

similar velocity widths of  $\sim 100 \text{ km s}^{-1}$  exhibit vertical elongations of more than 200 pc with thin widths of 30 pc. These vertical features are consistent with the foot points of a magnetic loop in its late stage of time evolution (13). The NANTEN data set indicates that a weak loop-like feature that has a large velocity gradient is “bridging” these vertical features at higher latitudes between  $l \approx 3^\circ$  and  $5^\circ$ , suggesting the existence of another magnetic loop in the positive longitude (SOM text and fig. S6). In addition, the model offers naturally substantial heating of the warm molecular gas at the foot points; the velocity dispersion ( $\sim 15$  to  $30 \text{ km s}^{-1}$ ) of the broad CO features corresponds to kinetic temperature higher than about  $10^4 \text{ K}$  if the shock is completely converted into thermal energy at the foot points. We suggest, therefore, that the present model has the potential to be applied to the other salient broad velocity features in the galactic center and to the heating of the molecular gas at their foot points.

The present model shares the common physics of solar loops. It has been well understood that solar loops are the results of magnetic flotation driven by the motion of the magnetic field lines anchored onto the solar granules (26). The size scale of the solar loops is  $\sim 12$  orders of magnitude smaller than those in the galactic center, and this is a natural consequence of the scale height on the solar surface, only  $\sim 200 \text{ km}$ , which determines the height of the loop. Despite the large difference in size,

the Alfvén speed is  $\sim 10 \text{ km s}^{-1}$  as calculated from the density ( $\sim 2.5 \times 10^{-7} \text{ g cm}^{-3}$ ) and the magnetic field strength ( $\sim 500 \text{ G}$ ) on the solar surface (26). The small size of the solar loop determines the much shorter time scale of the solar flotation as  $\sim 10^3 \text{ s}$ , as compared with  $\sim 10^{14} \text{ s}$  in the galactic center loops.

#### References and Notes

1. F. Yusef-Zadeh, M. Morris, D. Chance, *Nature* **310**, 557 (1984).
2. M. Morris, E. Serabyn, *Annu. Rev. Astron. Astrophys.* **34**, 645 (1996).
3. T. N. LaRosa *et al.*, *Astrophys. J.* **626**, L23 (2005).
4. D. T. Chuss *et al.*, *Astrophys. J.* **599**, 1116 (2003).
5. M. Morris *et al.*, *Astrophys. J.* **399**, L63 (1992).
6. A. Mizuno, Y. Fukui, *The Structure and Evolution of Our Galaxy*, D. Clemens, R. Shah, T. Brainerd, Eds. (ASP Conf. Ser. No. 317, ASP, Boston, MA, 2004), pp. 59–65.
7. Materials and methods are available as supporting material on Science Online.
8. T. Oka, T. Hasegawa, F. Sato, M. Tsuboi, A. Miyazaki, *Astrophys. J. Suppl. Ser.* **118**, 455 (1998).
9. M. Tsuboi, T. Handa, N. Ukita, *Astrophys. J. Suppl. Ser.* **120**, 1 (1999).
10. Y. Fukui *et al.*, *Publ. Astron. Soc. Japan* **51**, 751 (1999).
11. K. Matsunaga *et al.*, *Publ. Astron. Soc. Japan* **53**, 1003 (2001).
12. E. N. Parker, *Astrophys. J.* **145**, 811 (1966).
13. R. Matsumoto, T. Horiuchi, K. Shibata, T. Hanawa, *Publ. Astron. Soc. Japan* **40**, 171 (1988).
14. K. Shibata, R. Matsumoto, *Nature* **353**, 633 (1991).
15. M. Machida, M. R. Hayashi, R. Matsumoto, *Astrophys. J.* **532**, L67 (2000).
16. W. B. Burton, H. S. Liszt, *Astrophys. J.* **225**, 815 (1978).
17. G. Güsten, S. D. Philipp, in *Proceedings of the Fourth Cologne-Bonn-Zermatt Symposium*, Zermatt, Switzerland, 22 to 26 September 2003, S. Pfalzner, C. Kramer,

- C. Staubmeier, A. Heithausen, Eds. (Springer, Heidelberg, Germany, 2004), pp. 253–263.
18. T. Horiuchi, R. Matsumoto, T. Hanawa, K. Shibata, *Publ. Astron. Soc. Jpn.* **40**, 147 (1988).
19. J. Bally, A. A. Stark, R. W. Wilson, C. Henkel, *Astrophys. J.* **324**, 223 (1988).
20. K. I. Uchida, R. Güsten, *Astron. Astrophys.* **298**, 473 (1995).
21. T. Oka, G. R. Thomas, M. Goto, T. Usuda, J. B. McCall, *Astrophys. J.* **632**, 882 (2005).
22. I. A. Harris, T. D. Jaffe, M. Silber, R. Genzel, *Astrophys. J.* **294**, L93 (1985).
23. J. N. Rodríguez-Fernández *et al.*, *Astron. Astrophys.* **365**, 174 (2001).
24. M. T. Bania, *Astrophys. J.* **216**, 381 (1977).
25. M. Bitran, H. Alvarez, L. Bronfman, J. May, P. Thaddeus, *Astrophys. J. Suppl. Ser.* **125**, 99 (1997).
26. S. Nozawa, *Publ. Astron. Soc. Jpn.* **57**, 995 (2005).
27. We thank all of the members of the laboratory who contributed to these observations at Las Campanas Observatory (LCO). We appreciate the hospitality of all the staff members of LCO of the Carnegie Institution of Washington (CIW). This project is based on a mutual agreement between Nagoya University and the CIW. We also acknowledge that the operation of NANTEN is made possible by contributions from many Japanese public donors and companies. This work was financially supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports, and Culture of Japan and from Japan Society for the Promotion of Science.

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5796/106/DC1  
Materials and Methods  
SOM Text  
Figs. S1 to S6  
References

24 May 2006; accepted 14 August 2006  
10.1126/science.1130425

## Extending Top-Down Mass Spectrometry to Proteins with Masses Greater Than 200 Kilodaltons

Xuemei Han,<sup>1\*</sup> Mi Jin,<sup>1\*</sup> Kathrin Breuker,<sup>2</sup> Fred W. McLafferty<sup>1†</sup>

For characterization of sequence and posttranslational modifications, molecular and fragment ion mass data from ionizing and dissociating a protein in the mass spectrometer are far more specific than are masses of peptides from the protein’s digestion. We extend the  $\sim 500$ -residue,  $\sim 50$ -kilodalton (kD) dissociation limitation of this top-down methodology by using electrospray additives, heated vaporization, and separate noncovalent and covalent bond dissociation. This process can cleave 287 interresidue bonds in the termini of a 1314-residue (144-kD) protein, specify previously unidentified disulfide bonds between 8 of 27 cysteines in a 1714-residue (200-kD) protein, and correct sequence predictions in two proteins, one with 2153 residues (229 kD).

Identifying a protein and characterizing its structure from its mass and masses of its backbone fragments are major proteomics research capabilities. The “top-down” methodology (1) analyzes intact proteins by combining electrospray ionization (ESI) (2) with high-performance mass spectrometry (MS) such as Fourier transform (FT) MS (3). Five to 10 accurate mass values of backbone fragment ions from dissociation of an ionized protein are suf-

ficient to identify it from among the many thousands predicted by the DNA sequence (4, 5). However, the largest molecular ions that have yielded interresidue cleavages contain  $\sim 580$  and 663 residues (6, 7), with molecular sizes of 67 and 74 kD, respectively. Although no molecular weight ( $M_w$ ) limitation exists for the “bottom-up” approach (8–10) that initially digests the protein into small peptides, for identification of proteins (5) top-down MS gives fragments that

cover a much larger range of mass values, and these can be produced from a single protein’s  $M_w$ -selected molecular ions.

The dominant proteomics application of top-down MS has been for the characterization of posttranslational modifications and sequence errors (1, 4–7, 11). For bottom-up MS, the measured peptide mass values are matched against those of the peptides expected from the DNA-predicted proteins that are unmodified; to detect a modification, matching must also be done against the predicted peptide masses adjusted separately for all likely modifications. Further, bottom-up MS usually achieves only 40 to 90% sequence coverage (8–10). For top-down MS, in contrast, a discrepancy between the measured  $M_w$  value and that predicted by the DNA sequence directly shows the presence of a modification(s) or sequence error(s). The location of the modification(s) in the protein is then indicated by a corresponding discrepancy(s) in

<sup>1</sup>Department of Chemistry and Chemical Biology, Baker Laboratory, Cornell University, Ithaca, NY 14853, USA. <sup>2</sup>Institute of Organic Chemistry and Center for Molecular Biosciences Innsbruck (CMBI), University of Innsbruck, Innrain 52a, 6020 Innsbruck, Austria.

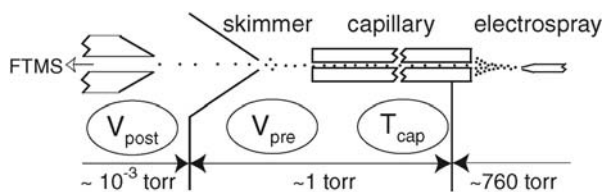
\*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: fwm5@cornell.edu

the fragment ion mass values (1, 4–7, 11). If, in an unknown protein, the measured fragment masses represent random backbone cleavages at 10% of the interresidue sites, it should be possible to locate a single protein modification to within 10 residues, on average. Mass spectra of the 76-residue ubiquitin protein (8.6 kD) gave fragment ions representing cleavage of all 75 interresidue bonds, making possible its de novo sequencing (12), but the largest (~580- and 663-residue) protein ions previously dissociated gave <20 cleavages each (6, 7).

For proteins larger than ~500 residues, the intractability for top-down MS appears to result from the increasing complexity of the gaseous molecular ion's tertiary conformer structure (13–17). Solvent removal from a denatured protein during ESI permits (13–17) much stronger electrostatic intramolecular interactions such as hydrogen bonds and salt bridges by removing the competition from water, although this process also weakens hydrophobic interactions.

**Fig. 1.** Prefolding dissociation. Protein solution was electro-sprayed from the right into the MS entrance capillary, then heated for ion desolvation and folding retardation; the ions were accelerated by  $V_{pre}$  through the pre-skimmer collision region of short mean-free path to cleave weak noncovalent bonds. Ions were then accelerated by  $V_{post}$  through the postskimmer region of long mean-free path to cleave strong backbone covalent bonds.

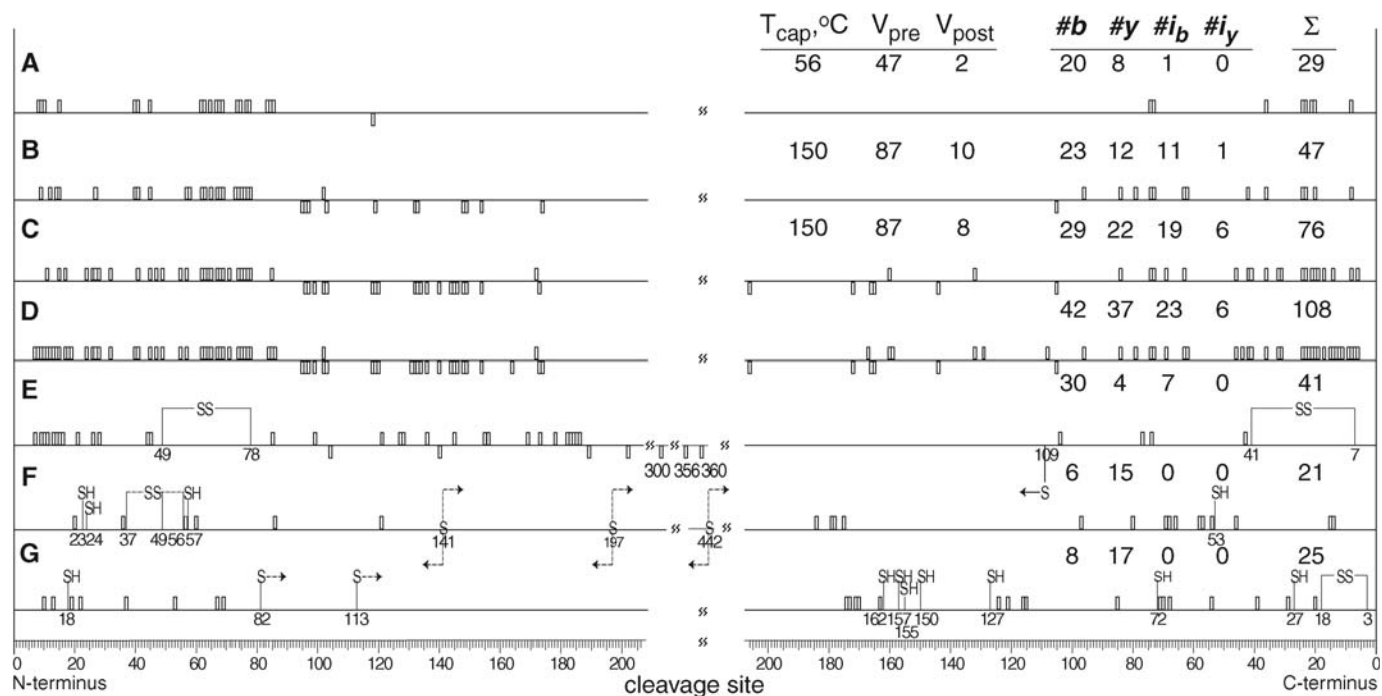


Immediately after desolvation of ubiquitin ions electro-sprayed from native solution, they were found (17) to undergo almost no nozzle-skimmer dissociation (18), presumably from fast folding of the nascent gaseous ions. This unusual ion stabilization could be greatly reduced and extensive fragment mass data obtained by the immediate addition of energy in several ways to the newly formed ions (17). Here this “prefolding dissociation” (PFD) (Fig. 1) and a method for conformer disruption that involves ESI solution additives are applied to large proteins, achieving cleavage of 108, 287, 87, and 62 interresidue bonds in the termini of 1023-, 1314-, 1714-, and 2153-residue proteins, respectively. A single PFD spectrum, consuming <1 pmol of protein (19), defines as many as 100 cleavages.

This intractability of large proteins with conventional ion-dissociation methods was re-investigated with the protein  $\beta$ -galactosidase ( $\beta$ -Gal; 1023 amino acids, 116 kD) (21). After

its molecular ions were trapped in the FTMS cell, they were subjected to infrared multiphoton dissociation (IRMPD) (17, 22) under strenuous conditions (80-ms, 27-W  $\text{CO}_2$  laser) that resulted in 29 interresidue cleavages (Fig. 2A). However, these conditions produced no cleavages from the three larger proteins studied here. The alternative PFD delays the postulated rapid formation of intramolecular noncovalent bonds both by (Fig. 1) heating the ESI inlet capillary (temperature,  $T_{cap}$ ) (16) and by collisional activation (acceleration by voltage  $V_{pre}$ ) in the subsequent preskimmer region (~1-Torr pressure, multiple low-energy collisions). Next, collisions (acceleration separately by  $V_{post}$ ) in the following ~ $10^{-3}$ -Torr postskimmer region are of higher energy (far longer mean free path) and thus provide more efficient cleavage of the stronger covalent backbone bonds of the protein ions (17). However, dissociation is effected before a specific molecular ion mass can be separated; precursor ion selection is required for MS/MS of mixture samples. Purified proteins and their binary mixtures were examined here.

Experimental variables of the new PFD methods were investigated first with the linear formylglycinamide ribonucleotide amidotransferase (PurL), predicted to have 1315 residues (23). Its ESI mass spectrum gave  $M_w = 143,500 \pm 23$ , inconsistent with the DNA-sequence predicted value of 143,635. Investigating the effect of capillary temperature,  $T_{cap} = 45^\circ\text{C}$ , gave no backbone cleavages for very vigorous subsequent collision



**Fig. 2.** (A) IRMPD mass spectrum of  $\beta$ -Gal (1023 residues). PFD mass spectra of (B to D)  $\beta$ -Gal [(B), single spectrum without and (C) with 500  $\mu\text{M}$  ammonium tartrate and (D), sum of 17 spectra] and (E)  $\beta$  (656 residues), (F)  $\alpha$  (767 residues), and (G)  $\gamma$  (291 residues) chains of human complement C4. Open vertical bars show backbone cleavages. Above the line: left, b ions; right, y ions; below the line: left, i<sub>b</sub>

ions; right, i<sub>y</sub> ions. For (E) to (G), the cysteines (numbered below the line from each terminus) that were found to be unsubstituted are indicated by a single vertical bar topped with -SH; other Cys residues are part of S-S bonds, with a solid line for the S-S bond indicating the identification of both Cys residues. The S-S bond in the  $\alpha$  chain is at 37-49, 37-56, or 49-56. Column captions are defined in Table 1.

conditions ( $V_{\text{pre}} = 170$  and  $V_{\text{post}} = 18$ ); even increasing  $T_{\text{cap}}$  to  $90^\circ\text{C}$  gave no cleavages at  $V_{\text{pre}} = 125$  and  $V_{\text{post}} = 13$ , but increasing  $V_{\text{pre}}$  to 145 yielded extensive product ions (Table 1). Of these mass values, 27 could be assigned to the predicted C-terminal sequence; the 23 other mass values were made assignable by removal of the predicted N-terminal methionine to give a revised  $M_w$  value of 143,504 versus 143,500 observed (1314 residues). Additional PFD spectra obtained at  $T_{\text{cap}} = 150^\circ$  (Fig. 3A),  $190^\circ$ ,  $295^\circ$ , and  $345^\circ\text{C}$ , but with lower collision energies ( $V_{\text{pre}} = 134$  to 49 V and  $V_{\text{post}} = 8$  to 2 V), each provided 9 to 15 “unique” cleavages (those not found in the other spectra), for a total of 126 different cleavages. At  $T_{\text{cap}} = 190^\circ\text{C}$ , three additional spectra in which  $V_{\text{pre}}$  was decreased from 133 to 100 V while  $V_{\text{post}}$  was increased from 5 to 13 V each yielded 9 to 17 unique cleavages, which

increased the total number of different  $190^\circ\text{C}$  cleavages from 60 to 94. One and two additional  $T_{\text{cap}} = 150^\circ$  and  $295^\circ\text{C}$  spectra, respectively, increased these different cleavages to 100 and 91. These 11 PFD spectra (Fig. 3B) designated a total of 173 different cleavage sites. Increasing the activation by any of the three parameters, but particularly by  $V_{\text{post}}$ , increased the relative intensity of ions of less than 50 residues,  $I_{<50\text{aa}}$  (Table 1).

We also tried a variety of small molecules added to the ESI solution to reduce this molecular-intractability. Glycerol, ribose, dimethyl sulfoxide, 1,4-cyclohexanedione, and ammonium benzoate gave no appreciable fragmentation improvement or had a negative effect, nor did prior 6 M urea denaturation or  $80^\circ\text{C}$  heating of the protein solution in the ESI tip. However, the ammonium salts of citrate, succinate, and tartrate ( $500 \mu\text{M}$ ;

Fig. 3, C to E) increased the interresidue cleavages of PurL in a single spectrum to as many as 100. These additives also produced a more stable electrospray but did not appreciably change the charge-state distribution of these molecular ions and those of cytochrome c or the abundance of background singly charged ions (19). Four spectra with the ammonium tartrate additive under a variety of conditions ( $T_{\text{cap}} = 150^\circ\text{C}$ ,  $V_{\text{pre}} = 100$  to 140 V,  $V_{\text{post}} = 5$  to 12 V) gave 174 different cleavages (Fig. 3F), and 10 spectra ( $T_{\text{cap}} = 150^\circ$  to  $190^\circ\text{C}$ ,  $V_{\text{pre}} = 65$  to 140 V,  $V_{\text{post}} = 3$  to 12 V) with additives gave 249 cleavages. All 21 spectra gave 287 different cleavages (Fig. 3G) that represent 73% of the first 100 N- and C-terminal interresidue bonds and 64% of the second 100. Such coverage should make extensive de novo sequence information possible for unknown proteins (12). The apparent “ball of spaghetti” tertiary structure only allows for cleavages in its unraveled ends, with  $>800$  central residues untouched despite these multiple activations.

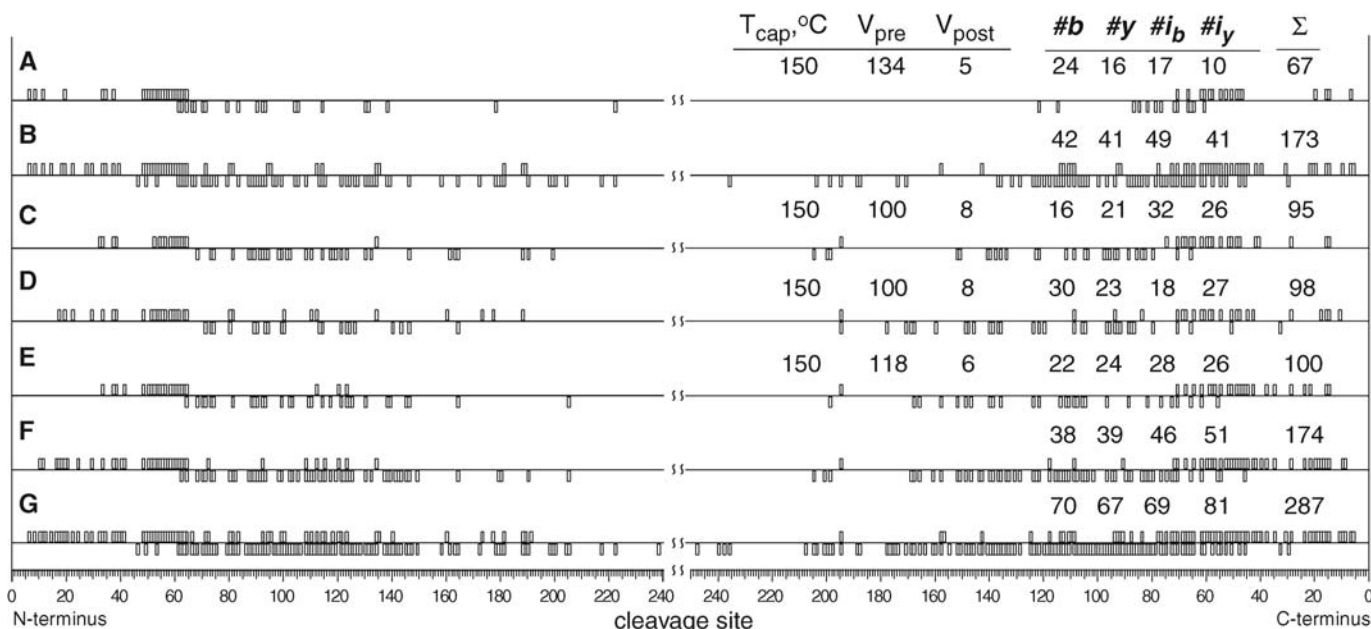
The ESI spectrum of the single-chain protein  $\beta$ -Gal (21) gave  $M_w = 116,355 \pm 12$  versus 116,352 predicted. PFD at intermediate energies gave 47 different cleavages (Fig. 2B). Added ammonium tartrate increased this number to 76 (Fig. 2C), whereas 17 spectra run under a variety of conditions gave 108 different cleavages (Fig. 2D). These included all 29 cleavages from IRMPD in the FTMS cell (Fig. 2A).

Human complement C4 glycoprotein (1714 residues,  $M_w = 200,000$  by SDS-polyacrylamide gel electrophoresis) consists of three chains linked by three disulfide bonds, but the locations of these or any other S-S bonds of

**Table 1.** Effect of excitation parameters on PFD mass spectra of PurL (1314 residues). Capillary temperature ( $T_{\text{cap}}$ ) and pre- and postskimmer collision voltages ( $V_{\text{pre}}$ ,  $V_{\text{post}}$ ) versus the number of N- and C-terminal (b, y) and secondary internal ( $i_b$ ,  $i_y$ ) product ions, the relative intensity ( $I$ ) of these ions of  $<50$  amino acids ( $I_{<50\text{aa}}$ ), and the total number of different backbone cleavages ( $\Sigma$ ), including cleavages unique ( $\Sigma_{\text{uniq}}$ ) among other values of  $T_{\text{cap}}$  or of  $V_{\text{pre}}$  and  $V_{\text{post}}$ \*

$T_{\text{cap}}$ ( $^\circ\text{C}$ )	$V_{\text{pre}}$	$V_{\text{post}}$	#b	#y	# $i_b$	# $i_y$	$I_{<50\text{aa}}$ (%)	$\Sigma$	$\Sigma_{\text{uniq}}$
90	145	13	11	17	12	10	49	50	15*
150	134	5	24	16	17	10	50	67	13*
295	60	8	23	17	12	17	36	69	12*
345	49	2	22	15	9	9	41	55	9*
190	133	5	24	14	15	7	36	60	17††
190	108	5	15	12	9	6	29	42	10†
190	105	8	15	13	10	8	49	46	9†
190	100	13	21	10	9	8	64	48	12†

\*Unique to other  $T_{\text{cap}}$  values. †Unique to other  $V_{\text{pre}}$  and  $V_{\text{post}}$  values. †† $\Sigma_{\text{uniq}} = 10$  at  $T_{\text{cap}} = 190^\circ\text{C}$ .



**Fig. 3.** PFD mass spectra of PurL (1314 residues): the effect of additives to the electrospray solution. (A) No additive. (B) Sum of 11 spectra without additives. Additives: (C)  $500 \mu\text{M}$  ammonium citrate, (D)  $500 \mu\text{M}$

ammonium succinate, and (E)  $500 \mu\text{M}$  ammonium tartrate. (F) Sum of four spectra with  $500 \mu\text{M}$  ammonium tartrate. (G) Sum of all 21 spectra. Designations as in Fig. 2.

its 27 Cys residues have not been specified (24). For such applications of bottom-up MS to distinguish S-H and S-S in glycosylated proteins (25), prior enzymatic or chemical treatment to remove glycans is necessary. By MS, Feng and Konishi measured a  $M_w$  of 196,863  $\pm$  29 (26), indicating 5.3% glycosylation of the predicted amino acid sequence  $M_w$  of 186,437. Our ESI mass spectrum gave no  $M_w$  data, presumably because of molecular-ion heterogeneity due to partial deglycosylation. Gentle PFD ( $T_{\text{cap}} = 115^\circ\text{C}$ ,  $V_{\text{pre}} = 50$ ,  $V_{\text{post}} = -4$ ) gave, unexpectedly, two major fragment ions of 20,838 daltons ( $b_{185}$  of the  $\beta$  chain) and 165,746  $\pm$  80 daltons, for a total of 186,584  $\pm$  80 daltons, indicating that the original  $\sim$ 7% glycosylation had been reduced to  $<$ 0.1% by PFD. Four spectra obtained with more strenuous activation ( $T_{\text{cap}} = 150^\circ$  to  $190^\circ\text{C}$ ,  $V_{\text{pre}} = 80$  to  $105$  V,  $V_{\text{post}} = 10$  to  $14$  V) identified 87 cleavages (Fig. 2, E to G) that yielded fragment ions without glycosylation.

These PFD data, simplified by the concomitant deglycosylation, can be used to assess possible disulfide bonding of the cysteine residues on the basis of two criteria. The presence of an S-S bond causes the experimental PFD mass to be 2 daltons less than the sequence-predicted value (27). Further, fragment ions are usually not observed from backbone cleavages between the Cys residues of an intrachain S-S bond, because this would require its additional dissociation. For the  $\beta$  chain with five cysteines (Fig. 2E), the masses of the  $b_{<49}$  ions are those predicted, whereas the masses of the  $b_{>78}$  ions are 2 daltons low, consistent with S-S bonding between the N-terminal Cys<sub>49</sub> and Cys<sub>78</sub> pair. The correspondingly low values for  $y_{>41}$  ions indicate the C terminus S-S bond from Cys<sub>41</sub> to Cys<sub>7</sub>. The remaining  $\beta$ -chain Cys that is 109 residues from the C terminus must provide the interchain S-S bond; consistent with this prediction, many bonds are cleaved in the backbone between the two proposed intrachain S-S bonds.

For the central  $\alpha$  chain (Fig. 2F), the  $b_{36}$  mass shows that Cys<sub>23</sub> and Cys<sub>24</sub> are unmodified, but the  $b_{56}$ ,  $b_{60}$ ,  $b_{86}$ , and  $b_{121}$  masses are 2 daltons less than expected, indicating S-S bond formation between two of the cysteines 37, 49, and 56. Similar evidence shows that the Cys that is 53 residues from the C terminus is unmodified, leaving the remaining cysteines 141, 197, and 442 as sites for previously identified S-S bonds to the outer  $\beta$  and  $\gamma$  chains (24). In the  $\gamma$  chain (Fig. 2G), eight fragment ions of  $b_{<69}$  have predicted mass values, indicating that Cys<sub>18</sub> is unmodified. However, the 17  $y$  ions of  $y_{\leq 174}$  have mass values 2 daltons less than that predicted, indicating a C-terminal S-S bond between Cys<sub>18</sub> and Cys<sub>3</sub> and -SH at the next seven cysteines. This result leaves only the N-terminal Cys<sub>82</sub> and Cys<sub>113</sub> as sites for the two interchain S-S bonds whose presence, but not locations, were found previously (24). These cysteines provide linkage to two of the proposed inter-

chain S-S sites of the  $\alpha$  chain, whose remaining site was linked to that of the  $\beta$  chain. For the fragment ion assignments shown in Fig. 2, E to G, the standard deviation of the mass difference in measured versus calculated values based on the S-S bonds postulated was  $\pm 0.14$  dalton. Of the 27 cysteines in the C4 glycoprotein, eight are in previously unidentified S-S bonds, modifying the deglycosylated  $M_w$  value to 186,429 (27).

In initial studies, the linear protein mycoerotic acid synthase (Mas) from the construct pQLAB-Mas (29) gave a measured  $M_w$  of 228,934  $\pm$  60, compared with the theoretical value of 229,067 (2154 residues). Five PFD spectra ( $T_{\text{cap}} = 150^\circ$  to  $190^\circ\text{C}$ ,  $V_{\text{pre}} = 70$  to  $125$  V,  $V_{\text{post}} = 4$  to  $15$  V) (fig. S1) gave 62 cleavages as far as 134 and 182 residues from the N and C termini, respectively, and the N-terminal cleavage data showed the absence of the predicted methionine (corrected theoretical  $M_w$  of 228,936 with 2153 residues).

For the identification of larger proteins by top-down MS (5), the limitation found here that primary  $b$ ,  $y$  ions are formed by dissociations in only the first 200 amino acids of each terminus ( $\sim$ 75% within 100 residues) has the advantage that the fragment ion mass values of the unknown sequences only have to be matched against masses of the correspondingly short terminal fragments of the DNA-predicted sequences, whereas the short peptides generated for bottom-up MS can be formed from any part of the protein. As a test of top-down identification of larger proteins (table S1), single PFD spectra of 1:1, 2:1, and 3:1 mixtures of PurL and  $\beta$ -Gal gave 11, 16, and 15, respectively, different mass values (1 to 10 kD, with an individual value often represented by multiple charge states) that matched those for primary backbone cleavages ( $b$ ,  $y$  ions) predicted for PurL. These mixtures also gave 17, 13, and 11  $b$ ,  $y$  ions (the largest had 84 residues) that matched those for  $\beta$ -galactoside, and all mass values fit those predicted with a standard deviation of 4.9 parts per million. These results suggest that reliable identifications (5, 12) for more complex mixtures of large proteins should be possible without initial proteolysis.

Top-down PFD characterization with our 6-T FTMS (12-T instruments now available) have far higher resolving power and sensitivity) can provide even  $\sim$ 70% sequence coverage on the first  $\sim$ 200 residues of each terminus of a large protein. This study indicates the further applicability of PFD to characterize stable posttranslational modifications such as methylation, acetylation, oxidation, and deamidation (27) in large proteins.

#### References and Notes

- N. L. Kelleher *et al.*, *J. Am. Chem. Soc.* **121**, 806 (1999).
- J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, C. M. Whitehouse, *Science* **246**, 64 (1989).
- K. D. Henry *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9075 (1989).
- F. W. McLafferty, E. K. Fridriksson, D. M. Horn, M. A. Lewis, R. A. Zubarev, *Science* **284**, 1289 (1999).

- N. L. Kelleher, *Anal. Chem.* **76**, 197A (2004).
- J. A. Loo, C. G. Edmonds, R. D. Smith, *Anal. Chem.* **63**, 2488 (1991).
- N. L. Kelleher *et al.*, *Protein Sci.* **7**, 1796 (1998).
- S. P. Gygi, R. Aebersold, *Curr. Opin. Chem. Biol.* **4**, 489 (2000).
- W. H. McDonald, J. R. Yates, *Dis. Markers* **18**, 99 (2002).
- W. J. Henzel, C. Watanabe, J. T. Stults, *J. Am. Soc. Mass Spectrom.* **14**, 931 (2003).
- Y. Ge *et al.*, *J. Am. Chem. Soc.* **124**, 672 (2002).
- D. M. Horn, R. A. Zubarev, F. W. McLafferty, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10313 (2000).
- C. S. Hoaglund-Hyzer, A. E. Counterman, D. E. Clemmer, *Chem. Rev.* **99**, 3037 (1999).
- K. Breuker, H.-B. Oh, D. M. Horn, B. A. Cerda, F. W. McLafferty, *J. Am. Chem. Soc.* **124**, 6407 (2002).
- H. Oh *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 15863 (2002).
- K. Breuker, F. W. McLafferty, *Angew. Chem. Int. Ed.* **44**, 4911 (2005).
- H. Zhai, X. Han, K. Breuker, F. W. McLafferty, *Anal. Chem.* **77**, 5777 (2005).
- J. A. Loo, C. G. Edmonds, R. D. Smith, *Science* **248**, 201 (1990).
- Unless stated otherwise, proteins were from Sigma-Aldrich (St. Louis, MO) and desalted with a Michrom BioResources (Auburn, CA) protein trap. Protein solutions were  $\sim$ 5  $\mu\text{M}$  in 50:45:5 MeCN:H<sub>2</sub>O:AcOH;  $\sim$ 1 s of ESI is necessary to fill the FTMS cell for measurement of each scan, with  $\sim$ 7  $\times$  10<sup>-13</sup> mol consumed for a 50-scan spectrum. Apparatus (Finnigan 6T FTMS) and experimental details were described recently (17). There, temperatures of the inlet capillary were measured at its flared entrance end just outside the heating current connection; here temperatures were measured in the stainless-steel body, with comparative values of 56°/58°, 62°/90°, 76°/150°, and 100°/190°C. Also as justified there (17), the predicted sequence is used to assign the observed masses to the N- and C-terminal products ( $b$  and  $y$  ions, respectively) and to internal ions ( $i_b$  and  $i_y$ ) of secondary  $b$ ,  $y$  dissociation. Excluding minor ions from H<sub>2</sub>O/NH<sub>3</sub> loss, this approach allowed  $>$ 95% assignment of observed mass values, although obviously the  $b$ ,  $y$  are more reliable than the  $i_b$  and  $i_y$  assignments. Singly charged fragment ions of low, but similar, relative intensities ( $\sim$ 1%; up to 10% at  $T_{\text{cap}} = 190^\circ\text{C}$ ,  $V_{\text{pre}} = 100$ ,  $V_{\text{post}} = 13$ ) appear at nearly every mass from 600 to 1500 daltons, even with ESI solution additives. The FTMS resolving power allows THRASH (20) to recognize and subtract these ions (17).
- D. M. Horn, R. A. Zubarev, F. W. McLafferty, *J. Am. Soc. Mass Spectrom.* **11**, 320 (2000).
- R. H. Jacobson, X.-J. Zhang, R. F. DuBose, B. W. Matthews, *Nature* **369**, 761 (1994).
- D. P. Little, J. P. Speir, M. W. Senko, P. B. O'Connor, F. W. McLafferty, *Anal. Chem.* **66**, 2809 (1994).
- R. Anand, A. A. Hoskins, J. Stubbe, S. E. Ealick, *Biochemistry* **43**, 10328 (2004).
- T. Seya, S. Nagasawa, J. Atkinson, *J. Immunol.* **136**, 4152 (1986).
- J. Qin, B. T. Chait, *Anal. Chem.* **69**, 4002 (1997).
- R. Feng, Y. Konishi, *Anal. Chem.* **64**, 2090 (1992).
- A similar top-down approach successfully identified five deamidation (-NH<sub>2</sub>  $\rightarrow$  -OH, +1 dalton) sites in ribonuclease A (13.7 kD) (28).
- V. Zabrouskov *et al.*, *Biochemistry* **45**, 987 (2006).
- K. C. Onwueme, C. J. Vos, J. Zurita, J. A. Ferreras, L. E. N. Quadri, *Prog. Lipid Res.* **44**, 259 (2005).
- We thank S. Ealick and L. Quadri for purified PurL and Mas. We also thank them and B. Baird, T. Begley, H. Zhai, and G. Infusini for helpful discussions. This work was supported by the Institute of General Medical Sciences (NIH grant GM16609 to F.W.M.) and the Austrian Fonds zur Förderung der Wissenschaftlichen Forschung and Tiroler Wissenschaftsfonds (grant T229 and UNI-0404/158 to K.B.).

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/314/5796/109/DC1  
Fig. S1  
Table S1

18 April 2006; accepted 16 August 2006  
10.1126/science.1128868



## Extending Top-Down Mass Spectrometry to Proteins with Masses Greater Than 200 Kilodaltons

Xuemei Han, *et al.*

*Science* **314**, 109 (2006);

DOI: 10.1126/science.1128868

**The following resources related to this article are available online at [www.sciencemag.org](http://www.sciencemag.org) (this information is current as of March 3, 2008):**

**Updated information and services**, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/cgi/content/full/314/5796/109>

**Supporting Online Material** can be found at:

<http://www.sciencemag.org/cgi/content/full/314/5796/109/DC1>

A list of selected additional articles on the Science Web sites **related to this article** can be found at:

<http://www.sciencemag.org/cgi/content/full/314/5796/109#related-content>

This article **cites 27 articles**, 8 of which can be accessed for free:

<http://www.sciencemag.org/cgi/content/full/314/5796/109#otherarticles>

This article has been **cited by** 23 article(s) on the ISI Web of Science.

This article has been **cited by** 1 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/cgi/content/full/314/5796/109#otherarticles>

This article appears in the following **subject collections**:

Chemistry

<http://www.sciencemag.org/cgi/collection/chemistry>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>