{31}/I^{32} are not involved in the formation of any other peptide. The I^{208}−L^{222} assignment was confirmed by MS/MS that provided 12 additional bond cleavages (Figure 3).

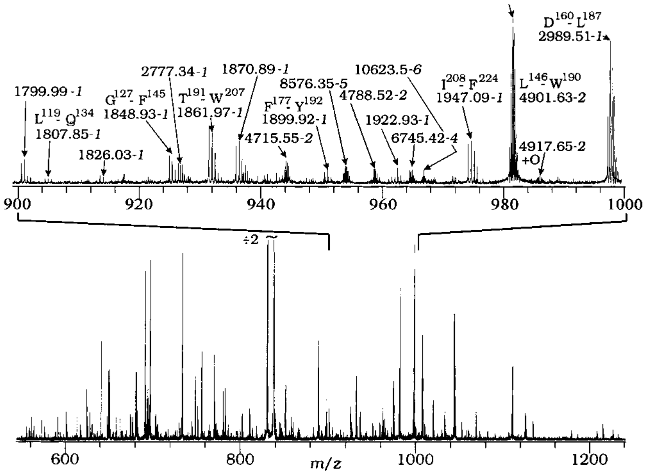

Figure 2. Broadband ESI/FTMS spectrum of all products from 10 min α-chymotrypsin digestion of partially denatured CA-B, 9 scans, 256 K data set; +O, oxidation (+15.99 Da); of a peptide with a corresponding amino acid sequence; arrows, most abundant isotopic peak.

Of the 258 amide bonds in CA-B, these 64 identified peptides represent cleavage of 62 bonds, an average of 4.2 residues per region (Figure 3). Thus the high resolution capabilities of FTMS offer more accurate assignment of fragment masses to the predicted sequence; this is of key importance for all 84 fragment masses to maximize the effectiveness of the bottom up approach. However, coverage is still only 95%; no combination makes a complementary set whose mass sum corresponds to the M_r value of the protein (an adventitious combination utilizing some of the 20 unassigned masses would obviously be misleading). No peptide contains the 14 N-terminal residues, although four peptides cover 16−113, four cover 118−259, and T^{107}-H^{118} covers the gap between these.

However, these data would be far less valuable if the measured M_r value 29024.3-*17* did not agree with the M_r of the

[20] As pointed out by Evan Williams, in the case of a single bad coin it is basically more efficient to separate the coins into three equally numbered groups and to weigh two of these against each other.

[21] Keil, B. In *Specificity of Proteolysis*; Springer-Verlag: New York, 1992.

Table 1. Peptides from Chymotryptic Proteolysis of Carbonic Anhydrase

mass	assign-ment	error, Da	mass	assign-ment	error, Da
590.34-*0*	$A^{140}-F^{145}$	0.00	1861.97-*1*	$T^{191}-W^{207}$	+0.02
656.33-*0*	$W^{243}-Q^{247}$	−0.01	1899.92-*1*	$F^{177}-Y^{192}$	+0.01
661.43-*0*	$R^{88}-F^{92}$	+0.04	1922.92-*1*^{a}	$D^{70}-Y^{87}$	−0.01
843.44-*0*	$N^{123}-F^{129}$	+0.06	1947.09-*1*	$I^{208}-F^{224}$	0.00
856.14-*0*^{a}	$A^{231}-L^{238}$	−0.01	2038.95-*1*	$V^{49}-N^{66}$	+0.01
945.51-*0*^{a}	$R^{88}-F^{94\,b}$	−0.01	2192.16-*1*	$D^{70}-L^{89}$	+0.06
	$E^{203}-L^{210}$	−0.01	2231.98-*1*	$H^{95}-Y^{113}$	−0.02
1051.58-*0*^{a}	$A^{241}-L^{249}$	+0.02	2251.05-*1*	$V^{59}-L^{78}$	+0.03
1108.63-*0*	$W^{243}-N^{251}$	+0.02	2263.28-*1*	$A^{241}-K^{259}$	+0.02
1149.69-*0*	$Q^{220}-N^{228}$	+0.06	2267.14-*1*^{a}	$D^{163}-L^{183}$	−0.01
1231.58-*0*	$N^{228}-L^{238}$	+0.01	2317.36-*1*	$I^{208}-L^{227}$	+0.04
1296.69-*0*	$A^{241}-N^{251}$	0.00	2355.11-*1*^{a}	$N^{123}-F^{145}$	−0.05
1295.66-*0*	$M^{239}-L^{249}$	−0.02	2387.23-*1*	$L^{202}-L^{222}$	−0.05
1298.72-*0*	$I^{208}-Q^{219}$	−0.02	2428.12-*1*	$N^{66}-Y^{87}$	−0.02
1345.69-*0*	$K^{211}-L^{222}$	0.00	2451.32-*1*	$S^{28}-Y^{50}$	−0.06
1351.77-*0*	$K^{147}-L^{159}$	−0.01	2507.37-*1*	$M^{239}-K^{259}$	−0.02
1378.80-*0*^{a}	$L^{119}-F^{129}$	+0.13	2516.13-*1*	$H^{93}-Y^{113}$	0.00
1406.77-*0*	$L^{240}-N^{251}$	−0.01	2690.41-*1*	$D^{163}-L^{187}$	+0.01
1408.78-*0*	$L^{238}-L^{249}$	0.00	2777.34-*1*	$V^{120}-F^{145\,b}$	−0.02
1426.76-*0*^{a}	$I^{208}-Q^{220}$	−0.04		$D^{160}-N^{185\,c}$	−0.05
1429.71-*0*^{a}	$T^{107}-H^{118}$	−0.06	2989.51-*1*	$D^{160}-L^{187}$	−0.03
1464.85-*0*	$L^{146}-L^{159}$	−0.02	3018.63-*1*	$V^{120}-K^{147}$	+0.13
1528.78-*0*	$G^{130}-F^{145}$	−0.02	3027.53-*1*	$H^{118}-F^{145}$	+0.03
1537.80-*0*	$M^{239}-N^{251}$	−0.02	3154.60-*1*^{a}	$D^{163}-W^{190}$	+0.03
1557.83-*0*	$I^{208}-M^{221}$	−0.01	3454.72-*2*	$D^{160}-W^{190}$	0.00
1596.79-*0*^{a}	$W^{15}-Q^{27}$	+0.01	3531.85-*2*^{a}	$I^{208}-L^{238}$	−0.05
1656.83-*0*	$F^{19}-D^{33}$	+1.00^f	3583.97-*2*	$H^{16}-L^{48}$	+1.03^d
1670.93-*0*	$I^{208}-L^{222\,b}$	0.00	4324.37-*2*	$K^{147}-L^{187}$	+0.05
	$K^{17}-N^{31}$	+0.09	4335.04-*2*^{a}	$V^{49}-Y^{87}$	0.00
1764.00-*0*	$L^{146}-L^{162}$	−0.01	4437.41-*2*	$L^{146}-L^{187}$	0.00
1781.80-*0*	$N^{123}-L^{139}$	−0.02	4788.52-*2*^{a}	$K^{147}-W^{190}$	+0.02
1791.97-*1*	$I^{32}-L^{48}$	−0.08	4901.63-*2*^{a}	$L^{146}-W^{190}$	+0.05
1807.85-*1*	$L^{119}-Q^{134}$	−0.02	6746.49-*4*	$L^{146}-W^{207}$	−0.02
1848.93-*1*	$G^{127}-F^{145}$	+0.02			

No match: 1402.62-*0*, 1458.75-*0*, 1562.89-*0*, 1654.96-*0*, 1669.08-*0*, 1708.89-*0*, 1799.99-*1*, 1826.03-*1*, 1870.89-*1*, 1921.00-*1*, 2005.13-*1*, 2026.11-*1*, 2064.12-*1*, 2085.95-*1*, 2438.15-*1*, 2575.32-*1*, 2986.60-*1*, 3086.44-*1*, 3176.60-*1*, 4226.28-*2*

[a] Ions appear only in the spectrum acquired with high trapping potentials. [b] Both bonds cleaved in forming this peptide are cleaved in forming others ("hot spots"). [c] One bond cleaved in forming this peptide is a "hot spot". [f] Consistent with Asn[31] deamidation.

DNA-predicted sequence. Now all of the 20 unassigned mass values (Table 1) would also have to be considered to explain the mass difference, making erroneous postulates probable, especially if the modification were in the 1−14 region not covered by peptide products. Consider the hypothetical CA-X case of the sequence predicted incorrectly as S,^{172} not T^{172} (M_r = 29010.7-*17*); now in the chymotryptic data only 54 peptides would be assigned to predicted sequence positions, with the 30 not assignable including the 10 peptides that actually represent regions 146−187, 146−190, 146−207, 147−187, 147−190, 160−187, 160−190, 163−183, 163−187, and 163−190. Assigning to these the +14 Da deviation indicated by the M_r value would be aided by the many common bond cleavages of these peptides, but some of the other 20 unassignable M_r values could also be rationalized with a 14.02 ± 0.04 Da difference, as 13.979, 14.016, or 14.052 Da is found for eight amino acid pairs (*e.g.*, A → G, 14.052). For example, the no-match peak m = 1458.75-*0* is actually 13.99 Da higher than the M_r value predicted for the peptide $W^{190}-L^{202}$. For this hypothetical CA-X sequence the peptide $L^{146}-W^{190}$ (4.9 kDa) would show measured versus predicted M_r of +14.07; the fragment masses of its MS/MS spectrum (Figure 3) would confirm this assignment except for the same deviation in residues 164−173 (b_{19} vs b_{28}). In this region, the error could result from seven replacements: I^{165} → V, $K^{166,168,\mathrm{or}170}$ → N, T^{167} or T^{172} → S, or D^{173} → T. (Only I^{165} → V and T^{167} or T^{172} → S would be possible with sufficient mass accuracy, achievable on the b_{28} peak with repeated heterodyne measurements.) [15] Bottom up proteolyses mainly yield M_r values <3 kDa; with this restriction here, error localization would be of far lower confidence. Although the number of peptides assigned here from a single proteolysis mixture (no separation) is far larger than normally reported, a recent publication shows identification of 123 peptide M_r values (of 143 measured masses) in a single MALDI spectrum (m = 700 to 2600) from a reflectron time-of-flight instrument.[5d] Indicating the potential for even larger proteins, a 9.4 T FTMS ESI spectrum of the proteolysis products of an 191 kDa protein shows 435 different mass values.[22]

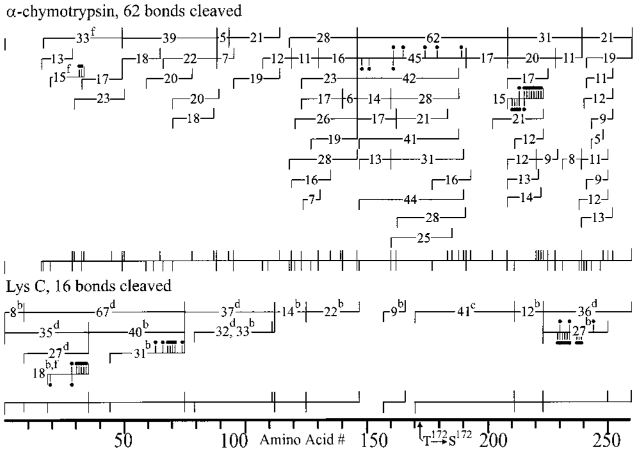

Figure 3. Map of peptides from extensive α-chymotrypsin and Lys-C proteolysis of CA-B (Tables 1 and 2). Numeric values indicate residue count; vertical bars above line, b-type ions or the C-terminus of a peptide; bars below line, y-type ions or a peptide's N-terminus; bars terminated by a solid dot, bonds cleaved from MS/MS of peptide ions. Superscripts: see footnotes for Table 2.

Lys-C Digestion. A more specific protease should yield fewer (and more abundant) peptides of more predictable M_r values, increasing the confidence of bottom up data assignments.[7] The protein region of mass modification 164−173 indicated by the bottom up chymotryptic data contains lysines K^{166}, K^{168}, and K^{170} (Figure 1), suggesting Lys-C proteolysis for more detailed characterization of this region. The ESI/FT mass spectrum of an unfractionated 8 h Lys-C digest of native CA-B (Figure 4, Table 2) shows nine masses (of a total of 12 measured) that are consistent (Figure 3, superscript b) with fragment sequences covering 165 of the protein's 259 amino acids, with MS/MS confirming the identities of $P^{44}-K^{75}$ and $F^{224}-K^{250}$ (Figure 5). No peptide masses corresponding to the regions 9−35, 76−79, 148−157, 167−211, and 251−259 were observed. However, the 41-mer $S^{171}-K^{211}$ was identified in the negative ion spectrum[23] (Figure 6); note the high ratio of acidic to basic (4:1) amino acids in this region (D^{173}, D^{178}, D^{188}, and E^{203} vs K^{2ll}).

For both the positive and negative ion spectra, peaks at 2001.99-*1* Da (Figure 4) and 2002.02-*1* Da (Figure 6) were not assignable as an expected Lys-C peptide. SORI fragmentation of the 2001.99 Da ion yields (Figure 7b) some ions consistent with those expected from $D^{18}-K^{35}$ (b_{11}, b_{13}), while b_{14} and b_{15} reveal deamidation of Asn[31] to cause the observed +0.99 Da mass error. Three peptides earlier along the Lys-C proteolytic

[22] Kelleher, N. L.; Belshaw, P. J.; Walsh, C. T. In preparation. Horn, D.; Zubarev, R. A.; McLafferty, F. W. *J. Am. Soc. Mass Spectrom.* In preparation.

[23] Loo, J. A.; Loo, R. R. O.; Light, K. J.; Edmonds, C. G.; Smith, R. D. *Anal. Chem.* **1992**, *64*, 81−88.

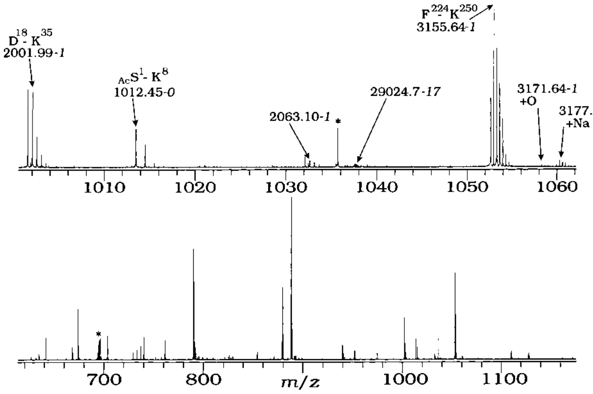

Figure 4. Broadband ESI/FTMS spectrum of products from 8 h Lys-C digestion of native CA-B, single scan, 512 K data set; +O, oxidation; +Na, substitution for H of another peptide; asterisk, noise peak.

Table 2. Peptides from Lys-C Proteolysis of Carbonic Anhydrase

mass^a	assign-ment	error, Da^a	mass^a	assign-ment	error, Da^a
972.55-*0*$^{b-d}$	$V^{158}-K^{166}$	0.00	4213.10-*2*^{d}	$A^{76}-K^{112}$	0.00
1012.45-*0*$^{b-d}$	$_{Ac}S^1-K^8$	0.00	4239.26-*2*^{d}	$F^{224}-K^{259}$	+0.01
1345.69-*0*$^{b-d}$	$E^{212}-K^{223}$	0.00	4435.17-*2*$^{b-e}$	$A^{36}-K^{75}$	0.00
1580.81-*0*$^{b-d}$	$Y^{113}-K^{125}$	0.00	4594.01-*2*^{c}	$S^{171}-K^{211}$	−0.01
2001.99-*1*$^{b-e}$	$D^{18}-K^{35}$	+0.99^f	7542.64-*4*^{d}	H^9-K^{75}	−0.02
2253.15-*1*$^{b-d}$	$Y^{126}-K^{147}$	0.00	8539.11-*5*^{e}	$_{Ac}S^1-K^{75}$	+1.01^g
3124.48-*1*^{d}	H^9-K^{35}	−0.01	9962.05-*6*^{e}	$A^{76}-K^{166}$	−0.06
3155.64-*1*$^{b-d}$	$F^{224}-K^{250}$	0.00	10330.2-*6*^{e}	$G^{169}-K^{259}$	−0.2
3513.64-*2*$^{b-d}$	$P^{45}-K^{75}$	−0.01	10558.7-*6*^{e}	$T^{167}-K^{259}$	−0.8^g
3673.75-*2*^{d}	$D^{80}-K^{111}$	+0.03	16281.2-*10*^{e}	$A^{76}-K^{223}$	−0.2
3801.80-*2*^{b,d}	$D^{80}-K^{112}$	−0.01	20503.7-*12*^{e}	$A^{76}-K^{259}$	+0.1
4119.93-*2*^{d,e}	$_{Ac}S^1-K^{35}$	−0.01			

No match: 938.30-*0*,^{c}972.55-*0*,^{c} 1748.98-*0*,^{e} 1877.03-*1*,^{b} 1885.95-*1*,^{d} 1925.94-*1*,^{d} 1983.98-*1*,^{c} 2063.10-*1*,^{b} 2159.06-*1*,^{d} 2538.43-*1*,^{e} 2619.28-*1*,^{d} 3004.49-*1*,^{d} 3100.51-*1*,^{d} 3112.24-*1*,^{d} 3308.72-*2*,^{d} 3378.81-*2*,^{d} 3451.82-*2*,^{d} 3728.94-*2*,^{d} 3915.00-*2*,^{d} 3955.38-*2*,^{e} 4656.28-*2*,^{c} 10384.2-*6*,^{e} 15589.4-*9*^{e}

[a] Data from first spectrum indicated. [b] Native CA-B, 8 h digest, positive ions (Figure 4). [c] Native CA-B, 8 h digest, negative ions (Figure 5). [d] Partially denatured CA-B, 20 min digest, positive ions (spectrum no shown). [e] Native CA-B, 30 min digest (Figure 8). [f] Consistent with Asn^{31} deamidation. [g] S/N ratio <3:1.

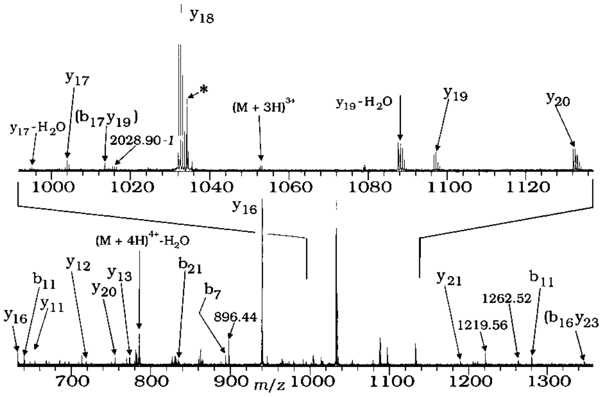

Figure 5. MS/MS spectrum (SORI dissociation, single scan) of *m/z* 790^{4+} ions in Figure 4; asterisk, noise peak.

pathway show little evidence of this deamidation, although the low signal/noise could mask the presence of a minor amount (Figure 7a). This NH_2 → OH change could result from either nonenzymatic autodeamidation[24] or a unique reactivity of Lys-C toward this peptide. To test the latter, a synthetic 18-mer with the DNA-predicted sequence $D^{18}-K^{35}$ *(i.e.,* with N^{31}) was treated with Lys-C. No evidence for deamidation was observed; conceivably deamidation could have accompanied enzymatic cleavage of the $K^{17}-D^{18}$ or $K^{35}-A^{36}$ bonds. Incubation of the peptide at pH 8.5 overnight also did not cause deamidation. Although two Asn-Gly dipeptidyl sequences ($N^{23}G^{24}$, $N^{61}G^{62}$) of holo-CA-B are known to autodeamidate,[24] $Asn^{31}Ile^{32}$ has not been detected as $Asp^{31}Ile^{32}$. As a further anomaly, one of three chymotryptic peptides covering this region, $F^{19}-D^{33}$, is 1.00 Da higher than expected (Table 1). Thus, accurate mass MS/MS can derive useful information even from unexpected (and unexplained) anomalies.

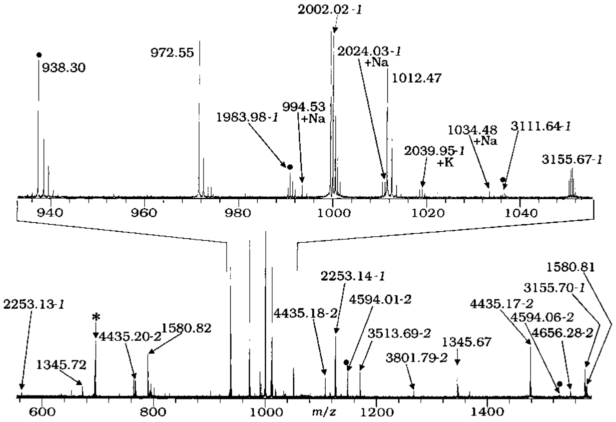

Figure 6. Negative ion spectrum of a 8 h Lys-C digest of native carbonic anhydrase, single scan. Dot, peptide not observed in any positive ion data; +Na, +K, substitution for H of another peptide; asterisk, noise peak.

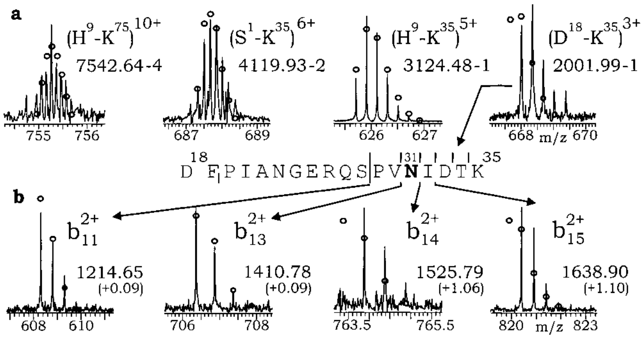

Figure 7. (a) Expanded molecular ion regions for peptides containing N^{31} or D^{31} from Figure 4 (single scan). (b) MS/MS fragment ions from SORI dissociation of 2001.99-*1* Da ions of (top right) 5 scans; parentheses, errors using external frequency calibration. Open circles: isotopic peak abundances calculated from the CA-B sequence.

A 20 min Lys-C digest of partially denatured CA-B produced masses corresponding to six additional peptides (Table 2); although $V^{148}-K^{157}$ and $T^{167}-K^{170}$ were not identified, peptides from the two digests represent (Figure 3, bottom) cleavages at 16 of 17 expected Lys sites (not K^{168}; K^{259} is C-terminal), sites not cleaved by chymotrypsin (Figure 3, top). Standard deviation for the Lys-C peptide mass values of S/N >3:1 (Table 2) is ±0.018 Da. For the hypothetical predicted sequence of CA-X, the negative ion M_r value of the Lys-C peptide $S^{171}-K^{211}$ would be 14.0 low; with the chymotryptic data, this would restrict the sequence error to residues 171−173, $S^{172} \rightarrow T^{172}$ ($\Delta m = 14.016$ Da) and $T^{173} \rightarrow D^{173}$ ($\Delta m = 13.979$ Da). This bottom up sequencing with chymotrypsin and/or Lys-C proteolysis is substantially aided by the high resolving power and mass accuracy of FTMS; however, assignment to the proposed sequence must be attempted for every measured mass (with MS/MS confirmation for doubtful cases) to find first the correctly predicted sequence regions.

[24] Wright, H. T. *Crit. Rev. Biochem. Mol. Biol.* **1991,** *26,* 1−52.

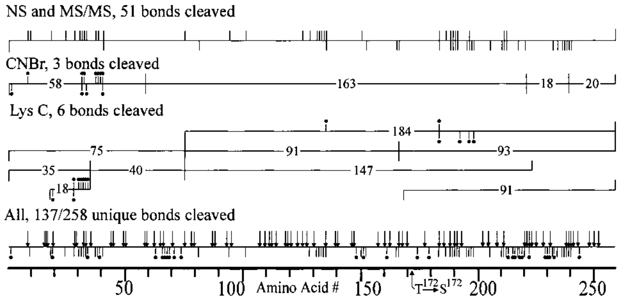

Figure 8. Map of “top down” mass data from CA-B of the following: (top) mass fragments from molecular ion dissocation, CNBr digestion, Lys-C, 30 min on native CA-B, designations as in Figure 3 and (bottom) summary of bond cleavages, arrows, CNBr and α-chymotrypsin and Lys-C proteolysis, vertical bars below line, MS/MS (dots are MS/MS of peptides).

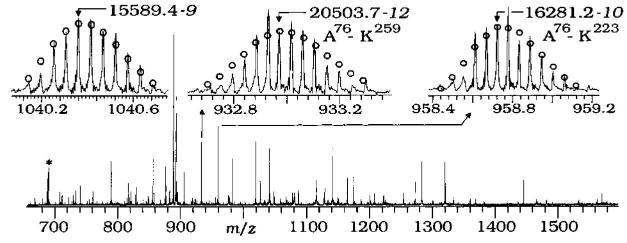

Figure 9. ESI/FT mass spectrum (single scan) of products from a 30 min Lys-C digest of native CA-B; insets, as in Figure 7; asterisk, noise peak.

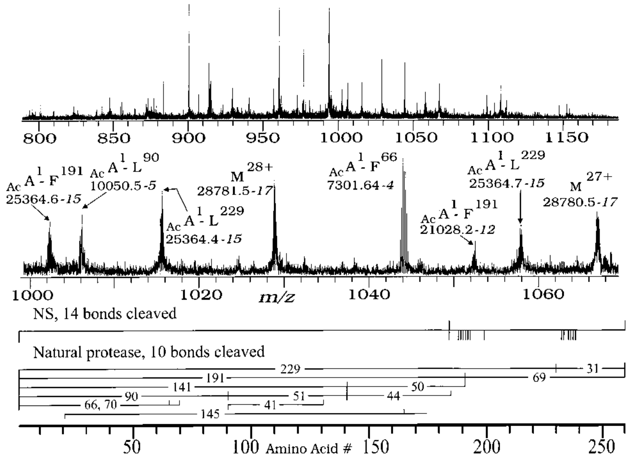

Figure 10. Partial proteolysis products from CA-human in an acidic blood extract, single scan. Bottom, maps of fragment ions from CA-human molecular ion dissociation and of observed peptide masses.

Alternative Top Down Sequencing. If a set of fragments can be found whose masses sum to the measured M_r value, only matching these to the predicted sequence identifies the correct and incorrect regions. Extensive MS/MS studies of nozzle-skimmer (NS)[15] and SORI[17] collisional dissociation of CA-B ions reported[11b] fragment ions representing the cleavage of 51 of the 258 bonds of CA-B (Figure 8, top); masses of eight complementary pairs sum to ∼29024.3-*17*, the measured M_r value (e.g., $b_{192} + y_{67}$, 21422.81-*13* + 7601.52-*4* = 29024.33-*17*; $b_{135} + y_{124}$, 15319.18-*9* + 13705.20-*8* = 29024.38-*17*). For the incorrect sequence CA-X, $S^{172} \rightarrow T^{172}$, only one ion of each of these pairs would match a predicted fragment sequence. For these, the largest N-terminal and C-terminal fragment ions are b_{135} and y_{76}, restricting the 14.0 Da error to the 135−183 residue region; the internal ion representing this region is also present, making a complementary trio with b_{135} and y_{76}. Internal ions representing P^{136}−D^{178} and L^{162}−D^{178} and the y_{93}, y_{94}, and larger y ions are 14.0 Da larger than the CA-X sequence prediction, isolating the error to the 167−178 residue region. The 51 bonds cleaved by MS/MS of the protein correspond to 5.2 residues per region.

Top Down Approach with Protein Degradation. MS/MS fragmentation of multiply charged protein ions does not always yield complementary ion fragments, e.g., protein A (45 kDa) and porcine serum albumin (67 kDa).[11c] Seeking a highly specific reagent for the “top down” approach, CA-B was subjected to the Met-specific reagent cyanogen bromide; methionine is a far less common amino acid than lysine, so that CNBr should generate much larger protein fragments. This chemical degradation gave four identified products of 2074.12-*1* (2092.12-*1*), 2376.41-*1*, 6539.33-*3* (6557.33-*3*), and 17942.3-*10* Da (plus the unidentified 1925.25-*1* and 2175.09-*1* Da); for some products both the homoserine lactone form and its ring opened form (+18.01 Da, in parentheses) were observed. Correcting for 30.00 Da (−SCH_3 replaced by −OH) added per amide bond cleaved, these four products sum to 29022.2-*15* (i.e., 29024.2-*17*), versus the predicted M_r value of 29024.7-*17*. MS/MS of the 6539.33 component yielded 11 identified fragment ions (Figure 8), verifying its 1−58 residue assignment. However, for the hypothetical CA-X sequence, the +14.0 Da error would only be isolated to a 163 residue region.

Proteolysis for shorter times should also produce larger peptide pieces. A Lys-C digest of native CA-B was halted after 30 min to yield (Figure 9) 63 isotopic distributions representing nine identified peptides of masses 1.7 to 20.5 kDa. Correlation of these to those possible from the sequence quickly maps the entire DNA-derived sequence of CA-B (Figure 8) with several complementary sets of peptide M_r values (e.g., 1−75, 76−166, and 167−259, with two H_2O added, sum to 29023.8-*17*). Assignment of the 20503.7-*12* component is confirmed by MS/MS yielding six fragment ions that include the b_{108} and y_{76} complementary pair. For the case of the incorrect DNA-derived sequence of CA-X, this pair and the G^{169}−K^{259} peptide would have limited the error to the region G^{169}−L^{183}.

Optimum Proteolysis-MS/MS Procedure. For the top down approach for the CA-B sample with the incorrect sequence CA-X ($T^{172} \rightarrow S^{172}$), CNBr produces four complementary fragments that isolate the error to the 59−221 residue region. Using instead MS/MS, a complementary ion trio immediately limited the error to the 135−183 region, with more extensive MS/MS shrinking this to a 167−178 region. Limited Lys-C proteolysis gave a complementary trio restricting the error to the 167−259 region, with further Lys-C proteolysis (Figure 3) or MS/MS shrinking this to either the 171−211 or the 169−183 region, respectively. With these restrictions, analysis of the 84 chymotryptic masses would be far easier, restricting the error further to the 171−176 residue region, with the MS/MS data shrinking this to a 171−173 region. Considering other hypothetical erroneous sequences, the largest region without a cleavage by any of these methods (Figure 8, bottom row) contains eight residues.

CA-Human. After extraction of CA from human red blood cells, ESI with nozzle-skimmer[15] fragmentation gave a measured M_r value of 28780.6-*17*, a +42.2 Da discrepancy from the predicted value,[4] plus 15 fragment ions including the b_{185}/y_{75} complementary pair (Figure 10, bottom) showing that the discrepancy is in the N-terminal fragment. In a routine confirmation, it was found unexpectedly that natural proteolysis during a 48 h storage of the same sample gave a spectrum (Figure 10) showing 14 peptides from 2.4 to 25.3 kDa. Thirteen of these correspond to peptides formed by cleavage on the

Table 3. Peptides from Natural Proteolysis of Human Carbonic Anhydrase

mass	assign-ment	error, Da[a]	mass	assign-ment	error, Da[a]
3433.79-*1*	$L^{230}-F^{260}$	+0.04	10050.3-*5*	$_{Ac}A^{1}-L^{90}$	+0.4
4728.21-*2*	$F^{91}-L^{131}$	+0.02	15676.1-*9*	$_{Ac}A^{1}-L^{141}$	+0.6
5369.05-*3*	$A^{142}-F^{191}$	+0.06	15846.7-*9*	$Y^{20}-L^{164}$	−0.2
5642.67-*3*	$F^{91}-L^{141}$	+0.03	21026.8-*12*	$_{Ac}A^{1}-F^{191}$	+0.3
7771.04-*4*	$W^{192}-F^{260}$	+0.11	25364-*15*	$_{Ac}A^{1}-L^{229}$	+0.8
7301.64-*4*	$_{Ac}A^{1}-F^{66}$	+0.07	28780.5-*17*	M^{+}	+0.1
7798.88-*4*	$_{Ac}A^{1}-F^{70}$	+0.08			

No match: 2435.21-*1*, 4721.23-*2*. Considering possible cleavage by other residues (*i.e.*, Y, W, and M), no additional matches were found within ±2 Da

[a] External calbration.

C-terminal side of F or L residues (Table 3). Three complementary sets were observed:1−229 + 230−260; 1−191 + 192−260, and 1−141 + 142−191 + 192−260. These adventitious data map the entire DNA-predicted sequence of human-CA (Figure 10, bottom) with localization of the +42.2 Da discrepancy to the N-terminal 19 residues (*N*-acetylation, +42.01 Da, would be consistent with that observed for CA-B). Peptides corresponding to the cleavage of 13 bonds covering 94% of human CA were identified from trypsin cleavage;[7b] here 100% coverage is obtained by accidentally cleaving 10 bonds with a natural protease. Although peptide MS/MS was not tried, this should localize the error further.

Top Down Approach for Larger Proteins. Initial mapping of the entire sequence should be even more valuable for proteins far larger than 29 or ∼43 kDa.[8−10] The DNA-predicted M_r value would be checked by ESI/FTMS and/or MALDI[1] to derive a value for the mass difference due to DNA-derived sequence errors or protein modifications. Next, MS/MS fragmentation and/or proteolysis would be tried to obtain a complementary set of fragments of sizes up to 50 kDa.[8−10] The DNA-derived sequence would indicate the protein fragments modified; for these, identifying the modified site would proceed as above with more extensive proteolysis and/or MS/MS. In a now published example of the 379-residue Thiaminase, the top down approach showed the location of multiple errors in the DNA predicted sequence[9a] and also that a suicide substrate binds covalently to C^{113}.[9c] For rabbit creatine Kinase,[10c] the 1:1 stoichiometry of the phenylglyoxal active site binding presumed to involve one of 18 arginines was shown to be due to partial binding at R^{291} and two of the three residues R^{129}, R^{131}, or R^{134}. For a 50-mer DNA,[25] assignment of MS and MS/MS spectral peaks characterized its complete sequence; a single mutation in a new 50-mer was characterized easily by observing a 9.0 Da reduction in M_r indicating A → T (313.06 → 304.05), with 5′-terminal fragments ($\underline{a}_{25}$) of the same mass (7932.49 vs 7932.36), but $\underline{a}_{27}$ fragments of 8535.57 vs 8526.49; because the original sequence has $C_{26}-A_{27}$, the mutation must be $A_{27} \rightarrow T_{27}$. For the 190 kDa immmunoglobulin IgE (two identical light and two identical heavy chains), with sequence predictions only for the nonvariant regions, to date the top down approach has given extensive sequence information for both the light and heavy chains.[26] A new MS/MS method, electron capture dissociation (ECD),[27] provides complementary, and far more extensive, sequence information versus conventional ion dissociation techniques. These techniques applied to ubiquitin, 8.6 kDa, have given its complete sequence except for four residue pairs. Although ECD was discovered after this work was completed, it could make the top down approach even useful for *de novo* sequencing.

Acknowledgment. The authors thank Tadhg Begley, Barbara Baird, Nathan Kruger, Petia Shipkova, David Horn, Evan Williams, and Michael Senko for experimental assistance and helpful discussions. We also thank the National Institutes of Health (F.W.M., GM16609) and the American Chemical Society (N.L.K., ACS Division of Analytical Chemistry full fellowship sponsored by Perkin-Elmer Corporation) for generous funding.

JA973655H

(25) Little, D. P.; Aaserud, D. J.; Valaskovic, G. A.; McLafferty, F. W. *J. Am. Chem. Soc.* **1996**, *118*, 9352−9359.

(26) Fridriksson, E. K.; Baird, B.; Holowka, D.; McLafferty, F. W. In preparation.

(27) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. *J. Am. Chem. Soc.* **1998**, *120*, 3265−3266. Zubarev, R. A.; Kruger, N. A.; Fridriksson, E. K.; Lewis, M. A.; Horn, D. M.; Carpenter, B. K.; McLafferty, F. W. *J. Am. Chem. Soc*. Submitted for publication. Kruger, N. A.; Zubarev, R. A.; Carpenter, B. K.; Kelleher, N. L.; Horn D. M.; McLafferty, F. W. *Int. J. Mass Spectrom*. In press. Kruger, N. A.; Zubarev, R. A.; Horn D. M.; McLafferty, F. W. *Int. J. Mass Spectrom*. In press.