# Peptide and protein *de novo* sequencing by mass spectrometry

Kenneth G Standing

Although the advent of large-scale genomic sequencing has greatly simplified the task of determining the primary structures of peptides and proteins, the genomic sequences of many organisms are still unknown. Even for those that are known, modifications such as post-translational events may prevent the identification of all or part of the protein sequence. Thus, complete characterization of the protein primary structure often requires determination of the protein sequence by mass spectrometry with minimal assistance from genomic data — *de novo* protein sequencing. This task has been facilitated by technical developments during the past few years: 'soft' ionization techniques, new forms of chemical modification (derivatization), new types of mass spectrometer and improved software.

**Addresses**
Department of Physics and Astronomy, University of Manitoba, Winnipeg, MB R3T 2N2, Canada
e-mail: standin@cc.umanitoba.ca

**Current Opinion in Structural Biology** 2003, **13**:595–601

This review comes from a themed issue on
Biophysical methods
Edited by Brian T Chait and Keith Moffat

0959-440X/$ – see front matter


**DOI 10.1016/j.sbi.2003.09.005**

**Abbreviations**

| | |
|---|---|
| **CID** | collision-induced dissociation |
| **ECD** | electron capture dissociation |
| **ESI** | electrospray ionization |
| **FTICR** | Fourier transform ion cyclotron resonance |
| **MALDI** | matrix-assisted laser desorption/ionization |
| **MS** | mass spectrometry |
| **MS/MS** | tandem mass spectrometry |
| ***m/z*** | mass/charge |
| **QqTOF** | quadrupole/time-of-flight |
| **TOF** | time-of-flight |

## Introduction

The advent of large-scale genomic sequencing has greatly simplified the task of determining the primary structures of peptides and proteins in many organisms, because open reading frames in the nucleotide sequence serve as templates for the construction of the corresponding proteins. The masses of peptides produced by proteolytic digestion of an unknown protein can be compared with those predicted to arise from each protein in the database; this 'mass mapping' is often sufficient to identify any protein whose full-length sequence is contained therein [1–3,4••].

Nevertheless, the genome sequences of most organisms are still unknown. Even for those that are known, modifications such as post-translational events may prevent the identification of all or part of the protein sequence, or at least the definition of the modifications. Thus, complete characterization of the protein primary structure often requires determination of the protein sequence with minimal assistance from genomic data — *de novo* protein sequencing. Early *de novo* protein sequencing measurements relied on Edman degradation of the protein [4••], but mass spectrometry (MS) has reduced the need for this technique because it is more sensitive and provides higher sample throughput. It can also cope better with protein mixtures and with modifications to the protein N terminus.

*De novo* protein sequencing by MS dates back more than 30 years [5,6,7•], first in combination with Edman degradation and then on its own. Early measurements were mostly made on sector mass spectrometers or triple quadrupoles, and with electron ionization, which almost completely fragmented the proteins (normally involatile and labile compounds), unless they were chemically modified (derivatized) [8]. However, more recent determinations have been greatly facilitated by subsequent technical developments (see [1–3,4••]). First, 'soft' ionization techniques have been developed, particularly electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI), which produce ions from peptides and proteins with much less fragmentation. Second, workers now have available other types of mass spectrometer, such as ion traps, time-of-flight (TOF) devices and Fourier transform ion cyclotron resonance (FTICR) instruments, each of which has advantages for sequencing measurements. Third, more efficient derivatization reactions have been developed. Finally, there have been great improvements in software to enable automated descrambling of mass spectra by computer, thus reducing the demands on manual interpretation of the data.

These developments in *de novo* peptide and protein sequencing by MS are reviewed in the following sections.

## Peptide sequencing

Straightforward MS measurement of the mass spectrum of the collection of proteolytic peptides resulting from enzymatic digestion of an unknown protein is often sufficient for protein identification, as noted above. Along with a MALDI measurement of the overall protein mass, it is also the usual first step in *de novo* sequencing of the protein, yielding the masses of the individual peptides.

The next step is to determine the sequences of the peptides. The MS measurement itself may yield considerable information about the amino acid compositions of these peptides, especially when it includes accurate measurements of low mass ions, such as immonium ions [6,9–11]. However, it cannot determine the order of the amino acid residues. Consequently, *de novo* peptide sequencing requires tandem mass spectrometry, usually denoted MS/MS [12]. In this technique, a given parent (precursor) ion is selected in one mass spectrometer and then broken up, usually by collisions [13•]. The *m/z* (mass/charge) values for the resulting daughter (product) ions are measured in a second mass spectrometer. Under favorable conditions, this procedure may yield a series of ions that contains sufficient information to determine the peptide sequence. Even more information can be obtained in certain instruments, ion traps for example, by breaking up the daughter ions themselves to yield a spectrum of granddaughter ions.

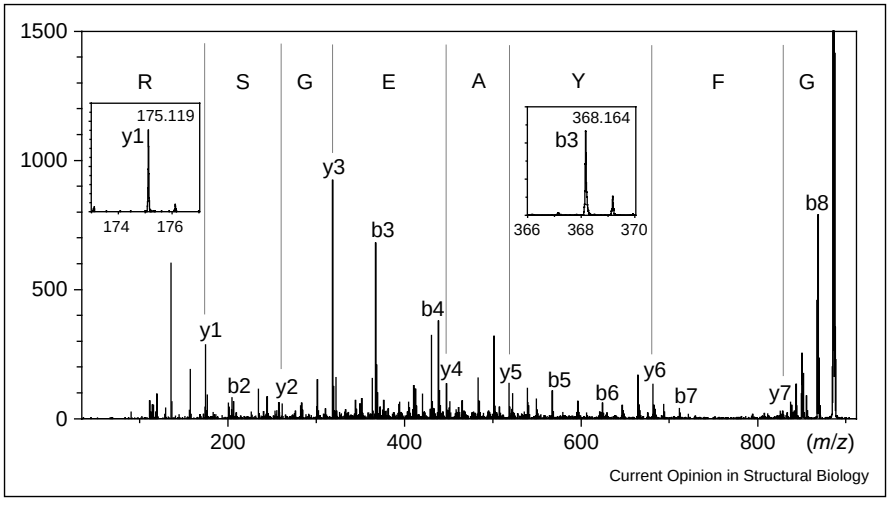
For underivatized and unlabeled peptides, ESI *de novo* sequencing measurements have been carried out following collision-induced dissociation (CID) in the triple quadrupole mass analyzer [14,15,16•] or in the quadrupole/TOF (QqTOF) instrument [15,16•,17–22], and also after resonant excitation in the quadrupole ion trap [23–25]. MALDI *de novo* sequencing has been carried out using CID and metastable decay in single- [26] and double-stage [27–29,30••] TOF instruments, using CID in QqTOF devices [31•,32,33,34••,35] and using resonant excitation/CID in ion traps [36••].

CID fragmentation mechanisms are extensively discussed in [12,13•,37•,38–45]. Extraction of the peptide sequence from the daughter ion spectra depends on the completeness of the observed series, the accuracy with which the mass differences between the relevant ion peaks are determined and the extent to which the fragmentation spectrum can be correctly interpreted. A peptide resulting from tryptic digestion normally has a basic residue (arginine or lysine) at its C terminus and it yields a prominent doubly charged ion peak when ionized by ESI. If this ion is chosen as the parent ion for an MS/MS measurement, the production of a series of y ion daughters (ions resulting from cleavage at the amide bonds and containing the C terminus) is favored and the resulting spectrum is likely to be easy to interpret [12,13•,14,15,16•,17–25,37•,38–42]. For this reason, ESI has been a popular choice for *de novo* peptide sequencing.

On the other hand, CID of MALDI ions (predominantly singly charged) tends to yield daughter ion spectra that include both y and b ions, the latter corresponding to cleavage at the peptide bond and containing the N terminus [26–29,30••,31•,32,33,34••,35,36••]. They also may be dominated by a few preferred fragmentation pathways [31•,43–45], particularly in ion traps. However, the presence of both b and y ions means that the spectra are rich in information, provided that the spectra can be interpreted. In this regard, sensitivity and accuracy of the instrument used for mass analysis are particularly important. A recent *m/z* spectrum obtained in our laboratory on a QqTOF instrument is shown in Figure 1 and compared with calculated mass values in Table 1. The good agreement between calculated and observed masses lends confidence to the assignments.

MALDI ionization has several attractive properties for the analysis of peptides. In particular, it has a relatively high tolerance to impurities and common biochemical

**Figure 1**




MS/MS *de novo* sequencing of the 886.407 Da (residues 170–177) GFYAEGSR peptide from SARS virus nucleocapsid protein. Series of y and b fragments are labeled. (Note that 886.407 was the measured mass during the HPLC MS run, but the calculated $MH^+$ is 886.406 Da.)

**Table 1**

**Calculated and measured masses of y and b series of fragments from MS/MS spectra of the 886.407 Da (residues 170–177) GFYAEGSR peptide from SARS virus nucleocapsid protein.**

| b fragments | | Residue | y fragments | |
|---|---|---|---|---|
| [MH$^+$] calculated | *m/z* (observed) | | [MH$^+$] calculated | *m/z* (observed) |
| 58.029 | – | G | 886.406 | 886.407 |
| 205.098 | 205.098 | F | 829.384 | 829.386 |
| 368.161 | 368.164 | Y | 682.316 | 682.318 |
| 439.198 | 439.204 | A | 519.253 | 519.247 |
| 568.241 | 568.247 | E | 448.216 | 448.217 |
| 625.262 | 625.270 | G | 319.173 | 319.177 |
| 712.294 | 712.297 | S | 262.152 | 262.152 |
| 868.395 | 868.398 | R | 175.120 | 175.119 |

additives and salts, reducing the demands on sample purification compared with ESI. Also, the MALDI sample is usable until it is depleted, so data can be collected from many different ion species produced from a single peptide mixture and until the desired quality of spectra has been obtained. These qualities have stimulated recent interest in MALDI sequencing, particularly after pretreatment of the sample by derivatization or isotopic labeling, as described below.

## Derivatization and labeling

As mentioned above, the use of electron ionization required chemical modification (derivatization) of the peptide to produce an informative mass spectrum. Although this is no longer an absolute requirement, the procedure may still yield marked improvement in mass spectral quality. Derivatization methods used until the late 1990s are reviewed in [8], but other techniques have been developed more recently [25,46••,47–55]. A notable example is sulfonation of the peptide N terminus, originated by Keough and his collaborators [46••,47–53]. This modification usually leads to the production of a series of C-terminal y daughter ions from singly charged parents and thus transfers the advantages of ESI to MALDI. Moreover, the required reagents can be obtained conveniently as a commercially available kit (the CAF MALDI sequencing kit from Amersham Bioscience). The main disadvantage of the method is a reduction in sensitivity by a factor ~10.

An alternative technique calls for isotopic labeling [56–63,64•,65–72], which helps to distinguish C-terminal from N-terminal ions, and also may provide information on quantitation. Particularly noteworthy is the use of labeling by $^{18}O$ [60–63,64•,65–72], usually carried out by doing the proteolytic digestion in a mixture of ordinary water and $H_2{}^{18}O$. Under these conditions, the C-terminal residue of each proteolytic fragment (except the one containing the C terminus of the protein) contains a mixture of $^{16}O$- and $^{18}O$-containing ions. Consequently, the C-terminal daughter ions alone are identified by this label and the N-terminal ions are unlabeled. A major advantage of the technique is that the normal digestion protocol is maintained and no sensitivity is lost. In addition to possible ambiguities arising from incorporation of either one or two $^{18}O$ atoms [69], the main disadvantage is the high cost of the $H_2{}^{18}O$.

## Mass accuracy

The accuracy with which *m/z* values are determined has a pronounced effect on the reliability of sequence assignments [9]. The 20 common amino acid residues have distinctive elemental compositions and consequently have distinctive masses, except for the L/I pair. However, a low accuracy measurement may be incapable of discriminating between D and N, or between E and Q/K (cases where $\Delta m \sim 1$ Da), and a much higher accuracy is needed to distinguish Q from K, where $\Delta m = 36$ mDa. In addition, there are combinations of amino acid residues that yield the same mass number [6] or even the same elemental composition [e.g. (G+G) = N = $C_4H_6N_2O_2$; (G+A) = Q = $C_5H_8N_2O_2$], and these may create ambiguities if the intermediate daughter ion is not seen. When modified peptides are considered, there are many additional possibilities.

However, there are several instrument types that are sufficiently accurate to resolve some of these uncertainties. Recent examples of success in QqTOF instruments include K versus Q, and VV versus PT (both with $\Delta m = 36$ mDa) [36••], as well as R–G versus V ($\Delta m = 11.2$ mDa) [33]. Phosphorylation has also been distinguished from sulfation ($\Delta m = 9.5$ mDa) by FTICR [73•,74].

## From peptide to protein sequence

The peptide sequences provide the raw material for determining the overall protein sequence. The problem is then one of ordering the individual peptide sequences. A purely MS method is to search for proteolytic fragments that span the initial ones. These may arise in the initial digestion (usually by trypsin) because of missed cleavages

**Table 2**

**Proteolytic fragments from the 'high plains virus' protein @ALSFKNSSGVLKAKTLKDGFVTSSDIETTVHDFSYEKPDLSSVDGFSLKS...***

| Enzyme | *m/z* (observed) | $MH^+$ (calculated) | Δm (mDa) | Fragment sequence |
|---|---|---|---|---|
| Trypsin | 3522.653 | 3522.644 | +9 | DGFVTSSDIETTVHDFSYEKPDLSSVDGFSLK |
| Asp-N | 3322.769 | 3322.753 | +16 | @ALSFKNSSGVLKAKTLKDGFVTSSDIETTVH |
| Glu-C | 2884.534 | 2884.531 | +3 | @ALSFKNSSGVLKAKTLKDGFVTSSDIE |
| Asp-N | 1834.080 | 1834.081 | –1 | @ALSFKNSSGVLKAKTLK |
| Trypsin | 1491.856 | 1491.854 | +2 | @ALSFKNSSGVLKAK |
| Trypsin | 1292.723 | 1292.722 | +1 | @ALSFKNSSGVLK |
| Trypsin | 607.350 | 607.346 | +4 | @ALSFK |

* The N-terminal portion of the original U60141 sequence is shaded. @A denotes acetylated alanine.

or they may be produced by digestion with other proteases. A simple example is shown in Table 2; a 14-residue peptide appeared, after tryptic digestion, in the same gel spot as peptides from a protein identified in the database as 'the putative protein of high plains virus'. Additional digestions then showed that this peptide was contained in an additional 18 amino acid residues that formed the actual N terminus of the putative protein.

The complete protein sequence can only be obtained by this method if the peptide coverage is also complete. This is difficult to obtain in a single digestion, but digestion by several different proteases is likely to improve the coverage. A useful check on the completeness of the coverage is provided by a MALDI measurement of the overall protein mass.

A combination of techniques may provide the complete protein sequence even if only partial peptide sequence information is available, as a single peptide sequence enables construction of a short oligonucleotide useful for the isolation of the gene that encodes the protein of interest. However, such cloning is most efficiently carried out if the peptide sequence is long, 100% accurate and encodes low degeneracy primers, that is, if it contains amino acids encoded by only one base triplet (Met, Trp) or two (Phe, Tyr, Gln, Glu, Asn, Asp, Cys); these requirements may not always be easy to fulfill.

## 'Top-down' sequencing

A recently developed method that circumvents some of the above difficulties is 'top-down' sequencing, pioneered by McLafferty and his collaborators [75–81]. The technique has been most frequently applied to measurements on FTICR mass spectrometers, because they provide very high resolution and mass accuracy. In this method, the ionized protein itself is introduced into the ICR cell. *M/z* measurements are then carried out in the same spectrometer both on the intact protein and on the products of its dissociation, thus avoiding the need to stitch together the products of peptide sequencing. In addition, the FTICR measurements are able to make use of various methods to manipulate the ions in the ICR cell, including CID and IR multiphoton dissociation. Most importantly, the ions can be broken up by the newly discovered electron capture dissociation (ECD) technique [76–82], which has only been available in FTICR instruments up to now. ECD induces far more general backbone cleavage than other methods, yielding extensive sequence information on proteins as large as approximately 40 kDa.

Nevertheless, although the fractional mass accuracy of the FTICR instrument is impressive, the measurements are made on very large ions, so the absolute mass accuracy, which determines the ability to distinguish between various sequence alternatives, is somewhat limited. This is one argument for using a combination of top-down sequencing with the usual 'bottom-up' method [83].

## Data interpretation and analysis

Several computer algorithms have been devised that seek to infer sequence *de novo* from MS/MS data [84–91,92••,93–95]. This work, together with methods that look for homology of the unknown protein with sequences of organisms that are already in the database, has been recently reviewed by Liska and Shevchenko [91,92••]. Clearly these techniques have great potential for improving *de novo* peptide and protein sequencing by MS.

## Conclusions

The recent developments summarized above have produced considerable improvements in our ability to carry out *de novo* peptide and protein sequencing by MS, although it is still not an easy task unless nature has designed the sequence to be particularly easy to unravel. Although mere protein identification will become more straightforward as more genomes are sequenced, complete protein characterization will continue to require *de novo* techniques; this will probably become more important as long lists of identifications are found to be inadequate for determining protein function. Fortunately, there is still room for ample improvement in all the methods described above. Thus, *de novo* peptide and protein sequencing by MS is likely to have a bright future.

## Acknowledgements

I thank Brian Chait and Andrej Shevchenko for helpful discussions and communication of their results before publication. I also thank Oleg Krokhin for supplying the data shown in Figure 1 and Table 1. Gideon Garland, James McNabb, Yi-min She and Victor Spicer provided invaluable technical support. This work was supported by grants from the US National Institutes of Health (GM 59240), and from the Natural Sciences and Engineering Research Council of Canada.

## References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Mann M, Hendrickson RC, Pandey A: **Analysis of proteins and proteomes by mass spectrometry**. *Annu Rev Biochem* 2001, **70**:437-473.

2. Aebersold R, Mann M: **Mass spectrometry-based proteomics**. *Nature* 2003, **422**:198-207.

3. Liebler DC: *Introduction to Proteomics: Tools for the New Biology*. Totowa, New Jersey: Humana Press; 2002.

4. •• Simpson RJ: *Proteins and Proteomics: a Laboratory Manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 2003.

A comprehensive and up-to-date account of the title subject, including many protocols. About half the volume is devoted to MS, although *de novo* sequencing is not specifically discussed. However, there are chapters on related techniques, such as gel electrophoresis, liquid chromatography and sample preparation, including proteolytic digestions, as well as one on Edman degradation.

5. Biemann K, Martin SA: **Mass spectrometric determination of the amino acid sequence of peptides and proteins**. *Mass Spectrom Rev* 1987, **6**:1-76.

6. Papayannopoulos IA: **The interpretation of collision-induced dissociation tandem mass spectrometry of peptides**. *Mass Spectrom Rev* 1995, **14**:49-73.

7. • Biemann K: **Four decades of structure determination by mass spectrometry: from alkaloids to heparin**. *J Am Soc Mass Spectrom* 2002, **13**:1254-1272.

An interesting account of the early development of protein sequencing, by the pioneer.

8. Roth KD, Huang ZH, Sadagopan N, Watson JT: **Charge derivatization of peptides for analysis by mass spectrometry**. *Mass Spectrom Rev* 1998, **17**:255-274.

9. Clauser KR, Baker P, Burlingame AL: **Role of accurate mass measurement (+/−10 ppm) in protein identification strategies employing MS or MS/MS and database searching**. *Anal Chem* 1999, **71**:2871-2882.

10. Nielsen ML, Bennett KL, Larsen B, Moniatte M, Mann M: **Peptide end sequencing by orthogonal MALDI tandem mass spectrometry**. *J Proteome Res* 2002, **1**:63-71.

11. Schlosser A, Lehmann WD: **Patchwork peptide sequencing: extraction of sequence information from accurate mass data of peptide tandem mass spectra recorded at high resolution**. *Proteomics* 2002, **2**:524-533.

12. Kinter M, Sherman NE: *Protein Sequencing and Identification Using Tandem Mass Spectrometry*. New York: Wiley-Interscience; 2000.

13. • Laskin J, Futrell JH: **Collisional activation of peptide ions in FT-ICR mass spectrometry**. *Mass Spectrom Rev* 2003, **22**:158-181.

A comprehensive review of the subject, including comparisons with theory. The reported measurements were made on an FTICR instrument, but most of the results are generally applicable.

14. Wilm M, Shevchenko A, Houthaeve T, Breit S, Schweigerer L, Fotsis T, Mann M: **Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry**. *Nature* 1996, **379**:466-469.

15. Shevchenko A, Chernushevich I, Wilm M, Mann M: ***De novo* peptide sequencing by nanoelectrospray tandem mass spectrometry using triple quadrupole and quadrupole/time-of-flight instruments**. *Methods Mol Biol* 2000, **146**:1-16.

16. • Shevchenko A, Chernushevich I, Shevchenko A, Wilm M, Mann M: ***De novo* sequencing of peptides recovered from in-gel digested proteins by nanoelectrospray tandem mass spectrometry**. *Mol Biotechnol* 2002, **20**:107-118.

A detailed discussion of *de novo* sequencing by ESI.

17. Morris HR, Paxton T, Panico M, McDowell R, Dell A: **A novel geometry mass spectrometer, the Q-TOF, for low-femtomole/attomole-range biopolymer sequencing**. *J Protein Chem* 1997, **16**:469-479.

18. Morris HR, Paxton T, Dell A, Langhorne J, Berg M, Bordoli RS, Hoyes J, Bateman RH: **High sensitivity collisionally-activated decomposition tandem mass spectrometry on a novel quadrupole/orthogonal-acceleration time-of-flight mass spectrometer**. *Rapid Commun Mass Spectrom* 1996, **10**:889-896.

19. van Der Wel H, Morris HR, Panico M, Paxton T, North SJ, Dell A, Thomson JM, West CM: **A non-Golgi alpha 1,2-fucosyltransferase that modifies Skp1 in the cytoplasm of Dictyostelium**. *J Biol Chem* 2001, **276**:33952-33963.

20. Romaris F, North SJ, Gagliardo LF, Butcher BA, Ghosh K, Beiting DP, Panic M: **A putative serine protease among the excretory-secretory glycoproteins of L1 *Trichinella spiralis***. *Mol Biochem Parasitol* 2002, **122**:149-160.

21. Vandenberghe I, Kim J-K, Devreese B, Hacisalihoglu A, Iwabuki H, Okajima T, Kuroda S, Adachi O, Jongejan JA, Duine JA *et al.*: **The covalent structure of the small subunit from *Pseudomonas putida* amine dehydrogenase reveals the presence of three novel types of internal cross-linkages, all involving cysteine in a thioether bond**. *J Biol Chem* 2001, **276**:42923-42931.

22. Ou K, Seow TK, Liang RC, Ong SE, Chung MC: **Proteome analysis of a human heptocellular carcinoma cell line, HCC-M: an update**. *Electrophoresis* 2001, **22**:2804-2811.

23. Arnott D, Henzel WJ, Stults JT: **Rapid identification of comigrating gel-isolated proteins by ion trap-mass spectrometry**. *Electrophoresis* 1998, **19**:968-980.

24. Zhang Z, McElvain JS: ***De novo* peptide sequencing by two-dimensional fragment correlation mass spectrometry**. *Anal Chem* 2000, **72**:2337-2350.

25. Sonsmann G, Romer A, Schomburg D: **Investigation of the influence of charge derivatization on the fragmentation of multiply protonated peptides**. *J Am Soc Mass Spectrom* 2002, **13**:47-58.

26. Schilling B, Wang W, McMurray JS, Medzihradszky KF: **Fragmentation and sequencing of cyclic peptides by matrix-assisted laser desorption/ionization post-source decay mass spectrometry**. *Rapid Commun Mass Spectrom* 1999, **13**:2174-2179.

27. Medzihradszky KF, Campbell JM, Baldwin MA, Falick AM, Juhasz P, Vestal ML, Burlingame AM: **The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer**. *Anal Chem* 2000, **72**:552-558.

28. Bienvenut WV, Deon C, Pasquarello C, Campbell JM, Sanchez J-C, Vestal ML, Hochstrasser DF: **Matrix-assisted laser desorption/ionization-tandem mass spectrometry with high resolution and sensitivity for identification and characterization of proteins**. *Proteomics* 2002, **2**:868-876.

29. Yergey AL, Coorssen JR, Backlund PS, Blank PS, Humphrey GA, Zimmerberg J, Campbell J, Vestal ML: ***De novo* sequencing of peptides using MALDI/TOF-TOF**. *J Am Soc Mass Spectrom* 2002, **13**:784-791.

30. •• Juhasz P, Campbell JM, Vestal ML. **MALDI-TOF/TOF technology for peptide sequencing and protein identification**. In *Mass Spectrometry and Hyphenated Techniques in Neuropeptide Research*. Edited by Silberring J, Ekman R. New York: Wiley; 2002:375-413.

A comprehensive discussion of the use of the new TOF/TOF instruments.

31. • Wattenberg A, Organ AJ, Schneider K, Tyldesley R, Bordoli R, Bateman RH: **Sequence dependent fragmentation of peptides**

**generated by MALDI quadrupole time-of-flight (MALDI Q-TOF) mass spectrometry and its implications for protein identification**. *J Am Soc Mass Spectrom* 2002, **13**:772-783.
A detailed discussion of peptide fragmentation in MALDI measurements.

32. She Y-M, Wang G-Q, Loboda A, Ens W, Standing KG, Burczynski FJ: **Sequencing of rat liver cytosolic proteins by matrix-assisted laser desorption ionization-quadrupole time-of-flight mass spectrometry following electrophoretic separation and extraction**. *Anal Biochem* 2002, **310**:137-147.

33. She Y-M, Haber S, Seifers DL, Loboda A, Chernushevich I, Perreault H, Ens W, Standing KG: **Determination of the complete amino acid sequence for the coat protein of brome mosaic virus by time-of-flight mass spectrometry: evidence for mutations associated with change of propagation host**. *J Biol Chem* 2001, **276**:20039-20047.

34. •• Shevchenko A, Sunyaev S, Loboda A, Shevchenko A, Bork P, Ens W, Standing KG: **Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching**. *Anal Chem* 2001, **73**:1917-1926.
One of the first demonstrations of the use of BLAST searches to obtain *de novo* sequence information.

35. Cramer R, Corless S: **The nature of collision-induced dissociation processes of doubly protonated peptides: comparative study for the future use of matrix-assisted laser desorption/ionization on a hybrid quadrupole time-of-flight mass spectrometer in proteomics**. *Rapid Commun Mass Spectrom* 2001, **15**:2058-2066.

36. •• Zhang W, Krutchinsky AN, Chait BT: ***De novo* peptide sequencing by MALDI-quadrupole-ion trap mass spectrometry: a preliminary study**. *J Am Soc Mass Spectrom* 2003, **14**:1012-1021.
A demonstration that ion traps can yield significant *de novo* sequence information.

37. • Tabb DL, Smith LL, Breci LA, Wysocki VH, Lin D, Yates JR III: **Statistical characterization of ion trap tandem mass spectra from doubly charged tryptic peptides**. *Anal Chem* 2003, **75**:1155-1163.
A summary of recent results on the title subject.

38. Loo JA, Edmonds CG, Smith RD: **Tandem mass spectrometry of very large molecules. 2. Dissociation of multiply charged proline-containing proteins from electrospray ionization**. *Anal Chem* 1993, **65**:425-438.

39. Tsaprailis G, Nair H, Somogyi A, Wysocki VH, Zhong W, Futrell JH, Summerfield SG, Gaskell SJ: **Influence of secondary structure on the fragmentation of protonated peptides**. *J Am Chem Soc* 1999, **121**:5142-5154.

40. Breci LA, Tabb DL, Yates JR III, Wysocki VH: **Cleavage N-terminal to proline: analysis of a database of peptide tandem mass spectra**. *Anal Chem* 2003, **75**:1963-1971.

41. Wysocki VH, Tsaprailis G, Smith LL, Breci LA: **Mobile and localized protons: a framework for understanding peptide dissociation**. *J Mass Spectrom* 2000, **35**:1399-1406.

42. Gu C, Tsaprailis G, Breci L, Wysocki VH: **Selective gas-phase cleavage at the peptide bond C-terminal to aspartic acid in fixed-charge derivatives of Asp-containing peptides**. *Anal Chem* 2000, **72**:5804-5813.

43. Yu W, Vath JE, Huberty MC, Martin SA: **Identification of the facile gas-phase cleavage of the Asp-Pro and Asp-Xxx peptide bonds in matrix-assisted laser desorption time-of-flight mass spectrometry**. *Anal Chem* 1993, **65**:3015-3023.

44. Qin J, Chait BT: **Preferential fragmentation of protonated gas-phase peptide Ions adjacent to acidic amino acid residues**. *J Am Chem Soc* 1995, **117**:5411-5412.

45. Qin J, Chait BT: **Collision-induced dissociation of singly charged peptide ions in a matrix-assisted laser desorption ionization ion trap mass spectrometer**. *Int J Mass Spectrom* 1999, **190**:313-320.

46. •• Keough T, Youngquist RS, Lacey MP: **Sulfonic acid derivatives for peptide sequencing by MALDI MS**. *Anal Chem* 2003, **75**:157A-165A.
An up-to-date review of the technique of sulfonation.

47. Keough T, Lacey MP, Youngquist RS: **Solid-phase derivatization of tryptic peptides for rapid protein identification by matrix-assisted laser desorption/ionization mass spectrometry**. *Rapid Commun Mass Spectrom* 2002, **16**:1003-1015.

48. Keough T, Lacey MP, Strife RJ: **Atmospheric pressure matrix-assisted laser desorption/ionization ion trap mass spectrometry of sulfonic acid derivatized tryptic peptides**. *Rapid Commun Mass Spectrom* 2001, **15**:2227-2239.

49. Keough T, Lacey MP, Youngquist RS: **Derivatization procedures to facilitate *de novo* sequencing of lysine-terminated tryptic peptides using postsource decay matrix-assisted laser desorption/ionization mass spectrometry**. *Rapid Commun Mass Spectrom* 2000, **14**:2348-2356.

50. Keough T, Lacey MP, Fieno AM, Grant RA, Sun Y, Bauer MD, Begley KB: **Tandem mass spectrometry methods for definitive protein identification in proteomics research**. *Electrophoresis* 2000, **21**:2252-2265.

51. Bauer MD, Sun Y, Keough T, Lacey MP: **Sequencing of sulfonic acid derivatized peptides by electrospray mass spectrometry**. *Rapid Commun Mass Spectrom* 2000, **14**:924-929.

52. Keough T, Youngquist RS, Lacey MP: **A method for high-sensitivity peptide sequencing using postsource decay matrix-assisted laser desorption ionization mass spectrometry**. *Proc Natl Acad Sci USA* 1999, **96**:7131-7136.

53. Hellman U, Bhikhabhai R: **Easy amino acid sequencing of sulfonated peptides using post-source decay on a matrix-assisted laser desorption/ionization time-of-flight mass spectrometer equipped with a variable voltage reflector**. *Rapid Commun Mass Spectrom* 2002, **16**:1851-1859.

54. Shen TL, Huang ZH, Laivenieks M, Zeikus JG, Gage DA, Allison J: **Evaluation of charge derivatization of a proteolytic protein digest for improved mass spectrometric analysis: *de novo* sequencing by matrix-assisted laser desorption/ionization post-source decay mass spectrometry**. *J Mass Spectrom* 1999, **34**:1154-1165.

55. Lindh I, Hjelmqvist L, Bergman T, Sjovall J, Griffiths WJ: ***De novo* sequencing of proteolytic peptides by a combination of C-terminal derivatization and nano-electrospray/collision-induced dissociation mass spectrometry**. *J Am Soc Mass Spectrom* 2000, **11**:673-686.

56. Muenchbach M, Quadroni M, Miotto G, James P: **Quantitation and facilitated *de novo* sequencing of proteins by isotopic N-terminal labeling of peptides with a fragmentation-directing moiety**. *Anal Chem* 2000, **72**:4047-4057.

57. Cagney G, Emili A: ***De novo* peptide sequencing and quantitative profiling of complex protein mixtures using mass-coded abundance tagging**. *Nat Biotechnol* 2002, **20**:163-170.

58. Gu S, Pan S, Bradbury EM, Chen X: **Use of deuterium-labeled lysine for efficient protein identification and peptide *de novo* sequencing**. *Anal Chem* 2002, **74**:5774-5785.

59. Gu S, Pan S, Bradbury EM, Chen X: **Precise peptide sequencing and protein quantification in the human proteome through *in vivo* lysine-specific mass tagging**. *J Am Soc Mass Spectrom* 2003, **14**:1-7.

60. Shevchenko A, Chernushevich I, Ens W, Standing KG, Thomson B, Wilm M, Mann M: **Rapid '*de novo*' peptide sequencing by a combination of nanoelectrospray, isotopic labeling and a quadrupole/time-of-flight mass spectrometer**. *Rapid Commun Mass Spectrom* 1997, **11**:1015-1024.

61. Qin J, Herring CJ, Zhang X: ***De novo* peptide sequencing in an ion trap mass spectrometer with $^{18}O$ labeling**. *Rapid Commun Mass Spectrom* 1998, **12**:209-216.

62. Wilm M, Neubauer G, Taylor L, Shevchenko A, Bachi A. ***De novo* sequencing of proteins with mass spectrometry using the differential scanning technique**. In *Proteome and Protein Analysis*. Edited by Kamp RM, Kyriakidis D, Choli-Papadopoulou T. Springer; 2000:65-79.

63. Uttenweiler-Joseph S, Neubauer G, Christoforidis S, Zerial M, Wilm M: **Automated *de novo* sequencing of proteins using the differential scanning technique**. *Proteomics* 2001, **1**:668-682.

64. Krokhin O, Li Y, Antonov A, Feldmann H, Flick R, Jones S,
• Stroeher U, Bastien N, Dasuri KVN, Cheng K *et al.*: **Mass spectrometric characterization of proteins from the SARS virus: a preliminary report**. *Mol Cell Proteomics* 2003, **2**:346-356.
Use of the $^{18}O$ tagging method to produce an almost complete set of peptides from an unknown protein.

65. Yao X, Afonso C, Fenselau C: **Dissection of proteolytic 18O labeling: endoprotease-catalyzed 16O-to-18O exchange of truncated peptide substrates**. *J Proteome Res* 2003, **2**:147-152.

66. Liu P, Regnier FE: **An isotope coding strategy for proteomics involving both amine and carboxyl group labeling**. *J Proteome Res* 2002, **1**:443-450.

67. Reynolds KJ, Yao X, Fenselau C: **Proteolytic 18O labeling for comparative proteomics: evaluation of endoprotease Glu-C as the catalytic agent**. *J Proteome Res* 2002, **1**:27-33.

68. Back JW, Notenboom V, de Koning LJ, Muijsers AO, Sixma TK, de Koster CG, de Jong L: **Identification of cross-linked peptides for protein interaction studies using mass spectrometry and 18O labeling**. *Anal Chem* 2002, **74**:4417-4422.

69. Stewart II, Thomson T, Figeys D: **18O labeling: a tool for proteomics**. *Rapid Commun Mass Spectrom* 2001, **15**:2456-2465.

70. Wang YK, Ma Z, Quinn DF, Fu EW: **Inverse 18O labeling mass spectrometry for the rapid identification of marker/target proteins**. *Anal Chem* 2001, **73**:3742-3750.

71. Yao X, Freas A, Ramirez J, Demirev PA, Fenselau C: **Proteolytic 18O labeling for comparative proteomics: model studies with two serotypes of adenovirus**. *Anal Chem* 2001, **73**:2836-2842.

72. Heller M, Mattou H, Menzel C, Yao X: **Trypsin catalyzed (16)O-to-(18)O exchange for comparative proteomics: tandem mass spectrometry comparison using MALDI-TOF, ESI-QTOF, and ESI-ion trap mass spectrometers**. *J Am Soc Mass Spectrom* 2003, **14**:704-718.

73. Marshall AG, Hendrickson CL, Stone D, Shi H: **Scaling MS**
• **plateaus with high resolution FT-ICRMS**. *Anal Chem* 2002, **74**:253A-259A.
An excellent discussion of the opportunities provided by the increase in mass accuracy of MS measurements, with particular attention to FTICR techniques.

74. Bossio RE, Marshall AG: **Baseline resolution of isobaric phosphorylated and sulfated peptides and nucleotides by electrospray ionization FTICR ms: another step toward mass spectrometry-based proteomics**. *Anal Chem* 2002, **74**:1674-1679.

75. Horn DM, Zubarev RA, McLafferty FW: **Automated *de novo* sequencing of proteins by tandem high-resolution mass spectrometry**. *Proc Natl Acad Sci USA* 2000, **97**:10313-10317.

76. Zubarev RA, Horn DM, Fridriksson EK, Kelleher NL, Kruger NA, Lewis MA, Carpenter BK, McLafferty FW: **Electron capture dissociation for structural characterization of multiply charged protein cations**. *Anal Chem* 2000, **72**:563-573.

77. McLafferty FW, Horn DM, Breuker K, Ge Y, Lewis MA, Cerda B, Zubarev RA, Carpenter BK: **Electron capture dissociation of gaseous multiply charged ions by Fourier-transform ion cyclotron resonance**. *J Am Soc Mass Spectrom* 2001, **12**:245-249.

78. Sze SK, Ge Y, Oh H, McLafferty FW: **Plasma electron capture dissociation for the characterization of large proteins by top down mass spectrometry**. *Anal Chem* 2003, **75**:1599-1603.

79. Ge Y, El-Naggar M, Sze SK, Oh HB, Begley TP, McLafferty FW, Boshoff H, Barry CE III: **Top down characterization of secreted proteins from *Mycobacterium tuberculosis* by electron capture dissociation mass spectrometry**. *J Am Soc Mass Spectrom* 2003, **14**:253-261.

80. Sze SK, Ge Y, Oh H, McLafferty FW: **Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue**. *Proc Natl Acad Sci USA* 2002, **99**:1774-1778.

81. Ge Y, Lawhorn BG, ElNaggar M, Strauss E, Park JH, Begley TP, McLafferty FW: **Top down characterization of larger proteins (45 kDa) by electron capture dissociation mass spectrometry**. *J Am Chem Soc* 2002, **124**:672-678.

82. Shi SD, Hemling ME, Carr SA, Horn DM, Lindh I, McLafferty FW: **Phosphopeptide/phosphoprotein mapping by electron capture dissociation mass spectrometry**. *Anal Chem* 2001, **73**:19-22.

83. VerBerkmoes NC, Bundy JL, Hauser L, Asano KG, Razumovskaya J, Larimer F, Hettich RL, Stephenson JL Jr: **Integrating 'top-down' and 'bottom-up' mass spectrometric approaches for proteomic analysis of *Shewanella oneidensis***. *J Proteome Res* 2002, **1**:239-252.

84. Taylor JA, Johnson RS: **Implementation and uses of automated *de novo* peptide sequencing by tandem mass spectrometry**. *Anal Chem* 2001, **73**:2594-2604.

85. Johnson RS, Taylor JA: **Searching sequence databases via *de novo* peptide sequencing by tandem mass spectrometry**. *Mol Biotechnol* 2002, **22**:301-315.

86. Taylor JA, Johnson RS: **Sequence database searches via *de novo* peptide sequencing by tandem mass spectrometry**. *Rapid Commun Mass Spectrom* 1997, **11**:1067-1075.

87. Fernandez-De-Cossio J, Gonzalez J, Satomi Y, Shima T, Okumura N, Besada V, Betancourt L, Padron G, Shimonishi Y, Takao T: **Automated interpretation of low-energy collision-induced dissociation spectra by SeqMS, a software aid for *de novo* sequencing by tandem mass spectrometry**. *Electrophoresis* 2000, **21**:1694-1699.

88. Dancik V, Addona TA, Clauser KR, Vath JE, Pevzner PA: **De *novo* peptide sequencing via tandem mass spectrometry**. *J Comput Biol* 1999, **6**:327-342.

89. Chen T, Kao M-Y, Tepel M, Rush J, Church GM: **A dynamic programming approach to *de novo* peptide sequencing via tandem mass spectrometry**. *J Comput Biol* 2001, **8**:325-337.

90. Shevchenko A, Sunyaev S, Loboda A, Shevchenko A, Bork P, Ens W, Standing KG: **Charting the proteomes of organisms with unsequenced genomes by MALDI-quadrupole time-of-flight mass spectrometry and BLAST homology searching**. *Anal Chem* 2001, **73**:1917-1926.

91. Liska AJ, Shevchenko A: **Expanding the organismal scope of proteomics: cross-species protein identification by mass spectrometry and its implications**. *Proteomics* 2003, **3**:19-28.

92. Liska AJ, Shevchenko A: **Combining mass spectrometry with**
•• **database interrogation strategies in proteomics**. *Trends Anal Chem* 2003, **22**:291-298.
An excellent review of recent progress in database searching, particularly in connection with BLAST homology.

93. Shevchenko A, Sunyaev S, Liska A, Bork P, Shevchenko A: **Nanoelectrospray tandem mass spectrometry and sequence similarity searching for identification of proteins from organisms with unknown genomes**. *Methods Mol Biol* 2003, **211**:221-234.

94. Sunyaev S, Liska AJ, Golod A, Shevchenko A, Shevchenko A: **MultiTag: multiple error-tolerant sequence tag search for the sequence-similarity identification of proteins by mass spectrometry**. *Anal Chem* 2003, **75**:1307-1315.

95. Cannon WR, Jarman KD: **Improved peptide sequencing using isotope information inherent in tandem mass spectra**. *Rapid Commun Mass Spectrom* 2003, **17**:1793-1801.