

UCSF

UC San Francisco Previously Published Works

Title

The Identification of Protein-Protein Interactions of the Nuclear Pore Complex of *Saccharomyces cerevisiae* Using High Throughput Matrix-assisted Laser Desorption Ionization Time-of-Flight Tandem Mass Spectrometry*

Permalink

<https://escholarship.org/uc/item/0ch59498>

Journal

Molecular & Cellular Proteomics, 1(6)

ISSN

1535-9476

Authors

Huang, Lan

Baldwin, Michael A

Maltby, David A

et al.

Publication Date

2002-06-01

DOI

10.1074/mcp.m200027-mcp200

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

The Identification of Protein-Protein Interactions of the Nuclear Pore Complex of *Saccharomyces cerevisiae* Using High Throughput Matrix-assisted Laser Desorption Ionization Time-of-Flight Tandem Mass Spectrometry*

Lan Huang‡, Michael A. Baldwin‡, David A. Maltby‡, Katalin F. Medzihradsky‡, Peter R. Baker‡, Nadia Allen§, Michael Rexach§, Ricky D. Edmondson¶, Jennifer Campbell||, Peter Juhasz||, Steven A. Martin||, Marvin L. Vestal||, and Alma L. Burlingame‡**

Mass spectrometry has become the technology of choice for detailed identification of proteins in complex mixtures. Although electrophoretic separation, proteolytic digestion, mass spectrometric analysis of unseparated digests, and database searching have become standard methods in widespread use, peptide sequence information obtained by collision-induced dissociation and tandem mass spectrometry is required to establish the most comprehensive and reliable results. Most tandem mass spectrometers in current use employ electrospray ionization. In this work a novel tandem mass spectrometer employing matrix-assisted laser desorption ionization-time-of-flight/time-of-flight operating at 200 Hz has been used to identify proteins interacting with known nucleoporins in the nuclear pore complex of *Saccharomyces cerevisiae*. Proteins interacting with recombinant proteins as bait were purified from yeast extracts and then separated by one-dimensional SDS-PAGE. Although peptide mass fingerprinting is sometimes sufficient to identify proteins, this study shows the importance of employing tandem mass spectrometry for identifying proteins in mixtures or as covalently modified forms. The rules for incorporating these features into MS-Tag are presented. In addition to providing an evaluation of the sensitivity and overall quality of collision-induced dissociation spectra obtained, standard conditions for ionization and fragmentation have been selected that would allow automatic data collection and analysis, without the need to adjust parameters in a sample-specific fashion. Other considerations essential for successful high throughput protein analysis are discussed. *Molecular & Cellular Proteomics* 1:434–450, 2002.

From the ‡University of California, San Francisco, California 94143-0446, §Stanford University, Stanford, California 94305, ¶Genomic Solutions, Inc., Ann Arbor, Michigan 48108, and ||Applied Biosystems, Framingham, Massachusetts 01701

Received, May 21, 2002, and in revised form, May 29, 2002

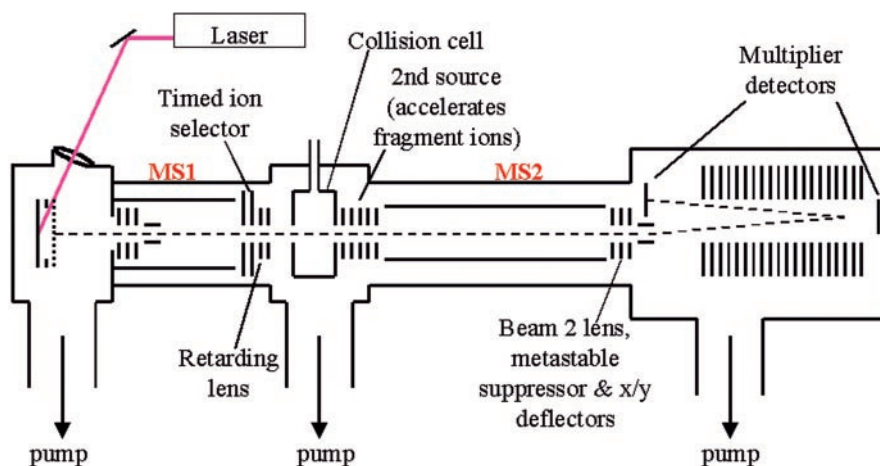
Published, MCP Papers in Press, May 30, 2002, DOI 10.1074/mcp.M200027-MCP200

Over the last five years, rapid advances in commercial instrumentation have brought the power of matrix-assisted laser desorption ionization (MALDI)¹ mass spectrometry (MS) into widespread use for protein identification in biomedical research. This growth is because of the fact that high sensitivity (low femtomole level) may be achieved, together with the ability to measure mass with high accuracy (<10 ppm) (1, 2). Hence, proteins separated by 1-D and two-dimensional SDS-PAGE and visualized by silver and/or fluorescence stains may be identified readily based on proteolytic digest mass maps and the use of database interrogation strategies (2–4). These features have accelerated the pace of research focused on examining the composition of complex mixtures of proteins including cell lysates (5–8), as well as protein-protein interactions (9–14).

There are many cases when results derived solely from database mass matching of digests are inadequate, incomplete, and/or even unreliable. In these cases, it is essential to have the ability to rapidly sequence peptides present in these digests. Although instrumentation has been available to make such measurements using electrospray ionization for some time (15, 16), instrumentation that could obtain peptide sequences with comparable sensitivities and spectral quality using MALDI has not. Hence, investigators have either had to switch ionization or have had to rely on the poor quality of spectra obtained from MALDI metastable fragmentation, so-called post-source decay spectra (17, 18). Either choice has drawbacks in practice, for example, the need to incorporate a clean-up step prior to being able to carry out a successful electrospray experiment, the problems of poor sensitivity, the variable and unpredictable quality of post-source decay mass spectra, and the long spectra acquisition time.

¹ The abbreviations used are: MALDI, matrix-assisted laser desorption ionization; 1-D, one-dimensional; MS, mass spectrometry; MS/MS, tandem MS; TOF, time-of-flight; TOF/TOF, tandem time-of-flight; CID, collision-induced dissociation; GST, glutathione S-transferase.

FIG. 1. Schematic diagram of an upgraded version of MALDI-TOF/TOF MS.



Based on a decade of experience employing four-sector tandem mass spectrometers for peptide sequence determination with singly protonated pseudomolecular ions (19, 20) and the advantage of the tolerance of MALDI of salts and buffers (21), the importance of gaining the ability to measure high quality sequence from MALDI was obvious. Proof of principle was established using the first orthogonal acceleration time-of-flight analyzer as MS-2 with a MALDI ion source on a double focusing magnetic sector instrument (MS-1) (22). Studies of peptides and modified peptides, as well as the co-factor of lysyl oxidase (23) and conotoxins (24), established the advantages of MALDI high energy CID spectra for reliable *de novo* peptide sequence and structure determination (23–26).

This work set the stage for the development of a less expensive MALDI tandem time-of-flight instrument with on axis ion optics capable of producing four-sector-like quality CID spectra for protein characterization while simultaneously realizing an increase in several orders of magnitude of absolute sensitivity (27). Our laboratories have reported the performance of an engineering prototype elsewhere (28).

In previous studies of the proteomics of the nuclear pore complex, we have employed manual digestions using MALDI mass mapping and/or liquid chromatography electrospray ionization CID MS as necessary for the identification of 135 proteins interacting with eight FG nucleoporins (11, 29). The present work describes the power and advantages of employing robotic sample processing, together with peptide sequence analysis using MALDI-TOF/TOF MS, to achieve very rapid analysis of the protein interactions detected using 22 different GST fusion proteins as “baits” in separate experiments.

EXPERIMENTAL PROCEDURES

Instrument Design—A preliminary research version of the TOF/TOF instrument based on the footprint of the standard Applied Biosystems Voyager-DE STR spectrometer was described previously (28). Subsequent improvements have increased the performance such that the resolving power in MS mode for peptide molecular ions in the typical mass range of 1 to 3 kDa is now 12,000–20,000 (full width half-maximum), and fragment ions above m/z 200 display a mass resolu-

tion of 4,000 to 8,000. For the spectra presented here, precursor ion mass selection for MS/MS operation without loss of sensitivity was 1%; subsequent changes have improved this to ~0.5%. With internal calibration the mass measurement accuracy for fragment ions is routinely in the region of 50 mDa for all masses, and 10 mDa may be achieved using multipoint calibration. External calibration typically gives errors within 100 to 200 mDa. All spectra shown here were obtained with a neodymium yttrium aluminum garnet (YAG) laser operating at a wavelength of 354 nm and at a frequency of 200 Hz. The sensitivity for interpretable CID spectra of peptide standards is in the range of 1 to 10 fmol deposited on the target, although this is sufficient material to provide multiple spectra. The instrument is rapidly switchable between MS and MS/MS modes without recalibration (in <2 min), which is valuable for recording survey scans to identify potential precursor ions for subsequent recording of CID MS. A schematic diagram that highlights changes in the design of the instrument is shown in Fig. 1.

The first mass analyzer used for precursor ion selection is essentially as described before, being a dual stage Wiley McLaren TOF mass spectrometer. Compared with the earlier version, higher resolution has been achieved by increasing the flight length from 26 to 40 cm. This TOF is combined with an improved timed ion selector consisting of two deflection gates, the opening and closing of which is under software control. Ion selection relies upon deflection off the ion optical axis of all ions other than those of the desired m/z value. The extent of this deflection is determined in part by the distance between the timed ion selector and the collision chamber, which in this case is 7 cm. The collision cell that was previously grounded is now electrically floating, allowing the accelerating voltage in TOF1 to be increased above its previous value of 3 kV without necessarily increasing the collision energy. The higher accelerating voltage improves ion transmission in TOF1. The collision energy is defined by the difference between the TOF1 accelerating potential and the potential of the collision cell, 8 and 7 kV, respectively, in these experiments. Thus the laboratory collision energy for all data presented here was 1 keV. After passing through the timed ion selector, a new decelerating lens ensures that the reduction in ion energy upon entry to the collision cell is completed with minimal defocusing and loss of signal. Collision gas is bled into the collision cell through an adjustable leak valve, and the pressure is maintained via differential pumping through $\frac{1}{8}$ -inch apertures at either end of the cell. No direct measure of the pressure in the collision cell is provided, but the amount of gas added is inferred from the increase in pressure measured by the ionization gauge in the high vacuum region close to the cell. For standard CID operation, the indicated pressure increases from 2.0×10^{-8} torr (no collision gas) to $\sim 2.0 \times 10^{-6}$ torr. The

second section is another pulsed TOF mass analyzer, which provides for recording of high resolution spectra of the fragment ions. This spectrometer has a two-stage acceleration region that increases the ion energy to 15 keV via a rapid high voltage power switch. Only ions that are in the volume of the second source at the time the pulse is applied are accelerated toward the ion mirror. Software calculates the flight time of the precursor ions to the second source; fragment ions formed in the collision cell have the same velocities and thus the same flight times and are accelerated similarly toward the ion mirror. The guide wire that was employed in the earlier instrument has been replaced by a second einzel lens, giving improved resolution and transmission. A pulsed metastable suppressor has been added that can be used to reject all ions with velocity equal to that of the precursor ion, *i.e.* the precursor ion and any fragment ions formed by its decomposition after acceleration in the second source. New X-Y electrostatic deflectors enable the reflector that was offset previously by 0.5° to be aligned with the major axis of the TOF.

Protein Purification, Separation, and Robotic In-gel Digestion—Nucleoporin-interacting proteins were purified by affinity-capture experiments as described previously (11). Twenty-two different recombinant GST-Nups and karyopherins were selected as bait. Briefly, affinity-capture experiments were performed in binding buffer (20 mM Hepes, pH 6.8, 150 mM KOAc, 2 mM Mg(OAc)₂, 2 mM dithiothreitol, 0.1% Tween 20). For each experiment, glutathione-Sepharose beads (Amersham Biosciences) and an *Escherichia coli* extract containing the desired GST-Nup were incubated for 20 min at 4 °C. Beads were collected by sedimentation at 2000 × *g* for 30 s, washed seven times by resuspension and sedimentation, and incubated with 1 ml of yeast extract (~10 mg). In some experiments, the yeast extracts were supplemented with 30 μg of HIS-Gsp1p-GTP (Q71L) 15 min prior to mixing with the immobilized GST-Nup. After 2 h at 4 °C, beads were washed six times with binding buffer, and bound proteins were eluted with 100 μl of 1 M NaCl and concentrated by trichloroacetic acid/sodium deoxycholate precipitation. Proteins that resisted salt extraction were subsequently extracted from the beads with SDS. All proteins were resolved by SDS-PAGE (7.5% gels) and visualized with Standard Coomassie Blue. 288 protein bands were selected from 10 different 1-D gels. Approximately one quarter of each band was cut out as a section of gel 1.2–1.5 mm square. The gel pieces were loaded into three 96-well plates, mounted in a Genomic Solutions ProPrep digestion robot, reduced and alkylated, and digested with trypsin in 25 mM NH₄HCO₃ for 4 h. The resulting unseparated digests were desalted robotically using micro C18 Ziptips. The peptides were washed from the Ziptips with a 1.8-μL elution solution containing diluted α-cyano-4-hydroxycinnamic acid matrix (six dilutions of saturated α-cyano in 60% acetonitrile/0.2% trifluoroacetic acid). The solutions were loaded directly onto three 96-well Teflon-coated MALDI plates for mass spectrometric analysis.

Mass Spectrometric Analysis—In the first pass of each sample plate, every sample spot was analyzed by MALDI-TOF MS in automated mode using an Applied Biosystems Voyager-DE STR reflectron mass spectrometer, operating under software control (Proteomics System 1). Spectra were accumulated for each sample until a specified signal strength and signal/noise were attained, and all spectra were submitted automatically to ProteinProspector for batch MS-Fit database searching. The acquisition time was ~1 h per sample plate. Spectra were also inspected visually, and a number of peptides were chosen for CID analysis by tandem mass spectrometry using the MALDI-TOF/TOF. For all tandem mass spectra reported here the laser frequency was 200 Hz, the laboratory collision energy was 1 keV, and air was used as collision gas. The laser energy employed for effective CID operation was typically ~30% higher than that used for normal MALDI operation. Variable numbers of laser shots were used to obtain each spectrum, depending upon the strength of the primary ion

beam, but acquiring the output from 1,000 shots required only 5 s, and as many as 10,000 shots could be accumulated within less than 1 min. Post-acquisition baseline correction and smoothing was carried out using software provided with the TOF/TOF instrument. Sequence tags were identified in CID spectra by visual inspection and selection of the major peaks for submission to MS-Tag. Some CID spectra were used to confirm assignments already made by peptide mass fingerprinting, and others were used to establish peptide identity for those not matched in the previous MS experiments. For those peptides that could not be assigned by MS-Tag, the CID spectra were used to carry out *de novo* peptide sequencing by manual interpretation. *De novo* assignments were important to monitor for nonspecific protein cleavage sites and the presence of chemical or post-translational modifications.

Software to Identify Sequence Tags from MALDI-TOF/TOF Spectra—Programs in the ProteinProspector software package that predict peptide fragmentation, such as MS-Pattern, or identify proteins based on their fragmentation, such as MS-Tag, were modified to accommodate the side chain fragments **d**, **v**, and **w** and other spectral features from high energy CID. The changes were based on the following rules for high energy CID fragmentation, derived from a compilation of spectra from four-sector tandem mass spectrometry (19, 20):

- No high energy fragments would be observed unless Arg, Lys, or His was present in the fragment, and that basic residue would not yield side chain fragments. For example, C-terminal Arg would not form **v** and **w** ions. However, it could form a **d** fragment if the peptide contained another basic amino acid. If two basic residues are present, both of them may produce side chain fragmentation.
- Ala, Gly, His, Phe, Pro, Trp, and Tyr, residues do not form **d** ions.
- Ala, Gly, His, Phe, Trp, and Tyr do not form **w** ions.
- Thr and Ile may each yield two **w** and **d** fragments because of the double substitution at the β-carbon in these amino acids.
- Gly and Pro do not form **v** ions.
- For amino acids N-terminally adjacent to Pro residues, no **c** fragments will be formed.
- Fragment **z**, defined as “**y** – 16 Da,” cannot be formed by Pro.
- If the fragment contains Arg, Asn, Gln, or Lys residues, a neutral loss of NH₃ (–17 Da) should be considered.
- If the fragment contains Asp, Glu, Ser, or Thr, neutral loss of H₂O (–18 Da) should be considered.
- The eighth and ninth rules apply to the internal fragment ions, as well.
- The first members of the N-terminal ion series will not be identified, because **b**₁ is not stable unless N-terminally modified, **a**₁ is observed as the immonium ion, and **c**₁ would be observed only rarely.

Typical MS-Tag results are presented and discussed below. The use of MS-Product is not illustrated here, but it should be noted that to generate a list of fragment ions including **d**, **v**, and **w** ions, the appropriate boxes must be checked on the ProteinProspector user interface.

RESULTS AND DISCUSSION

To evaluate the feasibility of carrying out high throughput protein identification using the MALDI-TOF/TOF tandem mass spectrometer, we have taken as an example the analysis of proteins interacting with known nucleoporin components in the nuclear pore complex (10). Target proteins were purified from yeast extracts using recombinant proteins as bait and then separated by 1-D SDS-PAGE. The criteria for selection of particular bands for analysis was the appearance or absence

FIG. 2. MALDI-TOF MS spectrum of unseparated tryptic digest from band B5 on plate 1 using GST-Nsp1 as bait. The inset displays part of the 1-D gel picture where the gel slice at ~120 kDa was sampled. Three gel lanes represent the following: (-), control without yeast extract; S30, normal yeast extract; *Ran*, yeast extract spiked with Gsp1p-GTP (Q71L). Detailed experimental procedure was described elsewhere (11). *T* represents trypsin autolysis products, and the rest of the peaks labeled were matched to karyopherin β -4 subunit. The peaks with * were chosen for CID analysis.

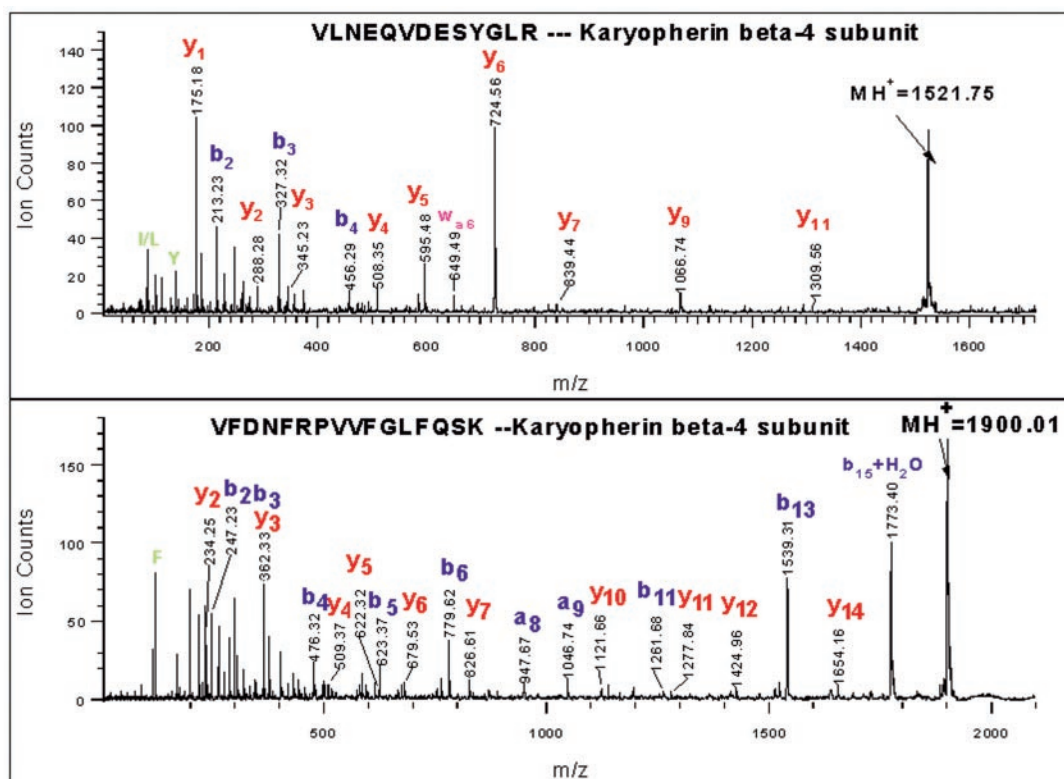
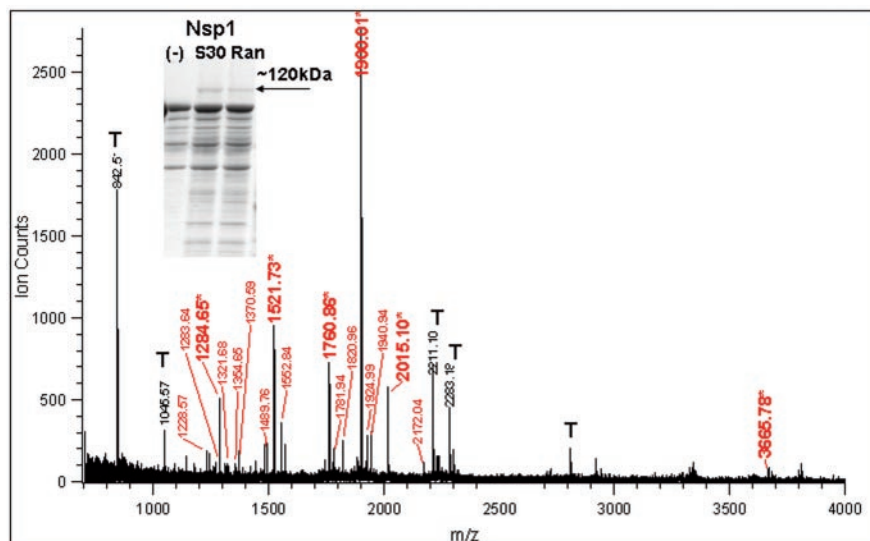


FIG. 3. MALDI-TOF/TOF CID spectra of two tryptic peptides from band B5 digest (shown in Fig. 2). Top, $MH^+ = 1521.7$; bottom, $MH^+ = 1900.0$. Both peptide sequences were matched to karyopherin β -4 subunit unambiguously after database searching using the MS-Tag program.

of protein bands based on a comparison of control and normal yeast extracts. On this basis, 288 protein bands were excised, digested with trypsin, and desalted, and the resulting peptide mixtures were analyzed by MALDI-TOF and MALDI-TOF/TOF. The results from selected gel bands are presented and discussed below.

Protein Identification by Peptide Mass Mapping, Confirmed by Peptide Sequencing—The inset of Fig. 2 shows protein enrichment of a yeast extract using Nsp1 as the recombinant nucleoporin bait. Compared with the control (*left lane*) designated “(-),” a protein band at ~120 kDa is clearly evident in the lane corresponding to the normal yeast extract (*middle*

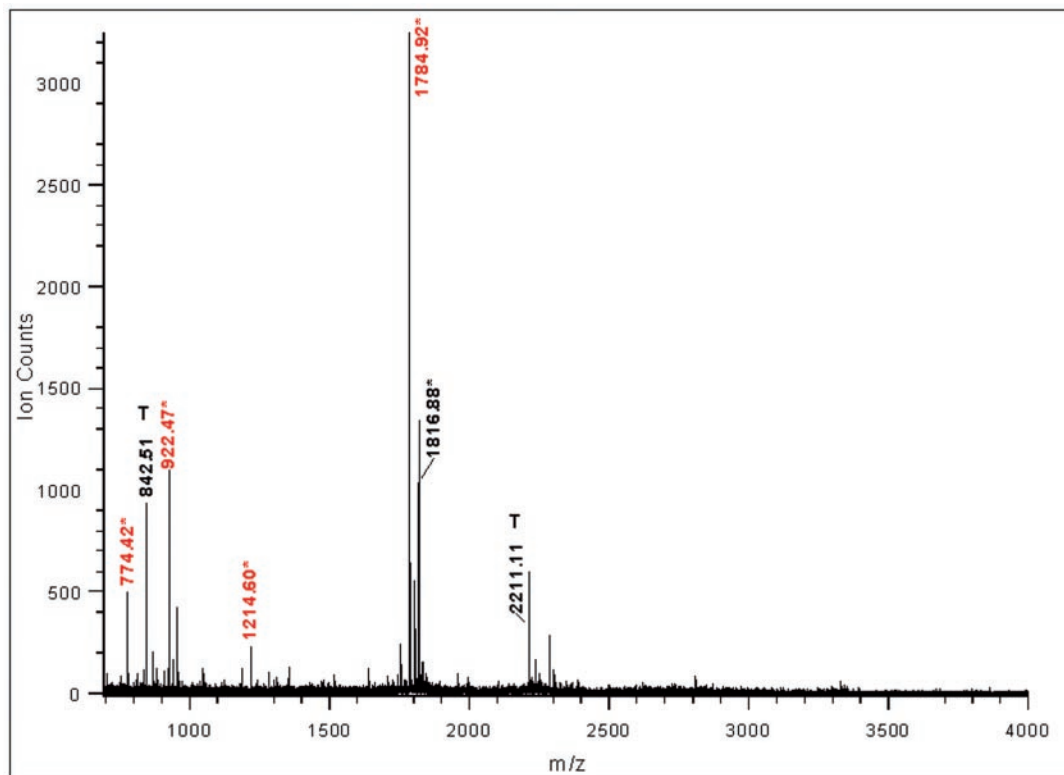


FIG. 4. MALDI-TOF MS spectrum of unseparated tryptic digest from band B8 on plate 2. *T* represents trypsin autolysis products, and five labeled peaks were chosen for CID analysis.

lane) and the extract spiked with Ran-GTP protein (Gsp1p-GTP) (right lane), as described previously (11). This band, named B5, was cut out, digested, and analyzed by MALDI-TOF MS, giving the spectrum illustrated in Fig. 2. The peptides labeled “*T*” represent trypsin autolysis products, which were used as internal standards for calibration of the mass scale. This calibration provided mass measurement accuracy within 20 ppm for database searching. This single mass spectrum contained 29 other peptides, all of which were submitted to MS-Fit for database searching. The search results gave one significant protein hit based upon 19 peptides that matched yeast karyopherin β -4 subunit labeled in the figure. To further establish the protein identity, MALDI-TOF/TOF CID spectra were recorded for those peptides labeled with *asterisks*. The results for two of these are illustrated in Fig. 3, showing the CID spectra of the tryptic peptides of MH^+ 1521.75 and 1900.01, respectively. The low mass region of Fig. 3, *top* for m/z 1521.75 shows strong immonium ions such as m/z 86 (Ile/Leu) and 136 (Tyr) that are typical of these CID spectra. Most of the members of the *y* ion series, y_1 to y_7 , y_9 , and y_{11} , and a few members of the *b* ion series are present. In the CID spectrum of the larger peptide of m/z 1900.01 (Fig. 3, *bottom*), most of the sequence ions (*b* and *y* ion series) are evident. Thus both of these CID spectra gave definitive fragment ion information, allowing the MS-Tag search program to identify the peptide sequences as VLNEQVDESYGLR (MH^+ =

1521.7) and VFDNFRPVVFGFLFQSK (MH^+ = 1900.01) from the yeast karyopherin β -4 subunit. The remaining CID spectra obtained also confirmed this assignment, consistent with the known function of this protein as a carrier involved in nucleocytoplasmic transport (11, 29–32).

Protein Identification in a Complex Mixture by Sequence Tags—Unseparated protein digests may be very complex, particularly from 1-D electrophoresis. A potential problem with protein identification by peptide mass fingerprinting of such mixtures may be a low fraction of the total peptide mass measured that match to any given single protein (less than 20%). In such cases, no clear candidate protein(s) may be selected reliably. The MALDI-TOF MS spectrum of the digest of band B8, plate 2 is shown in Fig. 4 and provides an example illustrating this point. The mass values of 25 peptides observed in this spectrum were submitted to MS-Fit for searching against a yeast database. No proteins were found to be matched with more than four of the total number of these peptides, whereas there were several hundred hits matching to four. Consequently no definitive candidate could be selected based on the MS-Fit search results alone. To obtain sequence information, TOF/TOF CID analysis was performed on the more abundant components labeled in Fig. 4, m/z 774.42, 922.47, 1214.60, 1784.92, and 1816.88. TOF/TOF CID spectrum of the tryptic peptide of MH^+ 774.4 is shown in Fig. 5A. The immonium ion region is typical in that it

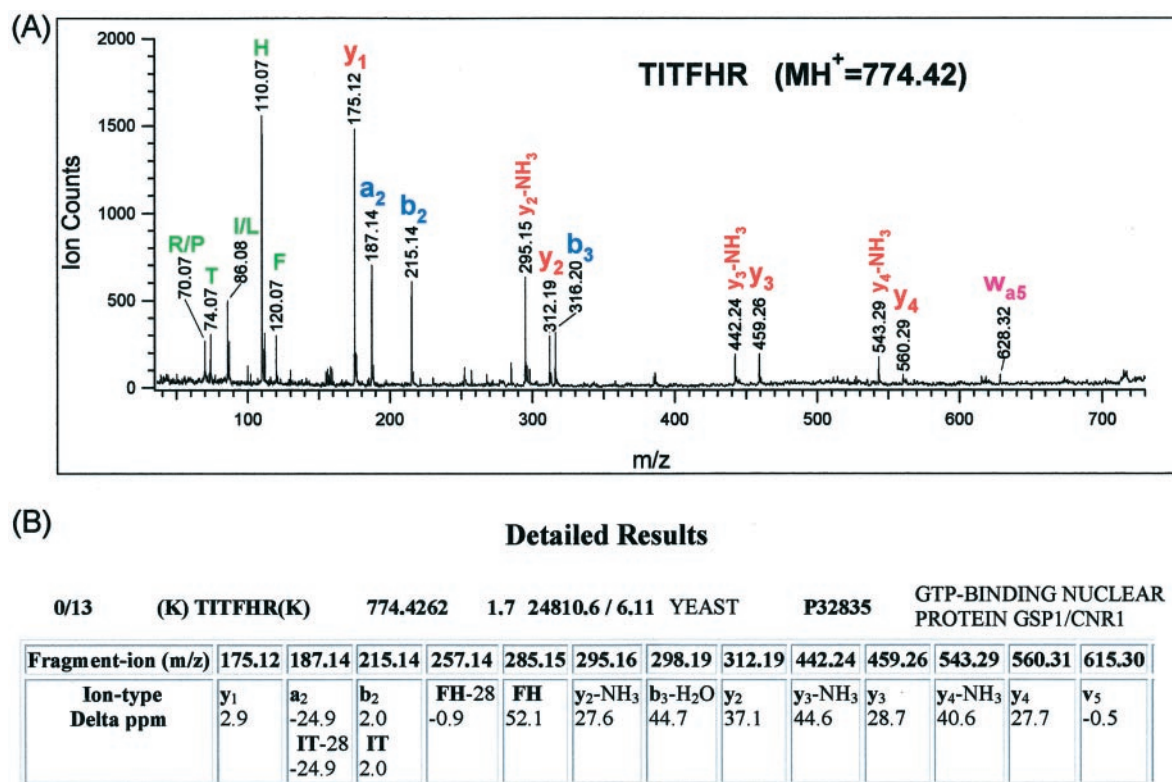


Fig. 5. A, MALDI-TOF/TOF CID spectrum of a tryptic peptide (MH^+ 774.4) from band B8 digest (shown in Fig. 4); B, detailed MS-Tag search result of A, indicating this peptide was matched to yeast GSP1.

indicates clearly the presence of certain amino acid residues, Arg/Pro (70), Thr (74), Ile/Leu (86), His (110), and Phe (120), and the y_1 ion of m/z 175 establishes the presence of an arginine residue at the C terminus. After carrying out a two-point internal calibration based on the y_1 and MH^+ ions, 14 of the fragment ions recorded were submitted to MS-Tag for database searching with a fragment ion mass measurement tolerance of 50 ppm. Every peak matched to a single protein, yeast GTP-binding nuclear protein GSP1, as shown in Fig. 5B, having an average mass error of 22 ppm. To illustrate the degree of coverage, the sequence ions (a , b , and y) and the immonium ions are labeled in Fig. 5A. The presence of the characteristic high energy CID fragment ion, w_{a5} , at m/z 628.3 would establish the presence of Ile rather than Leu at position 2 in this peptide sequence, determined as TITFHR.

Some of the CID spectra obtained from band B8 provided sufficient fragment ions for *de novo* peptide sequencing. An example is given in Fig. 6, showing the TOF/TOF CID spectrum of the tryptic peptide of MH^+ 1214.60. The immonium ion region indicates the presence of amino acid residues Ile/Leu (86), Gln/Lys (101), and Tyr (136). Sequence ion series, such as b_2 to b_8 and y_1 to y_9 , are observed, allowing the sequence of this peptide to be determined as N(I/L)(Q/K)YYD(I/L)SAK, and internal ions and satellite ions were consistent with this. The presence of w_{a4} and w_{b4} indicates Ile at the 4th position from the C terminus, whereas the presence of

w_{a9} indicates Leu at the 9th position from the C terminus. An MS-Tag search also identified the sequence as NLQYYDISAK matched to yeast GTP-binding protein GSP1. Furthermore, the TOF/TOF CID spectra of two further peptides of MH^+ at 922.47 and 1784.92 also gave unambiguous matches to this same protein. Thus peptides giving four of the strongest signals from band B8 unambiguously identified yeast GTP binding-protein GSP1 by peptide sequencing.

Another example of the difficulty of protein identification in cases where a mixture of proteins is present concerns a single 40-kDa band (G11) purified from a normal yeast extract using Rna1 as bait. The MALDI-TOF MS spectrum of the band digest was recorded. Sixty-six peaks were observed. After internal calibration, they were submitted to MS-Fit. Yeast elongation factor 1- α was the only significant hit, with 20 peaks matching, as labeled with a in Fig. 7. MS-Fit searches on the remaining peaks gave a variety of hits, but none were very significant. Selected ions were chosen for CID analysis, some to confirm the identification made by MS-Fit and others to identify additional protein components. The CID spectra of two peptides of MH^+ 3319.86 and 1326.75 (shown in Fig. 8, A and B, respectively) did indeed confirm the identification of yeast elongation factor 1- α . The peptides were determined by a MS-Tag search as TLEAIDAEQPSRPTDKPLRLPLQDVYK and EHALLAFTLGVR. Many sequence ions and a few side chain fragment ions were present in these CID spectra, e.g. in

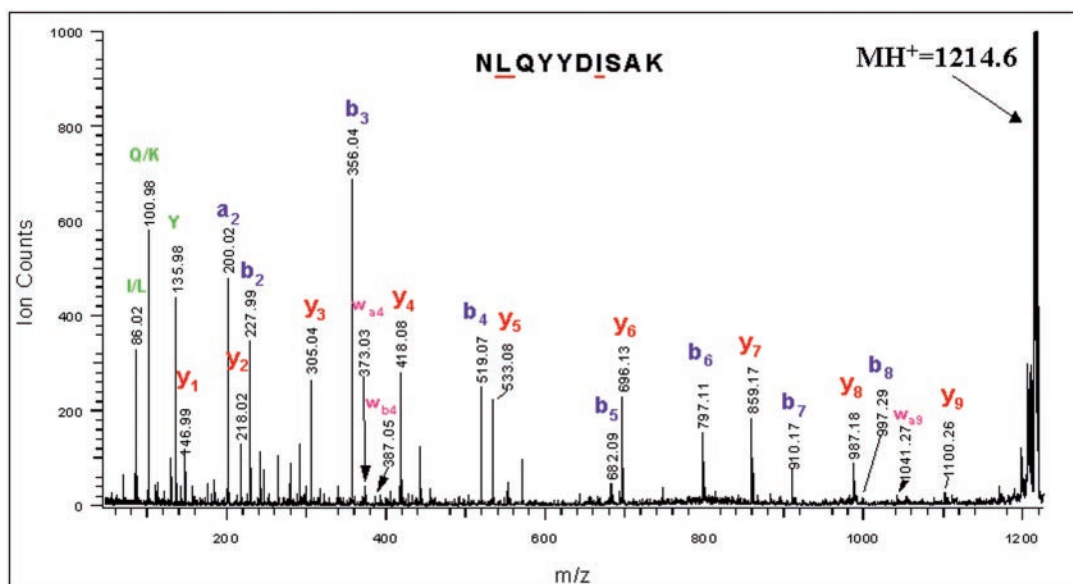


Fig. 6. MALDI-TOF/TOF CID spectrum of a tryptic peptide (MH^+ 1214.6) from band B8 on plate 2. The comprehensive fragmentation information allows *de novo* peptide sequencing with the capability of distinguishing Ile and Leu. The sequence was determined as NLQYYDISAK and was matched to GSP1 protein.

Fig. 8B, the presence of w_{a4} , w_{a8} , and w_{a9} generated by high energy CID distinguished Leu from Ile at the 4th, 5th, and 9th positions. Discussion of this sample continues below in the context of unresolved precursor ions.

CID of Peptides with Unresolved Precursor Ions—In a complex digest it is likely that some peptides will be sufficiently close in mass values that their isotope clusters will overlap. Unlike four-sector tandem mass spectrometry in which monoisotopic spectra could be obtained (20, 33, 34) the precursor mass window selected by the TOF/TOF is several mass units wide, and CID analysis may provide fragment ions originating from multiple precursors. Furthermore, the analysis is complicated as the overlapping peptides could originate from the same protein or from different proteins. Fig. 9 from band G11 represents an example of a CID spectrum of two ions differing in mass by only 1 Da, as shown in the *inset*. Although m/z 1355.70 was selected for CID analysis, a weaker peptide of MH^+ 1354.68 also contributed to the fragment ion spectrum. The higher mass peptide (peptide 1) was matched to elongation factor 1- α , which had already been identified (see above), and its sequence was assigned as YQVTVIDAPGHR. Although the lower mass spectrum was weaker and gave relatively few peaks, the sequence GFGREEFSQVAK was matched to serine methylase (mitochondrial). The ions attributed to each peptide are shown in Fig. 9, where the ions labeled with * were matched to both sequences. Although two proteins, elongation factor 1- α and serine methylase (mitochondrial), had been identified by peptide mass mapping and peptide sequencing, many peptides in the digest of band G11 were still unassigned. Some proteins were not well separated in 1-D SDS-PAGE, so the presence of overlapping peptides

was probed to assist in the assignment of the remaining peaks. Comparing the peptide masses for band G11 with those from adjacent bands revealed many coincident masses, indicating the presence of the same proteins. CID analysis of selected ions in the bands running before and after band G11 helped to confirm the previous two assignments and to identify three additional proteins as Ran GTPase activating protein 1, serine methylase (cytosolic), and MAS5. Ultimately 51 of the 66 peptides from G11 were attributed to the five proteins as summarized in Table I, each protein contributing a unique set of peptides as shown in Fig. 7. During high throughput analysis, a comparison of data from adjacent bands should increase the accuracy of assignments and eliminate unnecessary repetitive data acquisition and processing.

Automated CID analysis of unresolved peptides is potentially difficult as mixed fragmentation information can complicate database searches. However, sequential searches by MS-Tag on the same dataset but with different precursor masses specified may allow protein identification. Fig. 10 illustrates the TOF/TOF CID spectrum of peptides with MH^+ at 1962.01 and 1964.97 as shown in the *inset*. These two peptides separated by only 3 Da could not be resolved in MS1 of the MALDI-TOF/TOF. Consequently although several immonium ions were observed in the CID spectrum, they could not be easily assigned to either peptide. Eighteen fragment ions obtained in the spectrum were submitted for database searching with a mass tolerance for the internally calibrated parent ion of 20 and 300 ppm for the externally calibrated fragment ions. The maximum allowable number of unmatched ions was set to 40%, *i.e.* seven ions. The maximum number of missed cleavage sites was set to 1, and cysteines were alky-

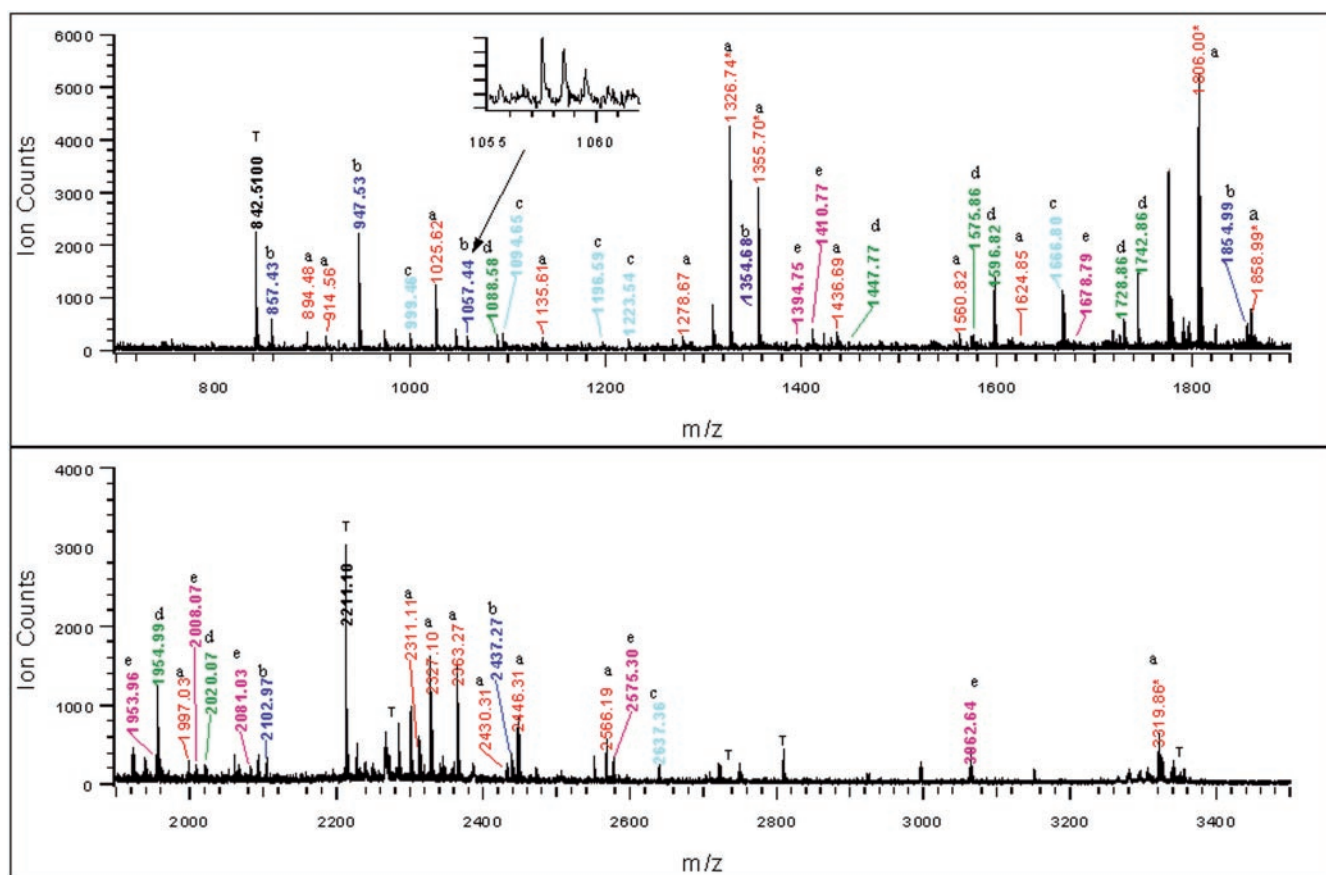


FIG. 7. MALDI-TOF MS spectrum of unseparated tryptic digest of band G11 on plate 2. *T* represents trypsin autolysis product. Peaks marked with * were selected for CID analysis. Five proteins were identified by peptide mass mapping and sequencing, and peaks matched to different proteins were labeled with a, EF-1- α ; b, serine methylase, mitochondrial; c, serine methylase, cytosolic; d, Ran GTPase activating protein 1; e, mitochondrial import protein MAS5. The detailed results are summarized in Table I.

lated by iodoacetamide. When m/z 1964.97 was used as the parent ion mass, eight hits were obtained from the SwissProt database. The four of these with the lowest number of unmatched ions (3 of 18) were all Elongation factor Tu from different species, which was the same match that had been obtained by MS-Fit. Most of the matched ions are sequence ions, *i.e.* **b** ions and **y** ions. The other four hits for different proteins with six or seven unmatched ions were almost certainly false positives as no other peptides could be found to match them. When 1962.01 was used as the parent mass for database searching, ten hits were returned. The top hit had six unmatched ions whereas the remainder all had seven. The three peaks previously unmatched by the first peptide were all matched only by two hits, also elongation factor Tu from two different species, suggesting the peptide sequence identity. In Fig. 10 all ions that were uniquely assigned to one or other peptide are marked with symbols (\wedge and \ast) whereas the ions without symbols were assigned to both sequences. In this example, the first sequence (MH^+ 1964.97) was easier to determine as it gave the top hit with the least unmatched ions. The second sequence was harder to distinguish from false positives, but all ions being assigned to one or other peptide

gave increased confidence, because all the ions should be interpretable. Furthermore, in this instance both sequences were derived from a common protein, and an MS-Fit search had previously given a clue to its identity.

Identification of Modified Proteins by Peptide Sequencing—Protein identification by peptide mass mapping is fast and simple. However, the identification process can be complicated because of the presence of protein mixtures (as discussed above), the presence of only a few peptides for database searching, presence of chemical modifications, and nonspecific cleavages. As discussed above, four of 25 peaks from band B8 tryptic digest (Fig. 4) matched to yeast GSP1 protein by peptide sequencing (Fig. 5) by MS-Tag. However, one abundant ion of MH^+ 1816.88 did not match this protein, and its CID spectrum could not be matched by MS-Tag. However, strong similarities between the CID spectra of MH^+ 1784.92 (Fig. 11A, SNYNFEKPFLLWLAR) and this peptide (Fig. 11B) were noted, suggesting the latter to be a modified peptide with a mass difference of +32. The **b** ions up to and including **b**₅ were observed to be the same in both cases, *i.e.* **b**₂ (202), **b**₃ (365), **b**₄ (479), and **b**₅ (626), as were **y**₁ (175) and **y**₃ (359), and lower mass internal ions were observed to be the

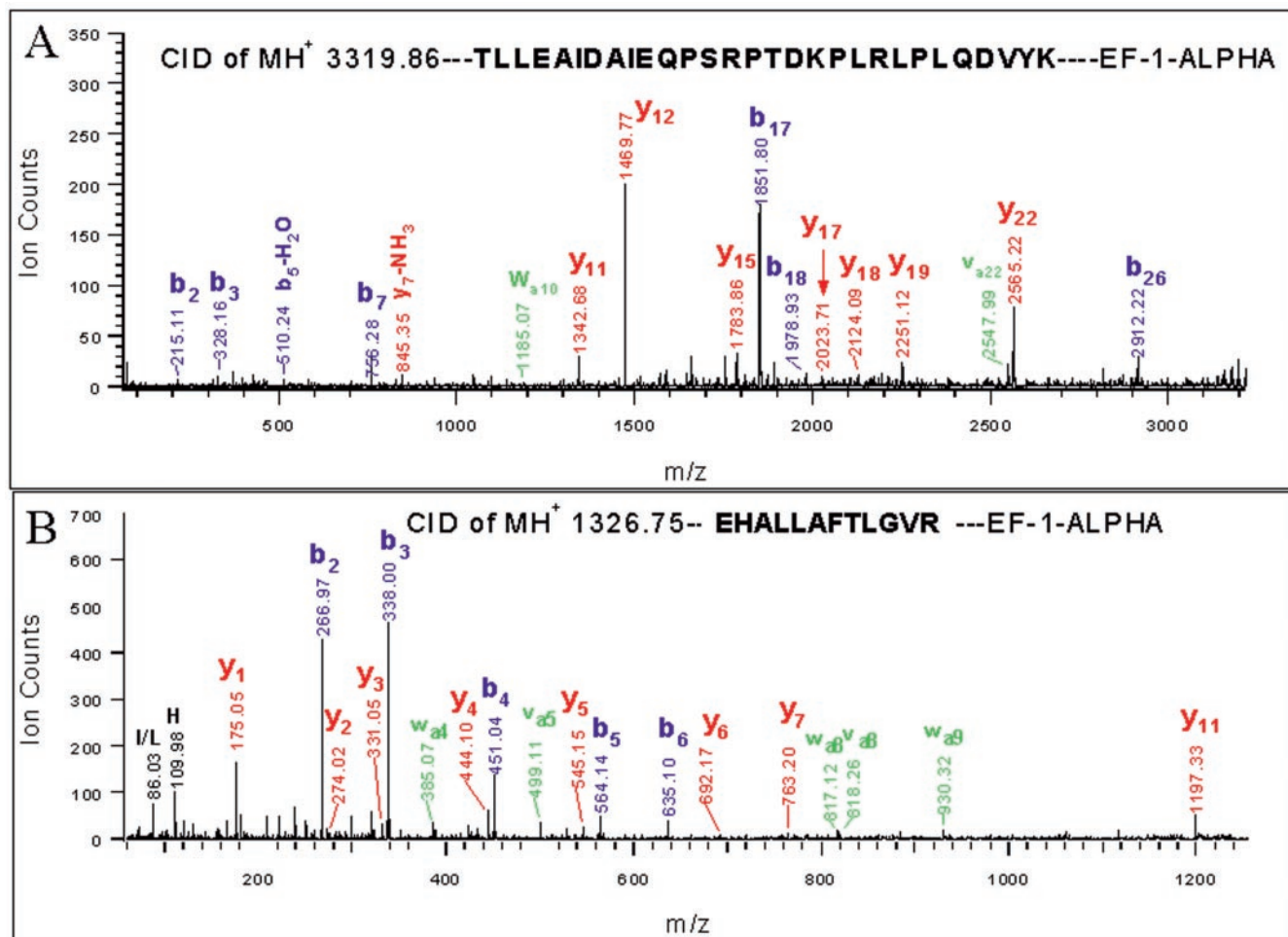


Fig. 8. MALDI-TOF/TOF CID spectra of two tryptic peptides from band G11 digest (shown in Fig. 6). A, $MH^+ = 3319.9$; B, $MH^+ = 1326.7$. Both peptide sequences were matched to EF-1- α unambiguously after database searching using the MS-Tag program.

same in both spectra. However, from y_4 onwards, y ions and associated ions were shifted higher by 32 mass units, e.g. y_4 (545) versus $y_4-NH_3^+$ (560), y_7 (902) versus y_7^+ (934), and y_8 (1030) versus y_8^+ (1062). Therefore the site of the modification was identified as the tryptophan at the 4th position from the C terminus. It is known that tryptophan can be oxidized through a Cu^{2+} -catalyzed reaction to form *N*-formylkynurenine (35), with the addition of two oxygen atoms, increasing the mass by 32 Da. Therefore, the peptide sequence of $MH^+ 1816.88$ was proposed as SNYNFEKPLW(2O)LAR, a modified peptide from GTP-binding protein GSP1. Because this protein modification is not common, the peptide mass would not be predicted by database searching tools. Nevertheless, by implementing this modification in the MS-Tag program, the peptide sequence was identified readily, thus all the major peptides were matched to the same protein.

Modified peptides were also found in band F9. As seen in the gel illustrated in Fig. 12, several unique low molecular weight proteins labeled F9 to F12 were pulled down using PrP20 as bait, contrasting with the larger proteins pulled

down by other baits. PrP20, also called RCC1 in vertebrates, is guanine nucleotide exchange factor, responsible for generating a Ran-GTP gradient across the nuclear envelope to facilitate karyopherin-mediated transport. From the MALDI-TOF MS spectrum of the unseparated tryptic digest from band F9 in Fig. 13, 24 peptide masses were submitted to MS-Fit for database searching. Once again, no more than four peptides matched to any one protein. CID analysis of peptides with MH^+ at 788.47 and 1004.57 (Fig. 14) identified the major protein in this band as yeast histone H3, defining their sequences as KLPFQR and YKPGTVALR, respectively. Based on masses alone, two other peptides at 831.49 and 1355.70 were consistent with this match with mass errors less than 10 ppm, but several other abundant peaks did not match the same protein. The MALDI-TOF MS spectrum of band F9 showed the peak at m/z 1335.70 to be the first in a cluster of four peaks, each separated by 14 Da, i.e. 1335.70, 1349.71, 1363.73, and 1377.73 as illustrated in the expansion in Fig. 13. Although the first of these peptides was relatively weak and was not selected for CID, in histone H3 the mass corre-

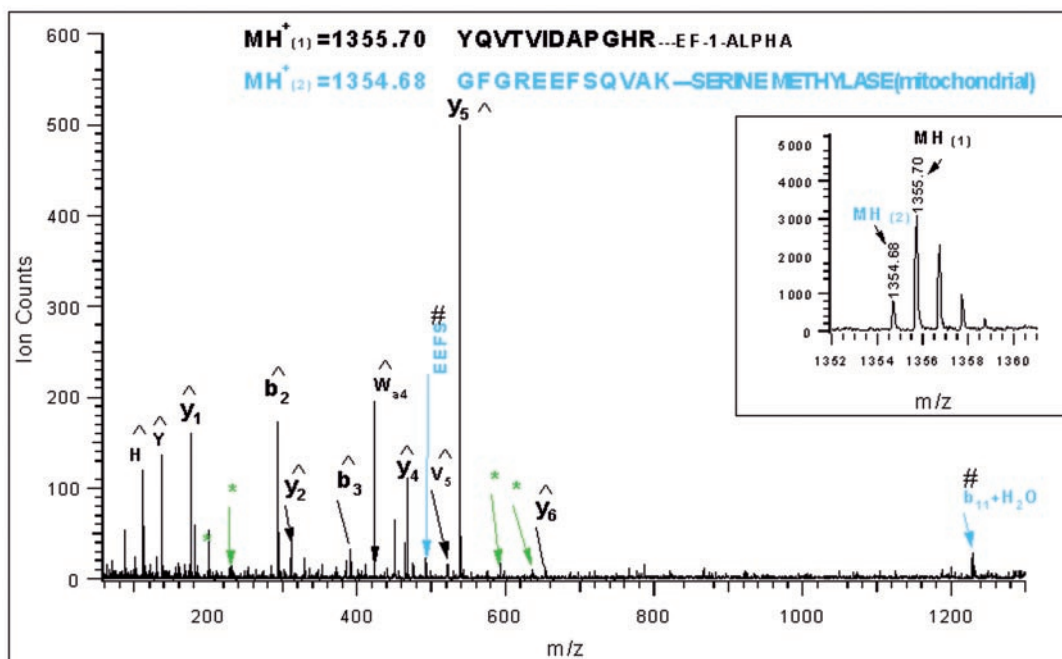


FIG. 9. MALDI-TOF/TOF CID spectra of two tryptic peptides ($MH^+_{(1)} = 1355.7$ and $MH^+_{(2)} = 1354.7$) from band G11 digest (shown in Fig. 6) 1 Da apart. The fragment ions marked were as follows: *, matched to both peptides; ^, only matched to the first peptide $MH^+_{(1)} = 1355.7$; #, only matched to the second peptide $MH^+_{(2)} = 1354.7$. The peptide sequences were determined as YQVTVIDAPGHR, matched to EF-1- α and GFGREEFSQVAK, matched to serine methylase, mitochondrial.

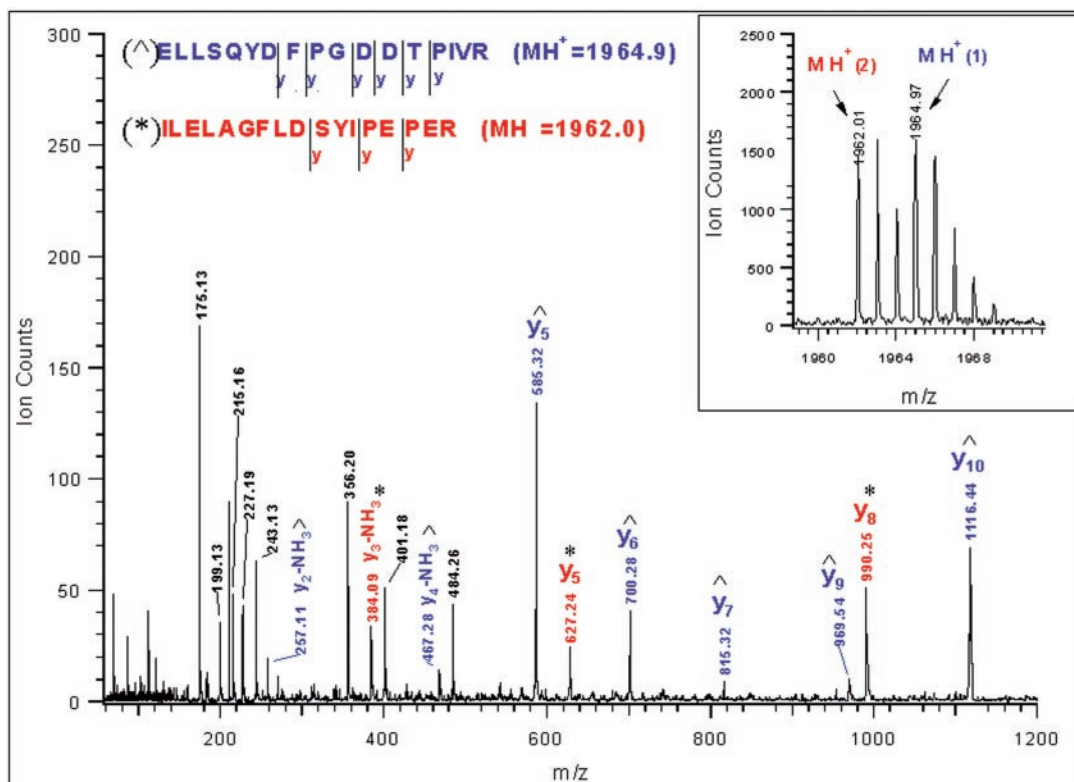


FIG. 10. MALDI-TOF/TOF CID spectra of two tryptic peptides ($MH^+_{(1)} = 1964.9$ and $MH^+_{(2)} = 1962.0$) from band C11 on plate 1. MS-Tag search results revealed that both peptides matched to elongation factor Tu. The sequences were determined as ELLSQYDFPGD-DTPIVR, $MH^+ 1964.9$; ILELAGFLDSYIPEPER, $MH^+ 1962.0$. The peaks labeled were as follows: ^, only matched to peptide (1); *, only matched to peptide (2). The rest of the peaks were matched to both sequences.

TABLE I

Detailed MS-Fit Results of Band G11 from plate 2

#1 20/66, ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA), Acc. # P02994, 50033.1 Da, pI = 9.14				#2 8/66, RAN GTPASE ACTIVATING PROTEIN 1, Acc. # P11745, 45816.2 Da, pI = 4.53.			
m/z	Delta	Peptide Sequence	Modifications	m/z	Delta	Peptide Sequence	Modifications
submitted	ppm			submitted	ppm		
894.4750	-20.8591	(R)YIEKFEK(E)		1088.5800	-38.2619	(K)LYQNGIRPK(G)	
894.4750	2.7626	(K)YAWVLDK(L)		1447.7943	-15.0261	(R)LENGSAVYLALGLK(S)	
894.4750	14.5551	(R)FAVRDMR(Q)		1575.8764	-8.5594	(K)GVATLIHYGLOYLK(N)	
914.5622	-5.7863	(R)QTVAVGVK(S)		1596.8233	1.5513	(K)AASKPFLETFICGR(N)	
1025.6147	3.8477	(K)JGGIGITVPVGR(V)		1728.8540	2.7979	(K)HLILSNNGMGPFAGER(I)	1Met-ox
1135.6346	-14.9306	(K)YAWVLDK(L)		1742.8590	-0.6339	(R)ESLVEVNFADLYTSR(L)	
1278.6616	22.3635	(R)FDELLEKND(R)		1954.9899	4.6009	(K)CPHLEIVNLSDNAFGLR(T)	
1326.7496	-2.8605	(R)EHALLAFTLGVR(Q)		2020.0865	0.5380	(K)LTSDDIKPYLEELAALK(T)	
1355.7010	-4.5536	(K)YQVTVIDAPGHR(D)		#3 7/66, CYTOSOLIC SERINE METHYLASE Acc. # P37291.52218.9 Da, pI = 6.98.			
1436.7081	-6.4756	(K)WDESRFQEI(VK)(E)		m/z	Delta	Peptide Sequence	Modifications
1560.8477	-1.3549	(K)SHINVVVIGHVDSGK(S)		submitted	ppm		
1624.8696	-0.1733	(K)ETSNFIKQVGYNPK(T)		857.4360	1.9018	(R)GAMIFFR(R)	1Met-ox
1806.0058	-0.1096	(R)FAVRDMRQTVAVGVK(S)	1Met-ox	999.4543	0.7160	(K)YSEGYPGAR(Y)	
1858.9897	4.3802	(K)YQVTVIDAPGHRDFIK(N)		1094.6512	33.5653	(R)LIDYKRM(R)(E)	
1997.0273	23.5673	(K)MDSYKWDSESRFQEI(VK)(E)		1196.6167	5.7935	(K)VLVAGTSAYCR(L)	
2311.1263	0.3372	(K)FVPSKPMCVEAFSEYPLGR(F)		1223.5596	12.3210	(R)YGGNEHIDR(M)	
2327.1222	0.7583	(K)FVPSKPMCVEAFSEYPLGR(F)	1Met-ox	1666.8130	0.7811	(K)ISAVSTYFSPYR(V)	
2363.2918	-0.6977	(K)TLLEAIDAIEQPSRPTDKPLR(L)		2637.3625	0.0696	(K)LITSHLVDTDPEVDSIIKDEIER(Q)	
2430.3147	-7.0395	(R)YETGVKIPGMVVTFAPAGVTEVK(S)		#4 7/66, Mitochondrial SERINE METHYLASE Acc. # P37292.53660.8 Da, pI = 8.83.			
2446.3121	-5.9775	(R)YETGVKIPGMVVTFAPAGVTEVK(S)	1Met-ox	m/z	Delta	Peptide Sequence	Modifications
2566.2065	2.6791	(K)SVMEMHHEQLEQVPGDNGVFNK(N)	1Met-ox	submitted	ppm		
3319.8657	11.7877	(K)TLLEAIDAIEQPSRPTDKPLRLPLQDVYK(I)		857.4360	1.9018	(R)GAMIFFR(K)	1Met-ox
				947.5361	4.9172	(K)SALFPSGLR(I)	
				1057.4673	7.7888	(K)YSEGYPGER(Y)	
				1354.6800	3.2932	(R)GFGREEFQVAK(Y)	
				1854.9930	-7.9633	(K)VIVAGTSAYSRLIDYAR(F)	
				2102.9809	-0.3772	(R)YGGNEHDKSESLSQAR(A)	
				2437.2691	0.1679	(K)LYSGGTDNHLVIDLSGTQVDGAR(V)	
#5 9/66, MITOCHONDRIAL PROTEIN IMPORT PROTEIN MASS (PROTEIN YDJ1), Acc. # P25491, 44671.1 Da, pI = 5.95.				#6 6/66, Mitochondrial SERINE METHYLASE Acc. # P37292.53660.8 Da, pI = 8.83.			
m/z	Delta	Peptide Sequence	Modifications	m/z	Delta	Peptide Sequence	Modifications
submitted	ppm			submitted	ppm		
1094.6512	-5.6361	(K)KLEELPPR(I)		2008.0939	-3.3710	(K)FYDILGVPVTATDVEIKK(A)	
1394.7748	-5.8653	(K)YGVIVPGEVIAPGMR(K)		2081.0300	-2.1051	(K)FTIKFPENHFTSEENLK(K)	
1410.7703	-5.3839	(K)YGVIVPGEVIAPGMR(K)	1Met-ox	2575.3158	0.0618	(R)GEADQAPDVIPGDVVFVYSERPHK(S)	
1678.8211	-4.5893	(K)EASAAEYELSDPEKR(D)		3062.6276	-1.2720	(R)IVFKGEADQAPDVIPGDVVFVYSERPHK(S)	
1953.9844	-3.9824	(K)FKEASAAEYELSDPEKR(D)					

51/66 Matched to 5 proteins!

sponds to a normal tryptic peptide EIAQDFKTDLR. Histone H3 can be methylated or acetylated at lysine residues (36), and the presence of a single lysine in this sequence is consistent with the 14-Da differences being because of single or double methylation or triple methylation/acetylation. The y ion series in the TOF/TOF CID spectra of the modified peptides displayed in Fig. 15 confirm that the modification is indeed localized on the lysine. The masses of the first four members of this series, y₁ (175), y₂ (288), y₃ (403), and y₄ (504), are the same in all three CID spectra, whereas from y₅ up they show systematic differences of 14 or 28 Da. In Fig. 15A the y₅ ion occurs with loss of ammonia (646-17 = 629), whereas the intact y₅ ions in Fig. 15, B and C occur at 660 and 674, and the y₆ ions at 793, 807 and 821 support the same conclusion. The b ions are less definitive, but b₂ (243) and b₃ (314) are the same in all spectra whereas strong b₁₀ ions at 1175, 1189, and 1203 show the expected mass shifts. Thus, the three peptides all correspond to lysine-modified forms of the peptide of MH⁺ 1355.70, arising from digestion of modified his-

tone H3. The results suggest that the two peptides of MH⁺ 1349.7 and 1363.7 are the modified peptides with mono- and dimethylated lysine, respectively. Based on the peptide mass, two possible modifications, i.e. trimethylation and acetylation, could occur on the lysine residue in the peptide of MH⁺ 1377.73, because the modified mass increments are isobaric, 42 Da. The peptide mass error is -6.9 ppm if lysine is trimethylated, whereas the mass error is 19.5 ppm if the lysine is acetylated. Even though all the peptides matched to histone H3 gave mass errors less than 10 ppm, suggesting the modification might be trimethylation, the possibility of acetylation could not be ruled out. However, the obvious difference between these two modifications is that trimethylated lysine carries a positive charge, whereas acetylated lysine is neutral. This chemical property would affect the fragmentation of the peptide dramatically in high energy CID. Biemann (37) has shown that a peptide with a positive charge at the N terminus causes the most side chain fragmentation, giving mainly d, a, w, v, and y ions, whereas a peptide with acetylation at the N

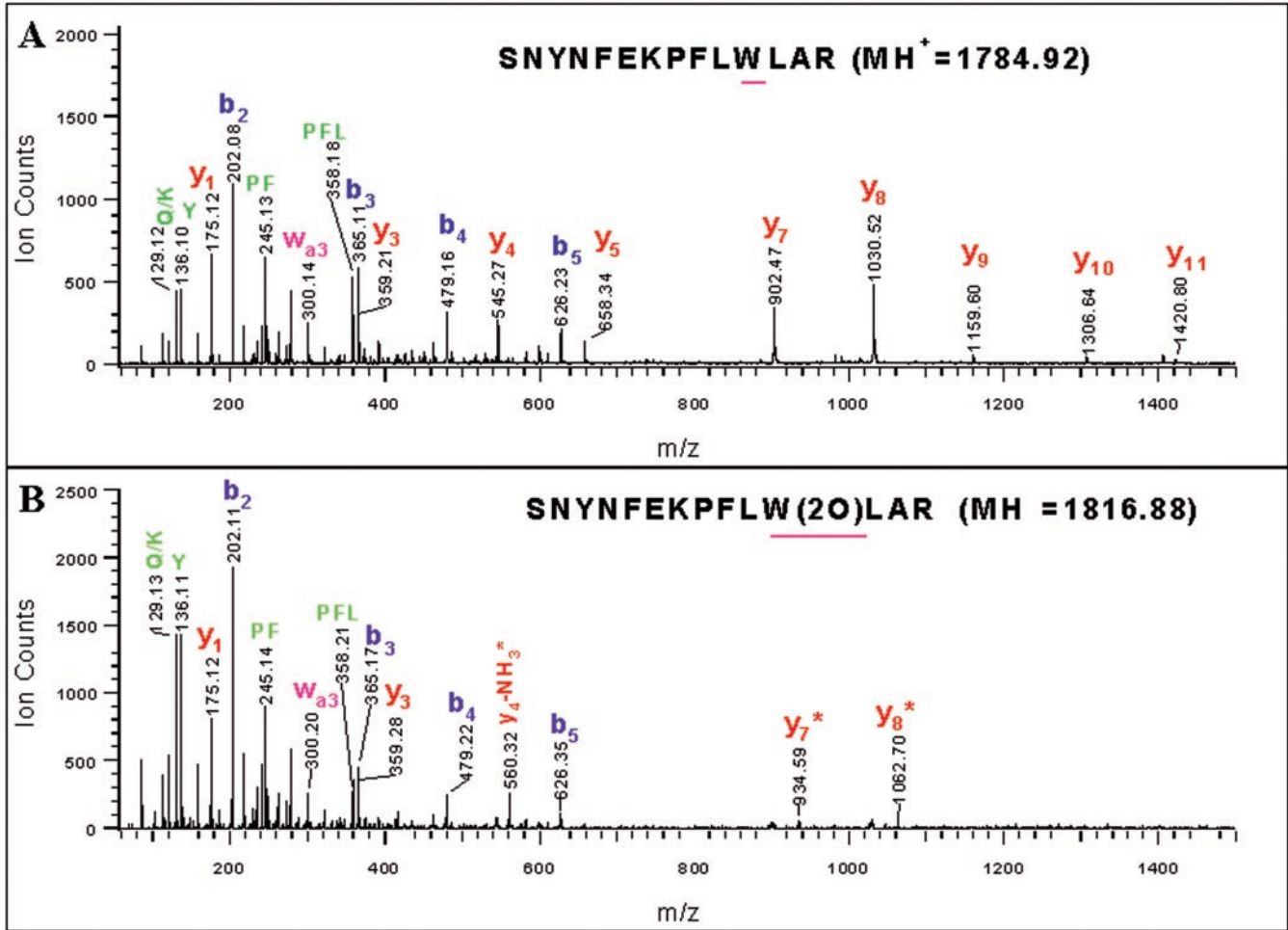


FIG. 11. Comparison of MALDI TOF/TOF CID spectra of two tryptic peptides from band B8 on plate 2. A, MH^+ 1784.9; B, MH^+ 1816.9. The second peptide corresponds to a modified form of the first sequence with oxidized tryptophan.

terminus causes very little side chain fragmentation, giving mainly **y** and **b** ions. Therefore, if this peptide of MH^+ 1377.73 is trimethylated, the fragmentation of this peptide would be different from the other two modified peptides (mono- and dimethylated), because they do not carry any charge. On the contrary, if this peptide is acetylated, the fragmentation pattern would be similar to those of the other two modified peptides. Comparing the three CID spectra in Fig. 15, **d** ions starting at the 8th position from the N terminus right after the modified lysine (at the 7th position), *i.e.* **d**_{a8}, **d**_{b8}, **d**₉, and **d**₁₀ were only observed in the CID spectrum of 1377.73 (Fig. 15C), whereas **w** ions up to the 5th position from the C terminus right before the modified lysine (at the 6th position), *i.e.* **w**_{a3}, **w**_{a4}, **w**_{a5}, and **v**₄ were only observed in Fig. 15, A and B. Furthermore, two fragment ions with higher intensity than the molecular ion, *i.e.* 1221.9 and 1318.9, were attributed to **b**₁₀ + H₂O and $MH^+ - 42 - 17$, which only occurred in Fig. 16C. Thus, these results suggest that the peptide of MH^+ 1377.73 most likely contains a trimethylated lysine residue. Furthermore, accurate mass measurement was also performed using

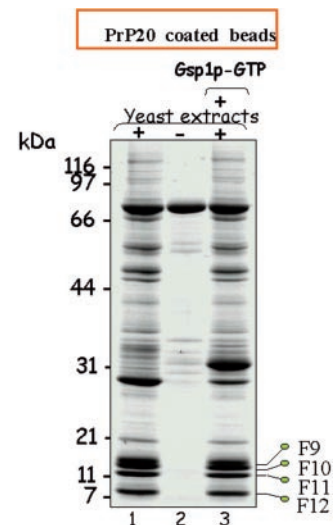


FIG. 12. 1-D gel picture of PrP20 interacting proteins. -, control; +, yeast extract; Gsp1p-GTP, yeast extract spiked with Gsp1p-GTP. Low molecular mass bands (7–15 kDa) labeled as F9, F10, F11, and F12 were selected for analysis.

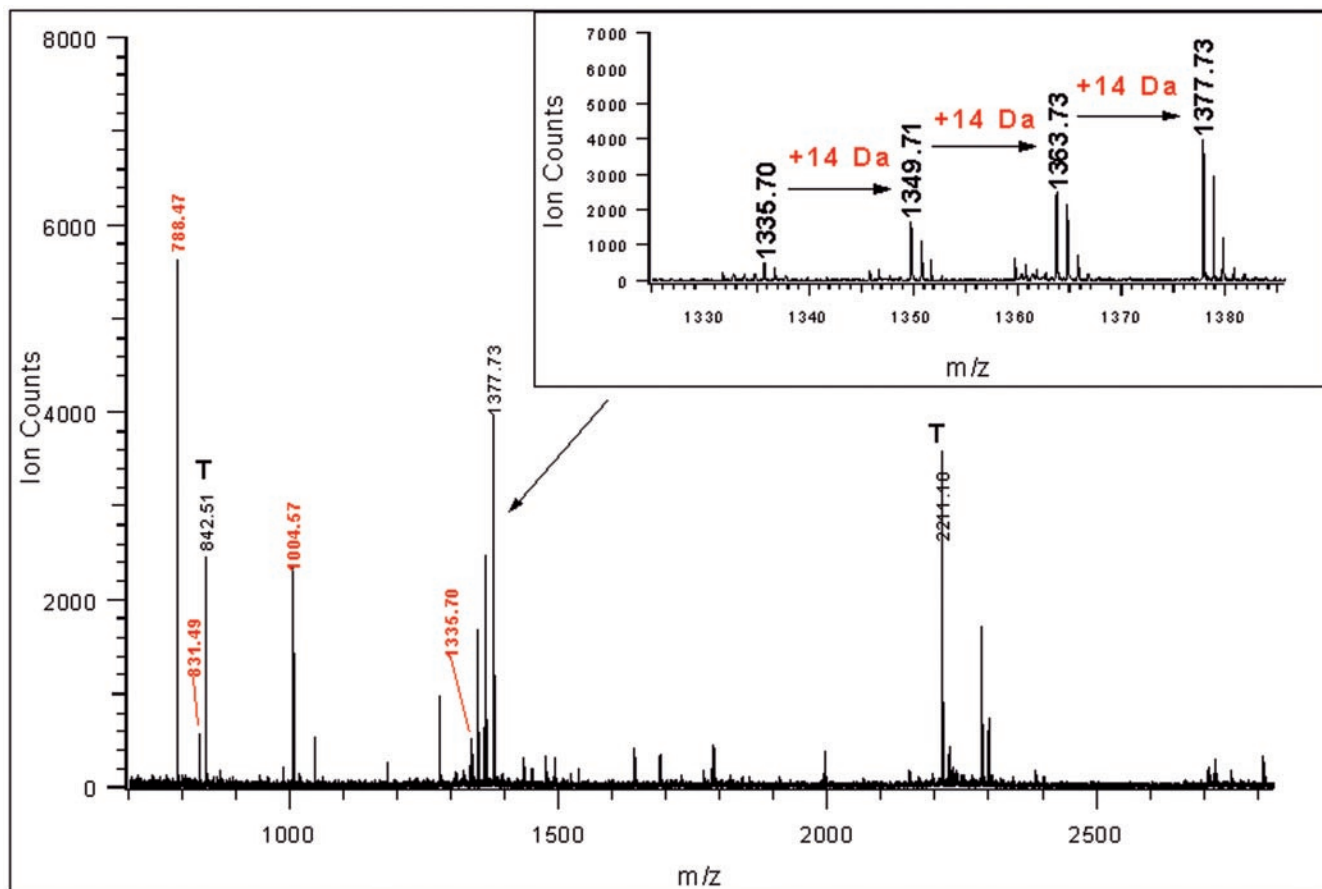


FIG. 13. MALDI-TOF MS spectrum of the unseparated tryptic digest of band F9 using PrP20 as bait. *T* represents trypsin autolysis products. The labeled peaks were analyzed by CID. The inset shows the mass region from 1325 to 1385, where adjacent peptides each differ by 14 Da.

either a single point (y_1 (175)) or two points (y_1 and MH^+) for internal calibration. The mass errors of fragment ions containing modified lysine were always less (~ 30 ppm) if lysine was trimethylated rather than acetylated. This further supports the proposal that the lysine in this peptide is trimethylated. It is believed that the formation of methylated lysine as catalyzed by histone methyltransferase in histone H3 at specific sites (36) may be associated with various biological processes ranging from transcriptional regulation to epigenetic silencing via heterochromatin assembly (38, 39). Analysis of bands F10, F11, and F12 identified them as histone H2B, H2A, and H4, respectively. Recently it was discovered that PrP20 (RCC1) associates with chromatin assembly through direct binding to mononucleosomes and histone H2A and H2B (40). The binding of PrP20 (RCC1) to nucleosomes or histones stimulates the catalytic activity of PrP20, generating the Ran-GTP gradient across the nuclear pore complex, which is essential for nuclear envelope assembly, nuclear transport, and other nuclear events. Further experiments need to be carried out to elucidate the exact role of the histone family members and their modifications in the nucleocytoplasmic transport mechanism.

Non-tryptic Cleavages—The presence of peptides formed by nonspecific digestion that does not obey the normal enzyme cleavage rules complicates protein identification by database searching using either peptide mass fingerprinting or identification of sequence tags. Although band B7 on plate 2 was found by both peptide mapping and peptide sequencing to contain a mixture of Ran-specific GTPase-activating protein and glyceraldehyde-3-phosphate dehydrogenase 3, a peptide of MH^+ 1596.72 did not match either of these proteins based on the MS-Fit search results. However, TOF/TOF CID analysis showed its sequence to be (D)PFIT-NDYAAymFK, formed by a nonspecific cleavage of glyceraldehyde-3-phosphate dehydrogenase 3, the N terminus being formed by digestion of an Asp-Pro bond as shown in Fig. 16. It is not clear whether this bond cleavage was caused by nonspecific enzymatic cleavage or chemical hydrolysis, because the Asp-Pro bond is known to be labile in acid. To identify such a peptide using the sequence tag approach, it is necessary to relax the normal enzyme cleavage rules. In general this can only be done on a limited basis as it greatly extends the search times. Note also that the presence of a methionine sulfoxide residue in this peptide is apparent from the loss of 64 Da.

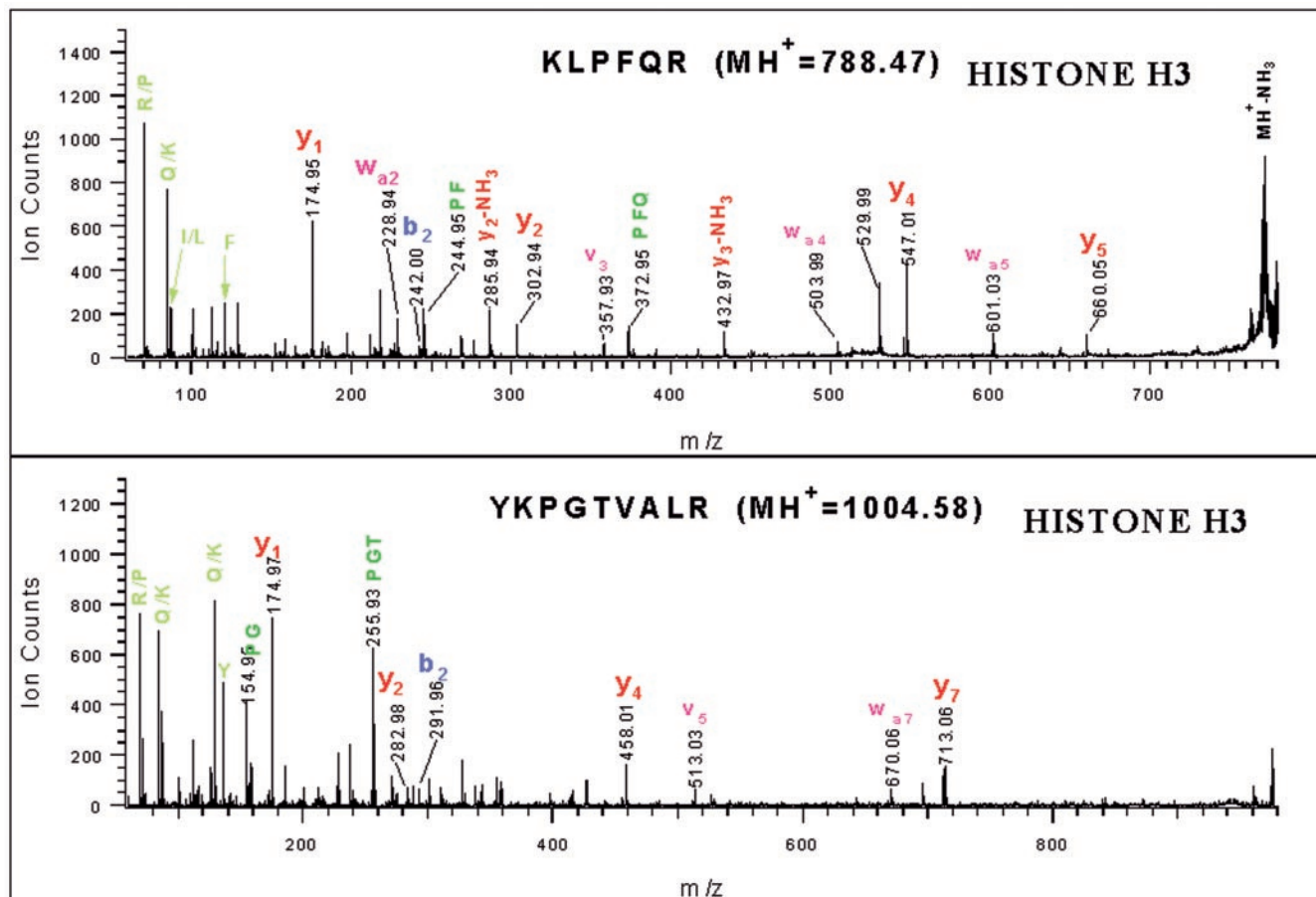


FIG. 14. MALDI TOF/TOF CID spectra of two tryptic peptides from band F9. A, $MH^+ = 788.5$; B, $MH^+ = 1004.6$. MS-Tag searches of these two peptides determined the protein identity as yeast histone H3.

CONCLUSIONS

This proteomic study sought to identify protein-protein interactions based on proteins in yeast extracts that associate with GST-fusion protein baits. As the yeast genome has been sequenced completely, all protein sequences should be present in the databases, therefore it might be anticipated that protein identification by peptide mass fingerprinting should have the maximum probability for success, and the requirement for peptide sequence data from CID experiments should be minimal. The fact that this was not the case in this study was because of the complexity of the mixtures, the occurrence of nonspecific cleavages, and the presence of modified proteins. Therefore tandem mass spectrometry proved to be an essential tool to resolve ambiguities reliably and rapidly. Here the combination of robotic sample preparation and analysis using the MALDI-TOF/TOF MS demonstrated their suitability for high throughput proteomic analysis. In this study MALDI-TOF spectra were collected on a separate instrument, and the tandem mass spectrometry was carried out manually over a two-day period. Nevertheless, it demonstrated the potential for realizing rapid MS and MS/MS using a single instrument giving comprehensive sequence information, high

sensitivity, and high mass accuracy. Furthermore, the MALDI-TOF/TOF is now equipped with software for automated collection of both MS and MS/MS spectra (41, 42). In the example described here, some 200 protein bands from the 288 selected samples were characterized by peptide mass mapping and peptide sequencing using the MALDI-TOF/TOF mass spectrometer. Many fractions contained overlapping proteins, and a total of 91 unique proteins were identified. Even though these proteins were all present in the databases, in many instances peptide sequencing was essential for unambiguous protein identification, especially for protein mixtures, proteins present only in low abundance, and modified proteins. The high quality MS and MS/MS spectra obtained using the MALDI-TOF/TOF MS proved to be suitable not only for database searching but also for *de novo* peptide sequencing for unknown proteins and the investigation of chemical modifications. Furthermore, high collision energy fragments generated by TOF/TOF tandem mass spectrometry could be used for more detailed structure determination, such as distinguishing between Ile and Leu.

It is anticipated that future automated operation of this instrument system with a multiplate cassette will increase

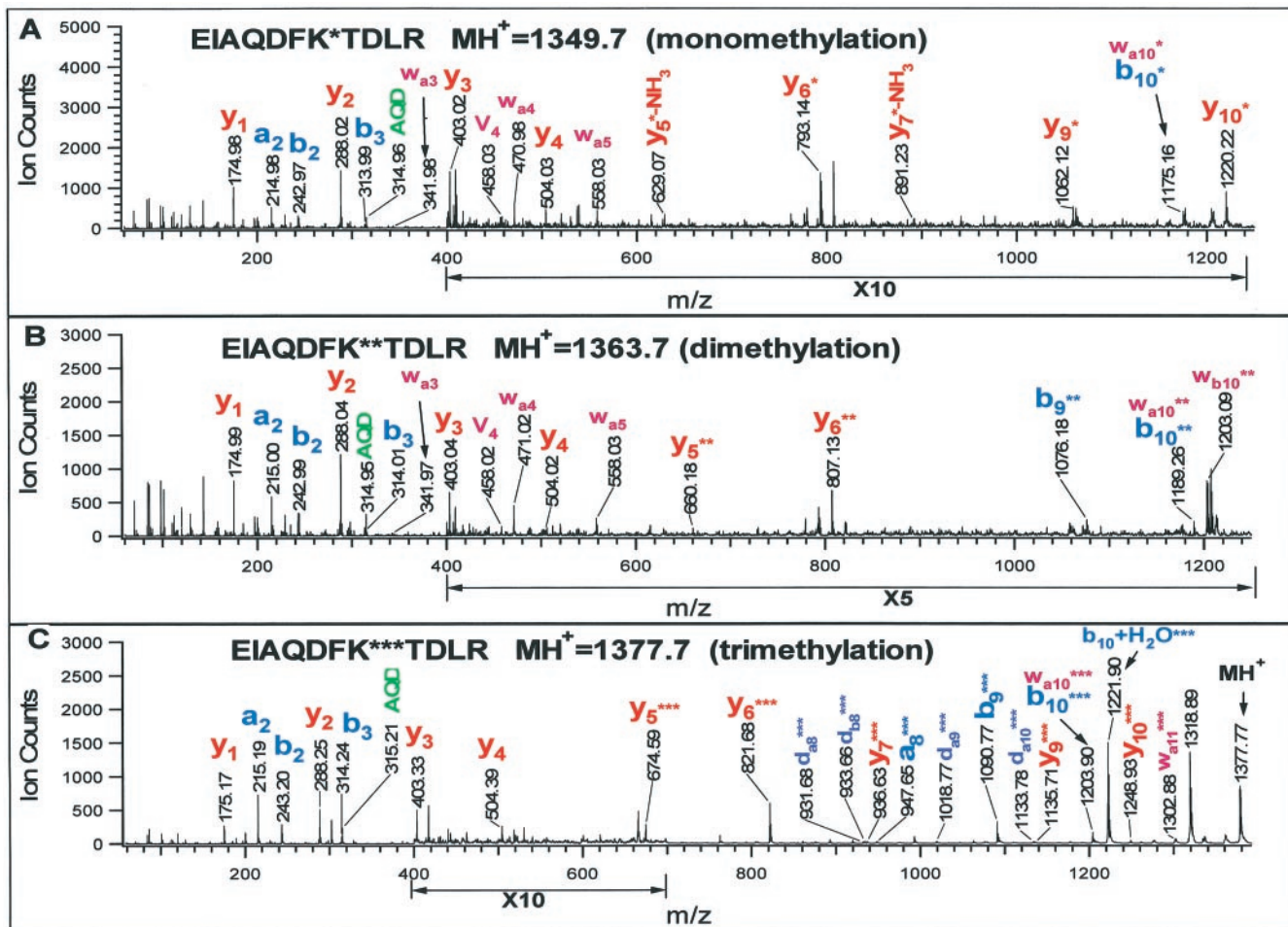
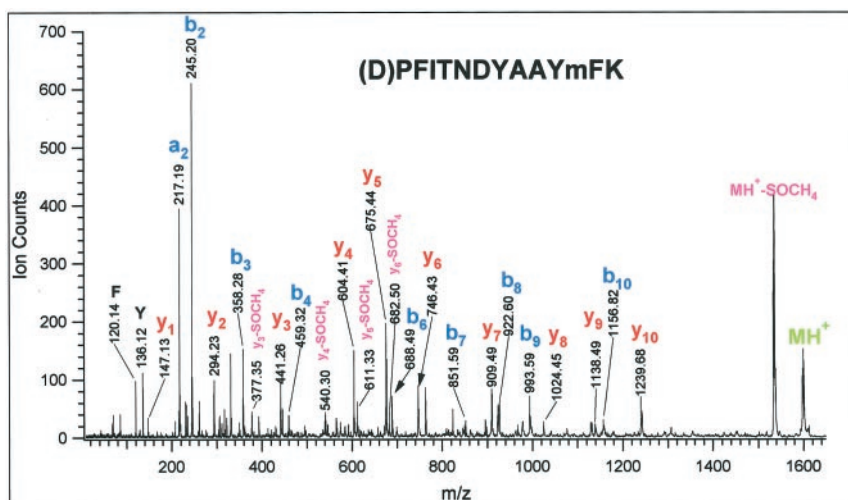


FIG. 15. Comparison of MALDI TOF/TOF CID spectra of three peptides from band F9 14 Da apart. A, $MH^+ = 1349.7$; B, $MH^+ = 1363.7$; C, $MH^+ = 1377.7$. The sequences were determined as EIAQDFK*TDLR with Lys modified as monomethylated (A), dimethylated (B), and trimethylated (C).

FIG. 16. MALDI-TOF/TOF CID spectrum of a peptide ($MH^+ 1596.7$) from band B7 on plate 2. The sequence was determined as (D)PFITNDYAA m FK, where m is oxidized methionine and the N terminus is a nonspecific cleavage site. The sequence matched to yeast glyceraldehyde-3-phosphate dehydrogenase 3.



greatly the capacity to map and identify some thousands of protein digests per day. However, in addition to the tandem mass spectrometer, many other components are critical to

the success of an integrated system for high throughput identification of proteins separated by gel electrophoresis. These include enhancements to the sample handling techniques of

gel cutting, digestion, sample cleaning and spotting, and to minimize sample loss and contamination, including effecting improvements to the robotic system to handle the very limited amounts of sample available for many critical studies. Software must be evaluated and optimized for MALDI MS and MS/MS spectral acquisition with automatic mass calibration, isotope stripping, and peak picking over the whole mass range. The data must be submitted for automatic database searching using both peptide mapping and sequencing, and feedback control is needed to ensure the efficient use of the samples and the instrument to identify the maximum number of proteins in any sample with the minimum effort. Further enhancements to ProteinProspector are required to optimize automated database searching tools suitable for protein mixtures, peptide mixtures, protein modifications, and peptides with nonspecific cleavages. Finally, it is important to develop a sample and data management system for sample tracking, data archiving, processing, and creation of output reports. Only then will automated high throughput proteomic analysis become a realistic possibility. Achieving this high capacity will permit studies of the temporal modulation of protein expression and modification of protein expression in response to both endogenous cell cycle signals and exogenous events such as heat shock or exposure to drugs.

* This work was supported in part by NCRR, National Institutes of Health Grant RR 01614 (to A. L. B.), NCRR, National Institutes of Health Shared Instrumentation Grant S10RR14606 (to A. L. B.), and NCI, National Institutes of Health Grant R33 CA86135 (to M. Pallavicini). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

** To whom correspondence should be addressed. Tel.: 415-476-5641; Fax: 415-476-0688; E-mail: alb@itsa.ucsf.edu.

REFERENCES

- Jensen, O. N., Podtelejnikov, A., and Mann, M. (1996) Delayed extraction improves specificity in database searches by matrix-assisted laser desorption/ionization peptide maps. *Rapid Commun. Mass Spectrom.* **10**, 1371–1378
- Clauser, K. R., Baker, P., and Burlingame, A. L. (1999) Role of accurate mass measurement (± 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal. Chem.* **71**, 2871–2882
- Jimenez, C. R., Huang, L., Qiu, Y., and Burlingame, A. L. (1998) MALDI mass analysis of peptides and proteins. *Curr. Prot. Protein Chem.* **16.3.1–16.3.6**
- Mann, M., Hendrickson, R. C., and Pandey, A. (2001) Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.* **70**, 437–473
- Page, M. J., Amess, B., Townsend, R. R., Parekh, R., Herath, A., Brusten, L., Zvelebil, M. J., Stein, R. C., Waterfield, M. D., Davies, S. C., and O'Hare, M. J. (1999) Proteomic definition of normal human luminal and myoepithelial breast cells purified from reduction mammaplasties. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 12589–12594
- Washburn, M. P., Wolters, D., and Yates, J. R., III (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **19**, 242–247
- Neubauer, G., Gottschalk, A., Fabrizio, P., Seraphin, B., Luhrmann, R., and Mann, M. (1997) Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 385–390
- Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H., and Mann, M. (1996) Linking genome and proteome by mass spectrometry: large-scale identification of yeast proteins from two dimensional gels. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 14440–14445
- Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A., and Mann, M. (1998) Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat. Genet.* **20**, 46–50
- Rout, M. P., Aitchison, J. D., Suprapto, A., Hjertaas, K., Zhao, Y., and Chait, B. T. (2000) The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* **148**, 635–651
- Allen, N., Huang, L., Burlingame, A. L., and Rexach, M. (2001) Proteomic analysis of nucleoporin interacting proteins. *J. Biol. Chem.* **276**, 29268–29274
- Ho, Y., Gruhler, A., Heilbut, A., Bader, G. D., Moore, L., Adams, S. L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., et al. (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183
- Gavin, A. C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J. M., Michon, A. M., Cruciat, C. M., Remor, M., Hofert, C., Schelder, M., Brajenovic, M., Ruffner, H., et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147
- Shevchenko, A., Schaft, D., Roguev, A., Pijnappel, W. W. M. P., Stewart, A. F., and Shevchenko, A. (2002) Deciphering protein complexes and protein interaction networks by tandem affinity purification and mass spectrometry: analytical perspective. *Mol. Cell. Proteom.* **1**, 204–212
- Wilm, M., Shevchenko, A., Houthaave, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996) Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379**, 466–469
- McCormack, A. L., Schieltz, D. M., Goode, B., Yang, S., Barnes, G., Drubin, D., and Yates, J. R., III (1997) Direct analysis and identification of proteins in mixtures by LC/MS/MS and database searching at the low-femtomole level. *Anal. Chem.* **69**, 767–776
- Spengler, B., Kirsch, D., Kaufmann, R., and Jaeger, E. (1992) Peptide sequencing by matrix-assisted laser-desorption mass spectrometry. *Rapid Commun. Mass Spectrom.* **6**, 105–108
- Kaufmann, R., Splengler, B., and Lutzenkirchen, F. (1993) Mass spectrometric sequencing of linear peptides by product-ion analysis in a reflectron time-of-flight mass spectrometer using matrix-assisted laser desorption ionization. *Rapid Commun. Mass Spectrom.* **7**, 902–910
- Biemann, K., and Scoble, H. (1987) Characterization by tandem mass spectrometry of structural modifications in proteins. *Science* **237**, 992–998
- Medzihradzsky, K. F., and Burlingame, A. L. (1994) The advantages of versatility of a high-energy collision-induced dissociation-based strategy for the sequence and structural determination of proteins. *Methods* **6**, 284–303
- Cohen, S. L., and Chait B. T. (1996) Influence of matrix solution conditions on the MALDI-MS analysis of peptides and proteins. *Anal. Chem.* **68**, 31–37
- Medzihradzsky, K. F., Bateman, R. H., Green, M. R., Adams, G. W., and Burlingame, A. L. (1996) Peptide sequence determination by matrix-assisted laser desorption ionization using a tandem double focusing magnetic/orthogonal acceleration time of flight mass spectrometer. *J. Am. Soc. Mass Spectrom.* **7**, 1–10
- Wang, S. X., Mure, M., Medzihradzsky, K. F., Burlingame, A. L., Brown, D. E., Dooley, D. M., Smith, A. J., Kagan, H. M., and Klinman, J. P. (1996) A crosslinked cofactor in lysyl oxidase redox function for amino acid side chains. *Science* **273**, 1078–1084
- Nakamura, T., Fainzilber, M., and Burlingame, A. L. (1996) Mass spectrometric-based revision of the structure of a cysteine-rich *Conus textile* peptide toxin with γ -carboxyglutamic acid. *Protein Sci.* **5**, 524–530
- Medzihradzsky, K. F., Maltby, D. A., Qiu, Y., Yu, Z., Hall, S. C., Chen, Y., and Burlingame, A. L. (1997) Protein sequence and structural studies employing matrix assisted laser desorption ionization —high energy colli-

- sion-induced dissociation. *Int. J. Mass Spectrom. Ion Process.* **160**, 357–369
26. Jiménez, C. R., Li, K. W., Dreisewerd, K., Spijker, S., Kingston, R., Bateman, R. H., Burlingame, A. L., Smit, A. B., Van Minnen, J., and Geraerts, W. P. M. (1998) Direct mass spectrometric peptide profiling and sequencing of single neurons reveals differential peptide patterns in a small neuronal network. *Biochemistry* **37**, 2070–2076
 27. Vestal, M., Juhasz, P., Hines, W., and Martin, S. (2000) in *Mass Spectrometry in Biology and Medicine*, pp. 1–16, Humana Press, Totowa, NJ
 28. Medzihradzky, K. F., Campbell, J. M., Baldwin, M. A., Falick, A. M., Juhasz, P., Vestal, M. L., and Burlingame, A. L. (2000) The characteristics of peptide collision-induced dissociation using a high performance MALDI-TOF/TOF tandem mass spectrometer. *Anal. Chem.* **72**, 552–558
 29. Denning, D., Mykytka, B., Allen, N., Huang, L., Burlingame, A. L., and Rexach, M. (2001) The nucleoporin Nup60p functions as a Gsp1-GTP sensitive tether for Nup2p at the nuclear pore complex. *J. Cell Biol.* **154**, 937–950
 30. Chook, Y. M., and Blobel, G. (2001) Karyopherins and nuclear import. *Curr. Opin. Struct. Biol.* **11**, 703–715
 31. Gilchrist, D., Mykytka, B., and Rexach, M. (2002) Accelerating the rate of disassembly of karyopherin-cargo complexes. *J. Biol. Chem.* **277**, 18161–18172
 32. Rexach, M., and Blobel, G. (1995) Protein import into nuclei: association and dissociation reactions involving transport substrate, transport factors, and nucleoporins. *Cell* **83**, 683–692
 33. Wen, D. X., Livingston, B. D., Medzihradzky, K. F., Kelm, S., Burlingame, A. L., and Paulson, J. C. (1992) Primary structure of the Gal β 1,3(4)GlcNAc α 2,3-sialyltransferase determined by mass spectrometry and cDNA cloning: evidence for a protein motif in the sialyltransferase gene family. *J. Biol. Chem.* **267**, 21011–21019
 34. Clauser, K. R., Hall, S. C., Smith, D. M., Webb, J. W., Andrews, L. E., Tran, H. M., Epstein, L. B., and Burlingame, A. L. (1995) Rapid mass spectrometric sequencing and peptide mass matching for characterization of human melanoma proteins isolated by two-dimensional PAGE. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 5072–5076
 35. Giessauf, A., van Wickern, B., Simat, T., Steinhart, H., and Esterbauer, H. (1996) Formation of *N*-formylkynurenine suggests the involvement of apolipoprotein B-100 centered tryptophan radicals in the initiation of LDL lipid peroxidation. *FEBS Lett.* **389**, 136–140
 36. Rice, J. C., and Allis, C. D. (2001) Histone methylation versus histone acetylation: new insights into epigenetic regulation. *Curr. Opin. Cell Biol.* **13**, 263–273
 37. Biemann, K. (1990) Sequencing of peptides by tandem mass spectrometry and high-energy collision-induced dissociation. *Methods Enzymol.* **193**, 455–479
 38. Nakayama, J., Rice, J. C., Strahl, B. D., Allis, C. D., and Grewal, S. I. (2001) Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* **292**, 110–113
 39. Rice, J. C., and Allis, C. D. (2001) Code of silence. *Nature* **414**, 258–261
 40. Nemergut, M. E., Mizzen, C. A., Stukenberg, T., Allis, C. D., and Macara, I. G. (2001) Chromatin docking and exchange activity enhancement of RCC1 by histones H2A and H2B. *Science* **292**, 1540–1543
 41. Huang, L., Chalkley, R. J., Baker, P. R., Maltby, D., Allen, N. P., Sihlbom, C., Rexach, M., Baldwin, M. A., and Burlingame, A. L. (2002) *50th ASMS Conference on Mass Spectrometry and Allied Topics, Orlando, FL, June 2–6, 2002*, in press
 42. Chalkley, R. J., Sihlbom, C., Maltby, D., Huang, L., and Burlingame, A. L. (2002) *50th ASMS Conference on Mass Spectrometry and Allied Topics, Orlando, FL, June 2–6, 2002*, in press