Technical brief

Samuel Purvine
Jason-Thomas Eppel*
Eugene C. Yi
David R. Goodlett

The Institute for Systems Biology, Seattle, WA USA

Shotgun collision-induced dissociation of peptides using a time of flight mass analyzer

Parallel collision-induced dissociation (CID) of peptides rather than serial, as is customary, results in loss of the obvious parent-fragment ion lineage available from CID on a single ion. We report proof-of-principle results suggesting the feasibility of parallel peptide CID, referred to here as shotgun CID, for protein identification when using the measured mass accuracies available from a time of flight mass analyzer and currently available search routines such as SEQUEST. Additionally, we report that parent-fragment ion lineage may be reconstructed from information encoded in the chromatographic single ion current traces of peptides.

Keywords: In-source collision-induced dissociation / Parallel peptide tandem mass spectrometry / Peptide mixtures PRO 0362

As an alternative to the 2-DE mass spectrometry (MS) method that has dominated the early days of proteomics, direct analysis by tandem mass spectrometry (MS/MS) of the peptides generated by the digestion of unseparated protein mixtures is in use by a number of laboratories. MS-only based methods are generally referred to as shotgun proteomics or by some as bottom-up proteomics [1–4]. The key feature of these methods is the bulk digestion (enzymatically or chemically) of many proteins simultaneously to peptides followed by multidimensional chromatographic methods to reduce the complexity of the peptide mixture prior to MS/MS sequencing of individual peptides. These shotgun proteomic methods have been most often coupled to microcapillary-HPLC (μLC)-ESI-MS/MS for protein identification rather than peptide mass mapping, because the tandem mass spectrum of a single peptide can be sufficient to identify the parent protein [5]. However, as shown recently [6, 7] these shotgun proteomic techniques that use serial data-dependent (DD) ion selection for CID during μLC introduction of the sample have notable limitations when used to analyze very complex mixtures. The limitations relate both to total proteome coverage in a single pass μLC-ESI-MS/MS experiment and to the reproducibility of detecting the same proteins in replicate analyses.

Correspondence: Dr. David R. Goodlett, The Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA
E-mail: goodlett@systemsbiology.org
Fax: +1-206-732-1299

Abbreviations: DD, data-dependent; **μLC**, microcapillary-HPLC

Incomplete coverage of the peptides, and by inference their parent proteins, in a single pass μLC-ESI-MS/MS analysis of a complex mixture means that proteins known to be present such as ^{32}P-labeled phosphopeptides are not detected [8]. In general there are a two obvious means to increase proteome coverage and reproducibility of the shotgun proteomic methods, but both will require technical breakthroughs. They are to (1) increase the peak capacity of the separation methods prior to the MS/MS analysis, and (2) increase the duty cycle at which the mass spectrometer can select a single ion, perform CID and then move to the next available ion. Given that gains in these two areas are not trivial, we would like to suggest an alternate solution that we demonstrate can work within the limitations of currently available peak capacity by maximizing mass spectrometric duty cycle. The method described herein involves parallel, rather than serial, CID of peptides (Fig. 1) TOF mass analyzer. We refer to it as shotgun CID to distinguish it from multiplex MS [9] carried out on a Fourier transform-ion cyclotron resonance mass spectrometer (FT-ICR MS) where very high mass accuracy and resolving power are available. Additionally, as opposed to multiplex MS where several ions are selected, shotgun CID maximizes MS/MS duty cycle by avoiding ion selection. Here we demonstrate that the shotgun CID method is feasible on an inexpensive TOF-MS instrument and is a viable alternative to the standard serial DD ion selection process.

Total yeast cell extracts were obtained by glass bead lysis described in detail elsewhere [7]. Pellets obtained from total yeast cell lysates were resuspended in 0.5 mL of 50 mM NH_4HCO_3 (pH 8.3), 0.5% SDS and solubilized by

* In memoriam J. T. Eppel (1971–2003)

 0173-0835/03/0606–847 $17.50+.50/0

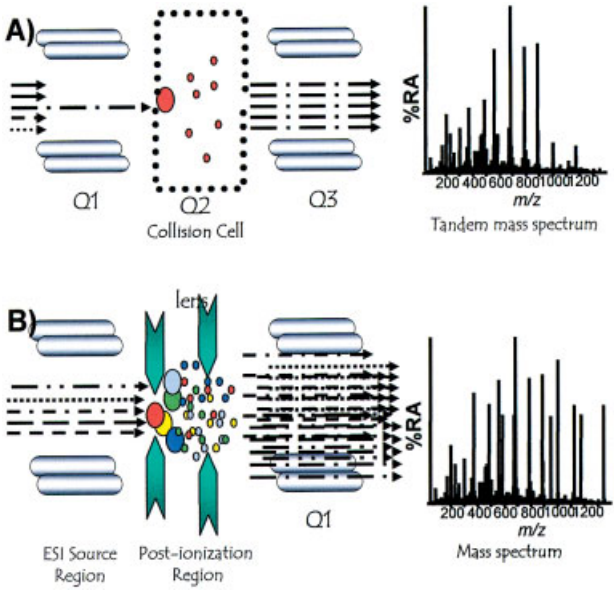

Figure 1. Comparison of serial to parallel MS/MS. Comparison of (A) traditional serial MS/MS where all fragment ions are presumed to be derived from CID of a single parent to (B) shotgun CID where peptides are subjected to CID in parallel and fragment ion spectra contain the fragments of multiple parent ions. Note that while shotgun CID (B) is shown occurring in-source it can also be carried out in a collision cell.

incubating for 30 min at 60°C with occasional vortexing. The resulting mixture was diluted to have 0.05% SDS with 50 mM NH_4HCO_3 (pH 8.3), modified trypsin (Promega, Madison, WI, USA) added at an enzyme:substrate ratio of 100:1 and incubated overnight at 37°C. Prior to LC-MS analysis resulting peptides were purified by OASIS® MCX (mixed-mode cation-exchange reversed-phase; Waters, Milford, MA, USA) following the manufacturer's protocol. At this point the sample was split and analyzed three times by μLC-MS; once on an ion trap mass spectrometer (IT-MS; ThermoFinnigan, San Jose, CA, USA) and twice on an orthogonal-time of flight mass spectrometer (TOF-MS; Applied Biosystems, Framingham, MA, USA). For μLC-MS analysis, peptide mixtures were injected onto a C18 trap cartridge (Michrome Bioresources, Auburn, CA, USA) for cleanup using a FAMOS autosampler (DIONEX, Sunnyvale, CA, USA) and then passed on to a 10 cm × 100 μm id μLC column by initiating linear gradient flow of acetonitrile from an Agilent (Santa Clara, CA, USA) 1100 HPLC. The effluent from the μLC column entered a miniaturized electrospray ionization (μESI) source (Brechbuehler, Spring, TX, USA) in which peptides were ionized and passed directly into either the IT- or TOF-MS instruments. The C18 trap cartridge, μESI-emitter/μLC column combination, a high voltage line for ESI charging and the waste line were each connected to one of four ports on a 4-way union (Upchurch, Oak Harbor, WA, USA) constructed entirely out of PEEK [7].

On the IT-MS instrument ion selection for CID was automated using top-down data-dependent (DD) ion selection of the three most intense ions in a survey scan including a 3 min dynamic exclusion period to prevent re-selection of previously selected ions and repeated continuously throughout the μLC-ESI-MS/MS analysis. For analysis by TOF-MS, the μLC setup was identical to that for the IT-MS, but the sample was analyzed twice using two different nozzle-skimmer voltages (V_{NS}) potentials; low V_{NS} resulted in no detectable fragmentation allowing parent ions to be measured and high V_{NS} caused sufficient fragmentation to induce fragment ions to form in the source region. These two analyses were necessary, because the Mariner ESI-TOF-MS could not switch between low and high V_{NS} at a rate that was compatible with maintaining chromatographic integrity. Future implementation of the method will be carried out by continuously alternating between low and high V_{NS} in a single μLC experiment. Tandem mass spectrometric results from the IT-MS/MS analysis were analyzed by SEQUEST against a yeast nonredundant sequence database to generate protein identifications [10]. In contrast there was no software for analysis of shotgun CID data. Thus, the shotgun CID data was analyzed on a limited basis by manually creating dta files for SEQUEST to use.

From the high V_{NS} μLC-ESI-TOF-MS analyses, a one-minute window of chromatographic time was arbitrarily chosen for examination. Figure 2A shows the summed shotgun CID mass spectrum from that one minute of chromatographic time. The two ions labeled I and II (Fig. 2A) were confirmed as parent ions from an identical analysis carried out at low V_{NS} μLC-ESI-TOF-MS where no fragmentation was observed. From this shotgun CID mass spectrum (Fig. 2A) two dta files were created for use in a SEQUEST database search. Each of the two dta files were identical except for the entry of the parent ion mass; *i.e.* either 1143.9 for the ion labeled I or 1289.0 for the ion labeled II was used. Both dta files had identical sets of fragment ions that included all ions present above a 5% relative abundance threshold. To identify the potential peptide sequences, each of the dta files was submitted to SEQUEST for a database search against the same yeast sequence database used for the IT-MS/MS data. The searches returned two peptide sequences, VINDAFGIEEGLMTTVHSLTATQK for peptide I and MIEIMLPVFDAPQNLVEQAK for peptide II (Fig. 2B), one of which was not identified in the single pass IT-MS/MS data set. This simple and limited data analysis demonstrates that (1) shotgun CID data can be used to

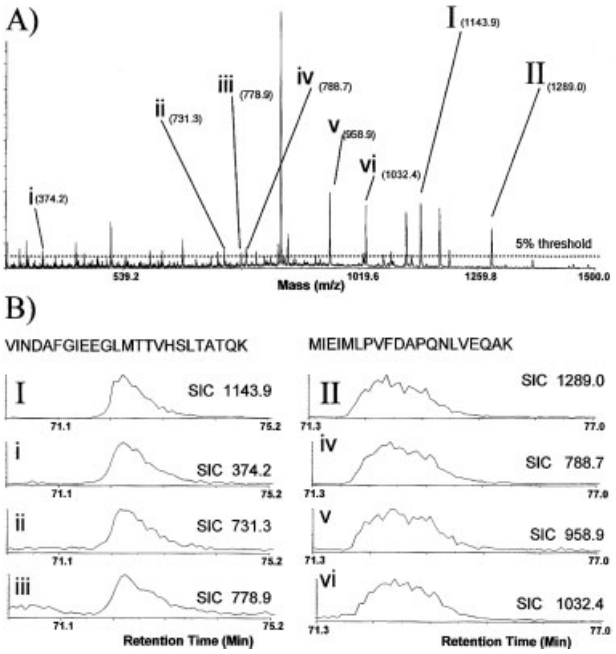

Figure 2. Example results from shotgun CID of yeast peptides. From a single analysis of peptides produced by enzymatic digestion of a whole cell yeast lysate by μLC-ESI-TOF-MS at a nozzle-skimmer voltage known to induce fragmentation of peptides in-source, one minute of chromatographic time was summed (A) to produce a mass spectrum containing multiple parent (I and II) and fragment (i, ii, iii, iv, v, vi) ions. Eight single ion current (SIC) traces are shown (B) demonstrating that parent-fragment ion lineage may be determined by matching SIC peak shape of parents to those of suspected fragment ions.

identify peptide sequence in a database on a manual basis with the SEQUEST database search engine; and (2) shotgun CID can, in a single pass μLC-ESI-MS analysis, identify peptide sequences that were missed due to the random and top-down nature of serial DD ion selection commonly employed in proteomic analyses.

One major problem with the data produced by shotgun CID is the resultant loss of parent-fragment ion lineage when multiple parent ions are fragmented together. While the above database search approach worked with this ideal and simple case, it might not always allow peptide sequences to be identified. Thus, it would be advantageous to be able to sort parent-fragment ion lineage by a method independent of the database search for the purpose of creating less complicated dta files. One obvious approach (data not shown) to sort parent-fragment ion lineage would be provided when analyzing samples labeled by one of the differential isotopic methods such as the isotope coded affinity tag method [4]. In these cases related fragment ions could be sorted due to their common isotopic pattern and knowledge of amino acid residue masses [11, 12]. However, for the shotgun CID method to be of general utility a generic means to sort parent-fragment ion lineage was sought.

During the course of manual analysis of the shotgun CID data such a method was discovered. It was observed (Fig. 2B) that parent-fragment ion lineage could be tracked by comparison of the chromatographic profile of the single ion current (SIC) trace for an individual parent ion to the SIC traces of potential fragment ions. For example in Fig. 2B, the SIC for the parent ion indicated by I with $[M+H]^+$ = 1143.9, has an SIC profile that matches the SIC traces marked as i, ii and iii, but not iv, v and vi. Likewise, the SIC trace for the parent ion marked as II with $[M+H]^+$ = 1289.0 matches the SIC traces for the ions marked iv, v and vi, but not those i, ii and iii. That the sorting of parent and fragment ions by this method was correct was confirmed by looking at the theoretically determined b- and y-ions from the two peptides identified in Fig. 2B. Also, this chromatographic method of sorting parent-fragment ion lineage could be of more general use for analysis of traditional tandem mass spectra where parent-fragment ion lineage is assumed based on "apparent" selection of a single parent ion but not always correct [13].

In conclusion we have presented the results of a feasibility study that suggest parallel CID of peptides, *i.e.* shotgun CID, on a TOF mass analyzer is viable. It can be used to circumvent the low reproducibility of ion picking by serial CID by maximizing MS duty cycle. The method has the single advantage over traditional serial ion selection for CID in that all of the MS duty cycle may be utilized; *i.e.* shotgun CID measures the masses of parent and fragment ions in parallel rather than in series. This should provide a significant advantage when coupled to HPLC separations of complex mixtures of peptides as often they are analyzed by shotgun proteomic methods. The method is not likely to completely replace serial CID of peptide ions but could be used in conjunction with it to more thoroughly characterize a sample. For instance shotgun CID analysis of a sample could be used as a primary screen of the proteins in a sample. Data from that primary screen would be used to generate both confident and degenerate protein identifications. The list of degenerate protein identifications (*i.e.* a list of potential parent proteins for which the shotgun CID results did not result in a single identification) while not the type of results typically provided can be argued to be of great value where sample is limited or fast analyses are important. The degenerate list of protein identifications could be reduced to a single correct identification by (1) generating a list of ions for which traditional serial CID carried out

in a second round of HPLC-MS analysis; and/or (2) the sample could be analyzed with relatively inexpensive assays, such as PCR [5] or Western blotting [14], to determine which of the proteins in the degenerate list is correct.

Finally, while we have demonstrated shotgun CID using in-source CID on a simple ESI-TOF MS instrument, it can also be carried out in a traditional collision cell (data not shown). Currently, the main limitations to practical application of shotgun CID are two-fold: (1) mass analyzers must be capable of measurement mass accuracy in the range of 5–50 ppm at 1000 U or better; and (2) automated software for analysis is not available. Fortunately, any number of currently available instruments (*e.g.* TOF-TOF, QTOF, QIT-TOF) can supply the required measured mass accuracy. To automatically identify parent proteins from shotgun CID data, we are developing software that will combine information from parent ion masses recorded across the entire chromatographic profile (*e.g.* as in common mass mapping approaches) with the fragment ion masses provided by shotgun CID.

The authors thank ISB founders Drs. R. Aebersold, A. Aderem and L. Hood for providing internal funds for exploratory research. More information on the ISB may be found at www.systemsbiology.org.

Received October 21, 2002

References

[1] McCormack, A. L., Schieltz, D. M., Goode, B., Yang, S. *et al.*, *Anal. Chem.* 1997, *69*, 767–776.

[2] Link, A. J., Eng, J., Schieltz, D. M., Carmack, E. *et al.*, *Nat. Biotechnol*. 1999, *17*, 676–682.

[3] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F. *et al.*, *Nat. Biotech.* 1999, *17*, 994–999.

[4] Aebersold, R., Goodlett, D. R., *Chem. Rev.* 2001, *101*, 269–295.

[5] Susin, S. A., Lorenzo, H. K., Zamzami, N., Marzo, I. *et al.*, *Nature* 1999, *397*, 441–446.

[6] Spahr, C. S., Davis, M. T., McGinley, M. D., Robinson, J. H. *et al.*, *Proteomics* 2001, *1*, 93–107.

[7] Yi, E. C., Lee, H., Purvine, S. O., Aebersold, R., Goodlett, D. R., *Electrophoresis* 2002, *23*, 3205–3216.

[8] Katze, M. G., Kwieciszewski, B., Goodlett, D. R., Blakely, C. M. *et al.*, *Virology* 2000, *278*, 501–513.

[9] Masselon, C., Anderson, G. A., Harkewicz, R., Bruce, J. E. *et al.*, *Anal. Chem.* 2000, *72*, 1918–1924.

[10] Eng, J. K., McCormack, A. L., Yates, J. R. III, *J. Am. Soc. Mass Spectrom.* 1994, *5*, 976–989.

[11] Goodlett, D. R., Bruce, J. E., Anderson, G. A., Rist, B. *et al.*, *Anal. Chem*. 2000, *72*, 1112–1118.

[12] Goodlett, D. R., Keller, A., Watts, J. D., Newitt, R. *et al.*, *Rapid Commun. Mass Spectrom.* 2001, *15*, 1214–1221.

[13] Lee, N., Goodlett, D. R., Marquardt, H., Geraghty, D. E., *J. Immunol.* 1998, *160*, 4951–4960.

[14] von Haller, P., Donohoe, S., Goodlett, D. R., Aebersold, R., Watts, J. D., *Proteomics* 2001, *1*, 1010–1021.