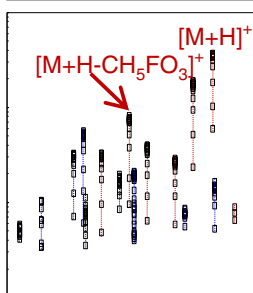CrossMark

RESEARCH ARTICLE

Extending a Tandem Mass Spectral Library to Include MS^2 Spectra of Fragment Ions Produced In-Source and MS^n Spectra

Xiaoyu Yang, Pedatsur Neta, Stephen E. Stein

Mass Spectrometry Data Center, National Institute of Standards and Technology, Mail Stop 8362, Gaithersburg, MD 20899, USA

Abstract. Tandem mass spectral library searching is finding increased use as an effective means of determining chemical identity in mass spectrometry-based omics studies. We previously reported on constructing a tandem mass spectral library that includes spectra for multiple precursor ions for each analyte. Here we report our method for expanding this library to include MS^2 spectra of fragment ions generated during the ionization process (in-source fragment ions) as well as MS^3 and MS^4 spectra. These can assist the chemical identification process. A simple density-based clustering algorithm was used to cluster all significant precursor ions from MS^1 scans for an analyte acquired during an infusion experiment. The MS^2 spectra associated with these precursor ions were grouped into the same precursor clusters. Subsequently, a new top-down hierarchical divisive clustering algorithm was developed for clustering the spectra from fragmentation of ions in each precursor cluster, including the MS^2 spectra of the original precursors and of the in-source fragments as well as the MS^n spectra. This algorithm starts with all the spectra of one precursor in one cluster and then separates them into sub-clusters of similar spectra based on the fragment patterns. Herein, we describe the algorithms and spectral evaluation methods for extending the library. The new library features were demonstrated by searching the high resolution spectra of *E. coli* extracts against the extended library, allowing identification of compounds and their in-source fragment ions in a manner that was not possible before.

Keywords: Tandem mass spectral library, In-source fragmentation, MS^n, Density-based clustering algorithm, Top-down hierarchical divisive clustering algorithm, Metabolite identification

Introduction

Electrospray ionization (ESI) tandem mass spectrometry, using ion-trap or beam-type collision cell instruments, is becoming a routine technique for compound identification in metabolomics, proteomics, lipidomics, and forensic science [1–5]. However, accurate identification of compounds from millions of mass spectra is highly challenging for data analysis [6–11]. Tandem mass spectral library searching matches acquired tandem mass spectra against reference spectra in libraries. This technique can provide an efficient and reliable method for identifying compounds in omics studies [12–18]. Nevertheless, the use of MS^2 spectra alone for identifying a compound may be insufficient because of the lack of specificity of the product ions. Identification may be difficult when multiple compounds share similar fragments in their MS^2 spectra. The use of MS^n spectra, in which the fragmentation pathways can be thoroughly probed, can help in distinguishing between similar compounds. The major peaks in an MS^2 spectrum are selected as precursors and fragmented to produce their MS^3 spectra. The top peaks in these spectra are, in turn, selected to obtain their MS^4 spectra. MS^n spectra with isomer-specific peaks can be used to distinguish between compounds with the same formula but different structures. For example, mass spectral libraries including MS^n were reported to aid identification of some specific types of compounds such as oligosaccharides and plant phenolics [19–24].

Simultaneous with ionization, an analyte may fragment within the ESI source, generating in-source ions, which are observed at the same retention time as the target analyte. In-source fragment ions are often observed in various mass spectrometers [25–27]. Although in-source fragments are the

Correspondence to: Xiaoyu Yang; *e-mail:* xiaoyu.yang@nist.gov

unwanted by-products in MS^1, their MS^2 spectra can be utilized for confirming the compound identification and must not be confused with an analyte having the same precursor mass and product ions. While MS^n spectra are collected in the ion-trap and their products are obtained at a single energy, the MS^2 spectra of in-source fragments are recorded in the collision cell with a number of collision energies. Together, these two measurements provide more robust identification.

The present measurements are carried out with an Orbitrap Elite mass spectrometer, where both an ion trap and a collision cell are used. Supplementing MS^2 spectra of the original compounds with the spectra of their in-source fragments can increase the confidence of metabolite identification in the sample. If an experimental spectrum matches with the spectrum of an in-source fragment, however, it is necessary to determine whether the identified compound really exists in the sample or is formed in the mass spectrometer.

We previously reported constructing a tandem mass spectral library with multiple precursor ion types [12] (such as $[M + H]^+$, $[M + H – H_2O]^+$, $[M + Na]^+$, $[M – H]^-$, $[M + Cl]^-$) . Herein we report on expanding this library with MS^2 spectra of in-source fragments and MS^n spectra of the most abundant fragment ions. Furthermore, consensus spectra were generated with a newly developed top-down hierarchical divisive clustering algorithm. These added features provide further signature peaks for metabolites, drugs, sugars, lipids, and other analytes. This is demonstrated by application of the extended library in the identification of *E. coli* metabolites and matching spectra of in-source fragments as well.

Experimental

Acquisition of Mass Spectra

Each authentic compound was dissolved in a water:acetonitrile (v:v = 1:1) solvent containing 0.1% formic acid and was diluted to a concentration of ≈1 μmol/L. The solution was infused into an Orbitrap Elite mass spectrometer to acquire several types of spectra. Low-resolution collision-induced dissociation (CID) MS^2, MS^3, and MS^4 spectra were acquired on the ion-trap (LTQ) section of the instrument, and high-resolution MS^2 spectra were acquired in the Orbitrap section with fragmentation using either higher-energy collision-induced dissociation (HCD) in the collision cell or CID in the ion trap (IT-FT). Spectra were recorded in the profile mode. The spectra were obtained in a data-dependent mode (using experimental run time of 50 min and exclusion time of 90 s) where MS^2 spectra were acquired for all major ions including in-source fragment ions. In the ion-trap measurements, MS^3 and MS^4 spectra were collected for the top three peaks in the corresponding MS^2 and MS^3 spectra. The resolution of MS^1 and MS^2 spectra acquired in the Orbitrap was set at 30,000.

Data Analysis

The procedure of constructing the tandem mass spectral library to include the MS^2 spectra of in-source fragments and MS^n spectra is summarized in Figure 1.

Clustering Precursors of MS^2 Spectra Each acquired MS^2 spectrum is associated with a precursor ion with a certain *m/z* value. To group the MS^2 spectra associated with the same precursor ion, all the precursor *m/z* values for the same compound, run at the same experimental condition, were collected and clustered using a simple density-based clustering algorithm as follows. First, for each precursor ion, we count the number of precursors that fall within 10 ppm in their *m/z* value (defined as precursor density). Then the precursors were grouped with the precursor with the highest precursor density. In this way, the MS^2 spectra with similar precursor *m/z* values were grouped into the same cluster. This process was repeated until all precursor ions observed were clustered. In a similar way, the precursors of MS^3 or MS^4 spectra, which were acquired in an ion trap and therefore of lower resolution, were clustered within 0.3 Da from the same cluster of MS^2 or MS^3 spectra, respectively.

The same procedure was used for clustering the precursor ions of in-source fragments, which were formed from the original target precursor. In this case, however, the spectra of the in-source fragment ions were retained only if the peak intensities of the corresponding ions in at least one MS^2 HCD spectrum and the IT-FT spectrum of the original target precursor were at least 20% of the base peak.

Clustering MS^2, MS^3, and MS^4 Spectra Spectra of the precursor ions identified in the above step varied somewhat because of instrument instability. In order to associate all spectra of a single precursor together, a sensitive top-down clustering algorithm was developed to cluster the MS^2, MS^3, and MS^4 spectra and to generate consensus spectra. This procedure is in the class of hierarchical divisive clustering algorithms [28]. Low quality spectra due to random mass shifts or peak intensity fluctuations were eliminated in this step. The basic idea of this algorithm is to first associate all spectra of the same precursor ion in one cluster and then separate them into sub-clusters of similar spectra as follows:

a) In each spectrum, each peak intensity is normalized based on the highest intensity in the spectrum.
b) Individual spectra are gathered to form a composite spectrum, which may include multiple peaks with identical *m/z* and intensity values and where each peak remains linked to its original spectrum.
c) All the peaks in the composite spectrum are sorted based on peak intensity in descending order (retaining the *m/z* value of each peak).
d) Starting with the most intense peak in the composite spectrum, all the peaks within 10 ppm (0.3 Da for low resolution ion-trap spectra) are compared with this most intense peak (if there is more than one most intense peak, the median *m/z* of these peaks is selected). Any spectra in which the peaks in this *m/z* range are much lower than this most intense peak (<20% of intensity) or do not have peaks in this range are eliminated from this composite. This procedure was repeated for the next most intense peak until the intensity of the query peak is lower than 10% of the most intense peak.
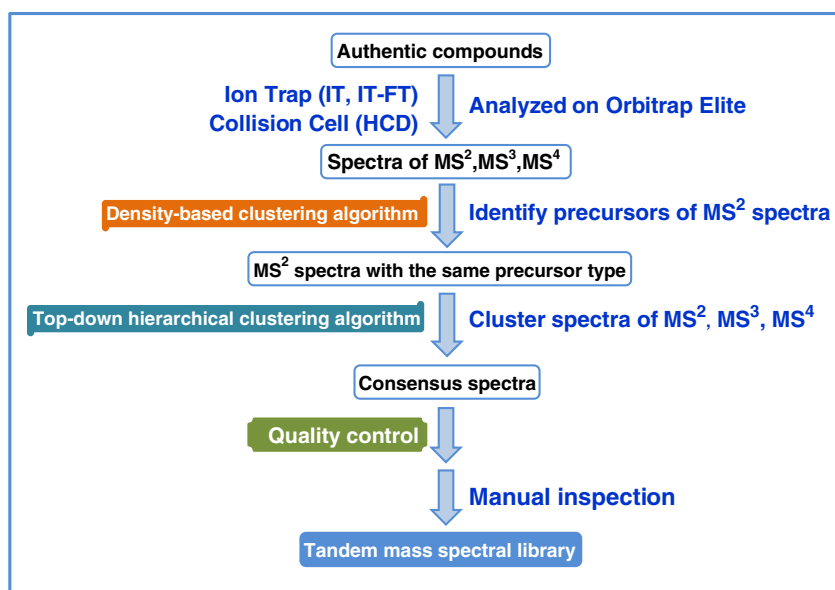
Figure 1. Procedure of building a tandem mass spectral library to include the MS^2 spectra of in-source fragments and MS^n spectra

e) Step (d) is applied to the unclustered spectra. This is repeated until no unclustered spectra remained.

f) For each cluster with at least three spectra, one consensus spectrum is generated. In a similar way as in step (d), starting with the most intense peak within 10 ppm (0.3 Da for low resolution ion-trap spectra), the median *m/z* and intensity is used as a peak for the consensus spectrum. This procedure is repeated for the next most intense peak until all peaks were handled for the consensus spectrum.

The best consensus spectrum is chosen for the library based on the quality control procedures described before [12]. The number of individual spectra is generally 10 to 60 for generating a consensus spectrum for the library.

Identifying Precursor and Product Ions   For high resolution spectra, all potential fragment formulas were listed and their theoretical *m/z* values were calculated based on the formula of each authentic compound that was analyzed [12]. The precursor candidates and the product ions were identified by matching the *m/z* values with the theoretical fragment *m/z* values within 10 ppm. If a precursor *m/z* matched with more than one precursor ion, the most reasonable one was chosen in subsequent manual inspection, based on the compound structure and fragmentations in the MS^2 and MS^n spectra.

Results and Discussion

Clustering and Identifying Precursor Ions

Precursor *m/z* values are clustered by using a density-based clustering algorithm. For example, 282 precursor candidates were found with different *m/z* values and abundances for 17β-

carboxy-17α-formyloxydexamethasone (M = $C_{22}H_{27}FO_6$) (Figure 2a). They were grouped into 16 clusters.

By comparing the *m/z* value of each cluster with the theoretical *m/z* values of the expected precursor ions and all possible in-source fragment ions, 12 precursor ions were identified for subsequent analysis. After quality control procedure and examination of the MS^2 spectra of these ions at the various collision energies, only seven ions were found to have spectra of sufficient quality for inclusion in the library. Sample spectra of four of these ions are shown in Figure 2b and three others are shown in Supplementary Figure S1.

Clustering Mass Spectra

The spectra obtained with the Orbitrap Elite mass spectrometer include high resolution MS^2 spectra from the Orbitrap (HCD at different energies and IT-FT) and low resolution MS^2, MS^3, and MS^4 spectra from the ion-trap (IT). These spectra were analyzed with our newly developed top-down hierarchical divisive clustering algorithm and the resulting consensus spectra were inspected for possible inclusion in the library. When more than one consensus spectrum is obtained, the correct one can be generally chosen based on the relative abundance or number of replicates. But occasionally, a more complex situation is encountered. For example, 31 IT-FT spectra acquired for 16α-hydroxyestrone were clustered into two consensus spectra with 11 and 20 spectra (A and B, respectively, in Figure 3). Spectrum B has a predominant peak with *m/z* 251.1431 (corresponding to loss of $2H_2O$ from the precursor), whereas in spectrum A, this peak is very small. Examination of the HCD spectra at various collision energies shows only one consensus spectrum in each energy, and the *m/z* 251.1431 peak was clearly observed (for example, spectrum C in Figure 3). The ion-trap spectrum (D in Figure 3) also shows a peak at *m/z* 251.1. Therefore, we chose spectrum B for inclusion in the
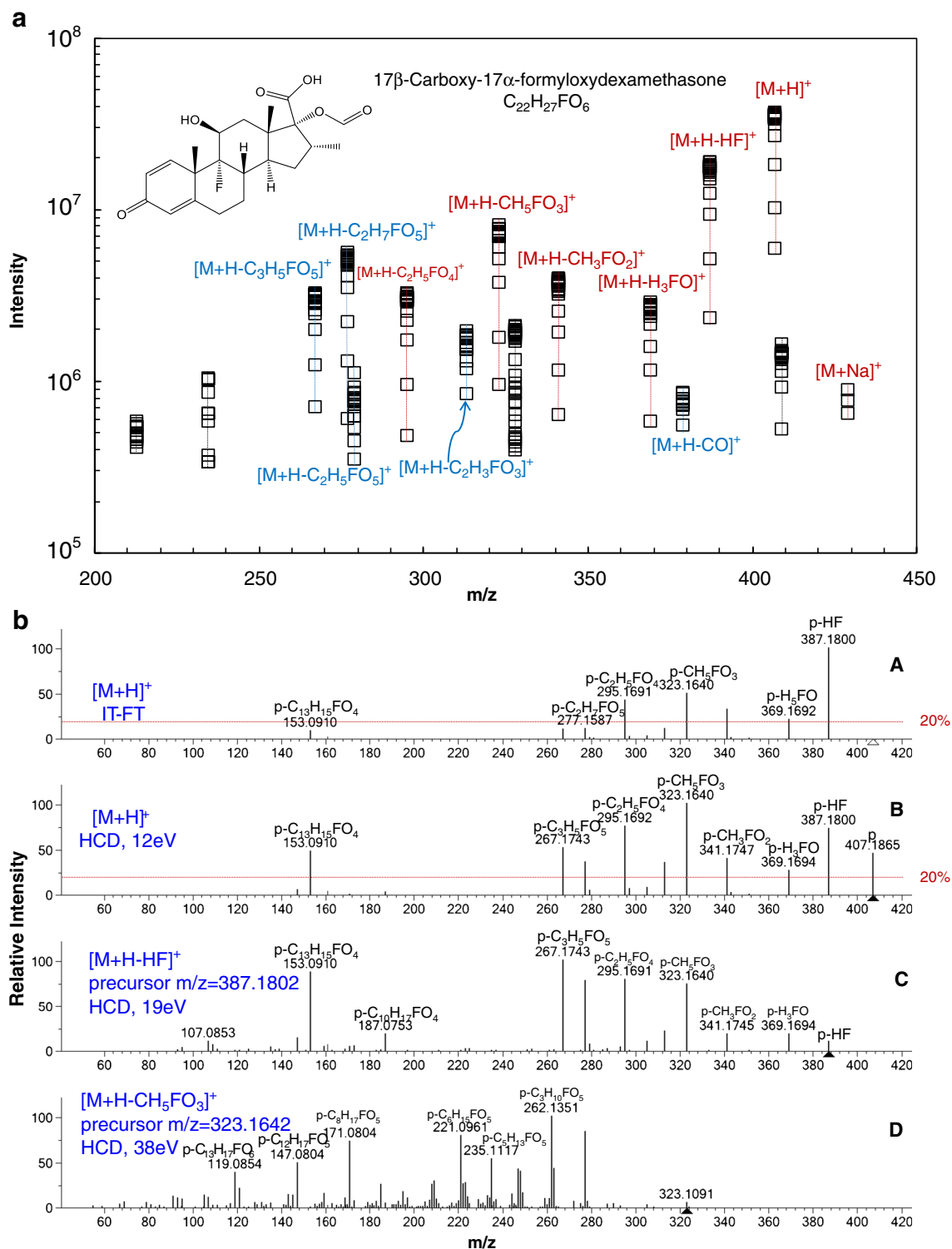
Figure 2. An example of clustering and identifying precursor ions and their tandem mass spectra. (**a**) Clustering and identifying precursor ions from a sample steroid. Each square corresponds to a precursor ion with *m/z* and intensity values. The dotted lines link similar precursors into the same cluster. Fragmentation spectra of the red-colored precursors are retained for the library. The blue-colored precursors are identified but not retained for the library because of low quality of spectra or low abundance. Unlabeled ones are unidentified. (**b**) MS^2 spectra of the primary precursor ion [M + H]$^+$ and the in-source fragment ions from the same compound. The precursors of spectra C and D are in-source fragments, which are also the major product ions (peak intensities >20% of the base peak) in the spectra of [M + H]$^+$ (spectra A and B) acquired from IT-FT and HCD instruments
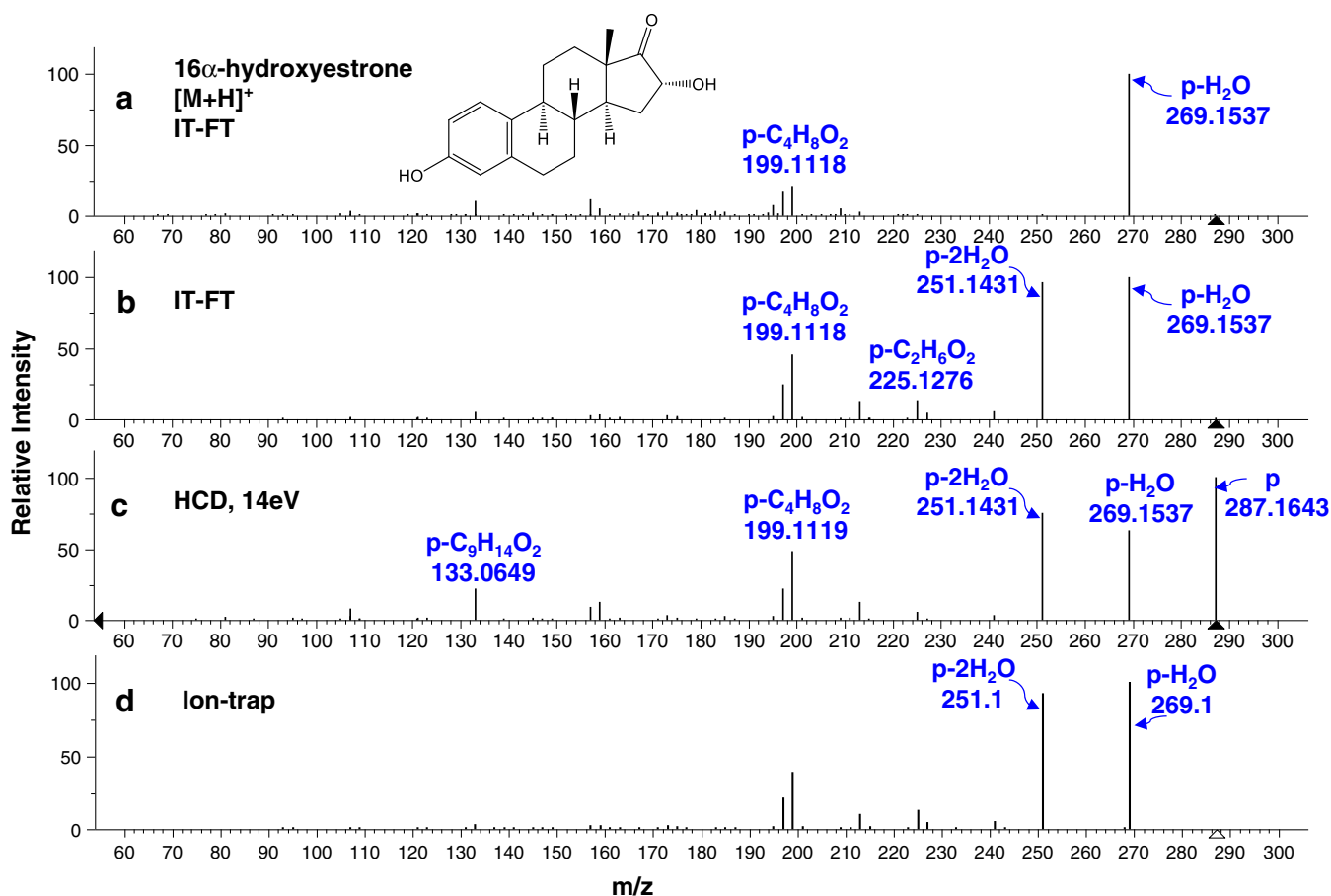
Figure 3. An example of generating consensus spectra after clustering MS^2 spectra by using the top-down hierarchical divisive clustering algorithm. Two consensus spectra were generated (A and B) for IT-FT spectra and one consensus was generated for HCD spectra at each collision energy and ion-trap spectra, respectively

library. A detailed clustering process for this example is described in Supplementary Figure S3.

The top-down hierarchical divisive clustering algorithm was tested using the spectra of 4480 authentic compounds that were analyzed on the Orbitrap instrument. The results show that this algorithm can cluster the spectra efficiently and accurately. In contrast to the bottom-up hierarchical agglomerative clustering algorithm, this top-down clustering algorithm clusters similar spectra directly based on the peak intensities without pair-wise comparison. The purpose of clustering is to group similar spectra into the same cluster. We only need to assure that the spectra in the same cluster are similar. Hierarchical divisive clustering halts if all similar spectra are grouped. It is, therefore, unnecessary to determine the similarity between each pair of spectra.

MS^n Spectra

MS^3 and MS^4 spectra were obtained in order to provide direct evidence of fragmentation pathways that can be utilized for distinguishing similar compounds. As an example, Figure 4 shows the MS^2 spectra of two compounds with the same formula but different structures, α-methylphenylalanine and

N-methylphenylalanine. In both cases, the $[M + H]^+$ ion dissociates to form the $[M + H - CH_2O_2]^+$ ion as the predominant product in the MS^2 spectra. But the MS^3 spectra of these product ions show that the $[M + H - CH_2O_2]^+$ ion produced from α-methylphenylalanine undergoes loss of NH_3 whereas the ion from N-methylphenylalanine undergoes loss of CH_3 and forms additional product ions as well. The loss of NH_3 vs. CH_3 can be used to distinguish these two isomeric compounds. MS^2, MS^3, and MS^4 of these two compounds are provided in the library. The MS^3 and MS^4 spectra were linked to MS^2 and MS^3 spectra, respectively.

In-Source Fragmentation

High resolution spectra of in-source fragment ions help identify compounds in a similar manner as the MS^n spectra. They provide better understanding of the fragmentation pathways of the compound in the mass spectrometer. Examples of MS^2 spectra of in-source fragments from a sample steroid were discussed above (Figure 2b, and Supplementary Figure S1). As another example, the spectra obtained with β-(2-thienyl)alanine are shown in the supplement (Supplementary Figure S2).
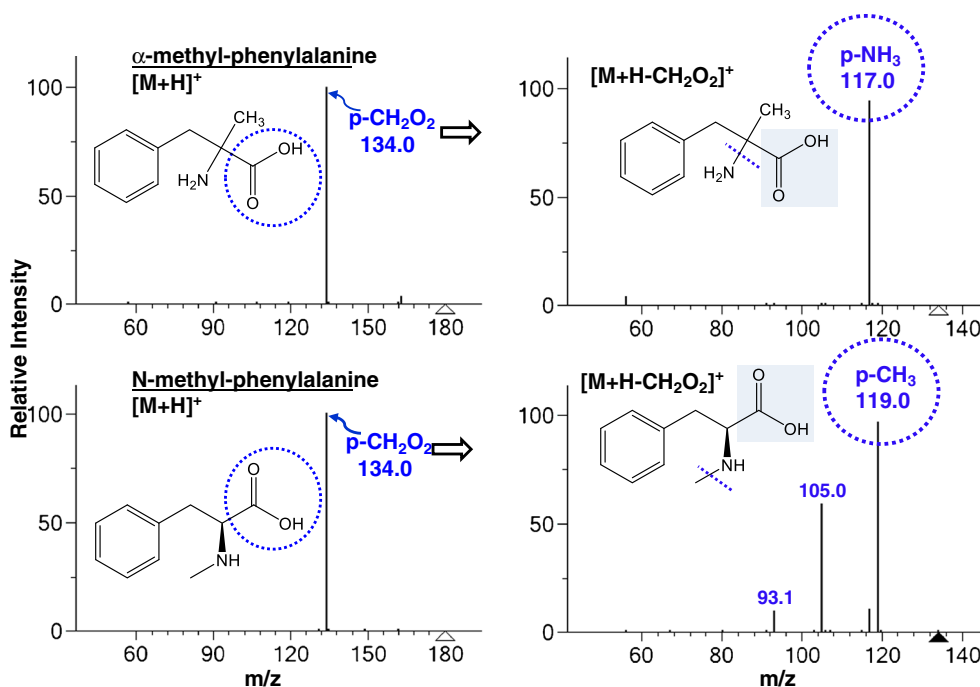
Figure 4. An example of MS^3 ion-trap spectra in the library to distinguish isomers

The MS^2 spectrum of an in-source fragment ion may, in certain cases, be equivalent to a spectrum of another compound, which helps in identifying the in-source fragment. For example, in Figure 5, spectrum A is a consensus spectrum that was generated from 31 individual HCD spectra of an in-source fragment ion of guanosine, $[M + H - C_5H_8O_4]^+$, whereas spectrum B is a consensus spectrum of guanine, which has the same precursor mass as that guanosine fragment. All spectra were acquired at collision energy 48 eV. These two spectra are almost identical and have the same quality. This comparison suggests that the in-source fragment of guanosine, $[M + H - C_5H_8O_4]^+$, is guanine. Therefore, the library contains two equivalent spectra, one of the authentic compound and the other of the in-source fragment ion from the larger compound. When searching an experimental spectrum against library spectra and finding guanine and guanosine at the same elution time, both spectra may have resulted from guanosine. Guanine can then be identified only if it is found at a different elution time or if guanosine is not present. So spectra of in-source fragments in the library can not only help identify compounds but also provide information for tuning the instrument to reduce in-source fragmentation, and identifying those high duplicate "unknown" spectra from in-source fragmentations.
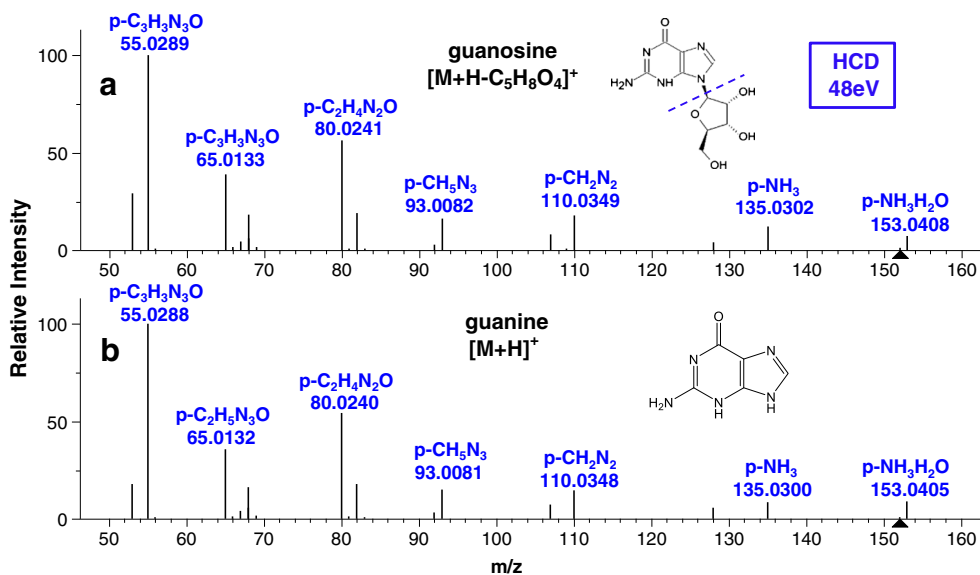

Figure 5. Spectral comparison of in-source fragment from guanosine and authentic compound guanine

An Extended Tandem Mass Spectral Library and Its Application in the Identification of Unknowns

A total of 4480 compounds have been examined with the above methods. The numbers of spectra obtained for the $[M + H]^+$ and $[M - H]^-$ ions were 53,880 and 15,261, respectively, and additional 46,892 spectra were collected for other common precursor ions described before [12]. By using the new processing method described in this paper, 51,960 MS^2 spectra of 1476 new precursor ions for in-source fragments and 40,061 MS^n spectra were identified and confirmed for these compounds. The spectral qualities were ascertained by manual inspection and following the NIST tandem mass spectral library quality control procedures [12]. The library was exported to an sdf-format file, including the annotated peak list, metabolite structure in molfile format, chemical name, synonym(s), formula, CAS number, precursor *m/z*, precursor ion, instrument type, etc. for each consensus spectrum. This comprehensive mass spectral library contains 652,475 high quality reference spectra obtained from 15,243 authentic compounds, including newly added MS^2 spectra of in-source fragments and MS^n spectra. The compounds include metabolites (of these, ~3500 are human metabolites), lipids, sugars, glycans, drugs, pesticides, surfactants, and other small molecules. In addition, the library also contains 84,446 MS^2 spectra of all dipeptides, all tryptic tripeptides, and many bioactive peptides.

The extended library was tested by searching the data sets from *E. coli* extracts that were run on the Q-Exactive mass spectrometer using the NIST MS Search (ver. 2.2) program. For example, 527 experimental MS^2 spectra were found to match the spectrum of oxidized glutathione (Figure 6a). In addition, 215 MS^2 spectra obtained at the same elution time matched the spectrum of the in-source fragment $[M + H - C_5H_7NO_3]^+$ from this compound (Figure 6b). This in-source fragment identification confirms the identification of oxidized glutathione. However, this result also showed that the substructure was generated in the mass spectrometer, and was not present in the original sample. Searching the library, including the spectra from in-source fragments, helps users to confirm compound identification and avoid misidentifying in-source fragments as components.

Conclusion

We describe a top-down hierarchical divisive clustering algorithm that can associate the MS^2, MS^3, and MS^4 spectra of
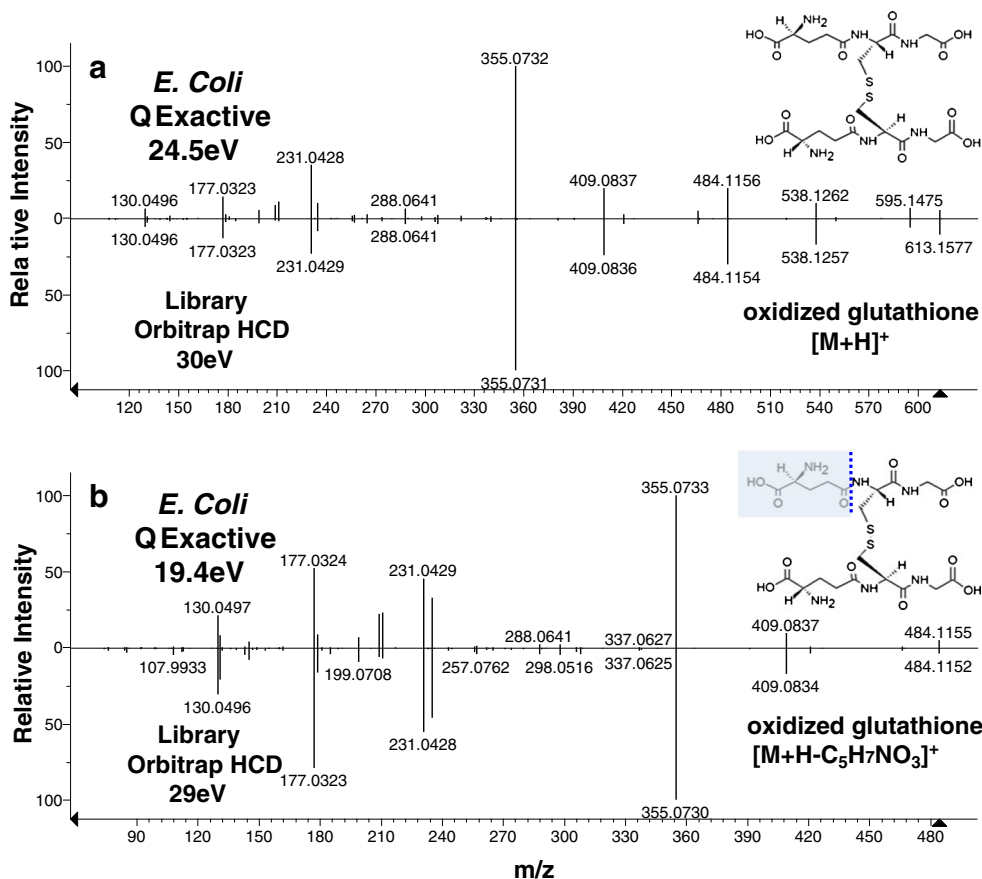

Figure 6. An example of searching the NIST Tandem Mass Spectral Library for identifying *E. coli* metabolites. (**a**) Identifying a metabolite, oxidized glutathione. (**b**) Identifying its in-source fragment. In both cases, the top spectrum is an experimental spectrum and the bottom one is a reference spectrum in the library. The scores for these matches were 951 and 874, respectively (max = 999)

original precursor ions and of their in-source fragment ions efficiently and accurately. A density-based clustering algorithm was used for clustering precursor ions, which were subsequently identified by matching with the calculated theoretical m/z values based on the formula of each analyzed authentic compound. We applied these algorithms to extend the tandem mass spectral library to include the spectra of in-source fragments and MS^n spectra. The extended library can be used for more accurate compound identification by searching of spectra of in-source fragment ions and MS^n spectra.

Acknowledgements

Compliance with Ethical Standards

Disclaimer   Certain commercial equipment or materials are identified in this paper to specify the experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose. The authors declare no competing financial interest.

References

1. Apfel, A., Chakel, J.A., Fischer, S., Lichtenwalter, K., Hancock, W.S.: Analysis of oligonucleotides by HPLC-electrospray ionization mass spectrometry. Anal. Chem. **69**, 1320–1325 (1997)
2. Tolstikov, V.V., Fiehn, O., Tanaka, N.: Application of liquid chromatography-mass spectrometry analysis in metabolomics: reversed-phase monolithic capillary chromatography and hydrophilic chromatography coupled to electrospray ionization-mass spectrometry. Methods Mol. Biol. **358**, 141–155 (2007)
3. Han, X., Aslanian, A., Yates, J.R.: Mass spectrometry for proteomics. Curr. Opin. Chem. Biol. **12**, 483–490 (2008)
4. Han, X., Gross, R.W.: Shotgun lipidomics: electrospray ionization mass spectrometric analysis and quantitation of cellular lipidomes directly from crude extracts of biological samples. Mass Spectrom. Rev. **24**, 367–412 (2005)
5. Steiner, I., Brauers, G., Temme, O., Daldrup, T.: A sensitive method for the determination of hordenine in human serum by ESI(+) UPLC-MS/MS for forensic toxicological applications. Anal. Bioanal. Chem. **408**, 2285–2292 (2016)
6. Wijetunge, C.D., Li, Z., Saeed, I., Bowne, J., Hsu, A.L., Roessner, U., Bacic, A., Halgamuge, S.K.: Exploratory analysis of high-throughput metabolomic data. Metabolomics **9**, 1311–1320 (2013)
7. Tsou, C.-C., Avtonomov, D., Larsen, B., Tucholska, M., Choi, H., Gingras, A.-C., Nesvizhskii, A.I.: DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. Nat. Methods **12**, 258–264 (2015)
8. Schubert, O.T., Gillet, L.C., Collins, B.C., Navarro, P., Rosenberger, G., Wolski, W.E., Lam, H., Amodei, D., Mallick, P., MacLean, B., Aebersold, R.: Building high-quality assay libraries for targeted analysis of SWATH MS data. Nat. Protoc. **10**, 426–441 (2015)
9. Gao, C., Weisman, D., Gou, N., Ilyin, V., Gu, A.Z.: Analyzing high dimensional toxicogenomic data using consensus clustering. Environ. Sci. Technol. **46**, 8413–8421 (2012)
10. Bartel, J., Krumsiek, J., Theis, F.J.: Statistical methods for the analysis of high-throughput metabolomics data. Comput. Struct. Biotechnol. J. **4**, e201301009 (2013)
11. Sugimoto, M., Kawakami, M., Robert, M., Soga, T., Tomita, M.: Bioinformatics tools for mass spectroscopy-based metabolomic data processing and analysis. Curr. Bioinform. **7**, 96–108 (2012)
12. Yang, X., Neta, P., Stein, S.E.: Quality control for building libraries from electrospray ionization tandem mass spectra. Anal. Chem. **86**, 6393–6400 (2014)
13. Yang, X., Neta, P., Stein, S.E.: Constructing MS^n mass spectral library for more accurate metabolite identification. Oral presentation (MOB 22), Proceedings of the 63rd ASMS Conference on Mass Spectrometry and Allied Topics. St. Louis. MO, May 31–June 4 (2015)
14. NIST Tandem Mass Spectral Library 2014. Available at: http://chemdata.nist.gov/mass-spc/msms-search/
15. Toghi Eshghi, S., Shah, P., Yang, W., Li, X., Zhang, H.: GPQuest: a spectral library matching algorithm for site-specific assignment of tandem mass spectra to intact *N*-glycopeptides. Anal. Chem. **87**, 5181–5188 (2015)
16. Lei, Z., Jing, L., Qiu, F., Zhang, H., Huhman, D., Zhou, Z., Sumner, L.W.: Construction of an ultrahigh pressure liquid chromatography-tandem mass spectral library of plant natural products and comparative spectral analyses. Anal. Chem. **87**, 7373–7381 (2015)
17. Shao, W., Lam, H.: Denoising peptide tandem mass spectra for spectral libraries: a bayesian approach. J. Proteome Res. **12**, 3223–3232 (2013)
18. Kind, T., Okazaki, Y., Saito, K., Fiehn, O.: LipidBlast templates as flexible tools for creating new in-silico tandem mass spectral libraries. Anal. Chem. **86**, 11024–11027 (2014)
19. Ridder, L., van der Hooft, J.J., Verhoeven, S., de Vos, R.C., Bino, R.J., Vervoort, J.: Automatic chemical structure annotation of an LC-MS(n) based metabolic profile from green tea. Anal. Chem. **85**, 6033–6040 (2013)
20. Rojas-Cherto, M., Peironcely, J.E., Kasper, P.T., van der Hooft, J.J., de Vos, R.C., Vreeken, R., Hankemeier, T., Reijmers, T.: Metabolite identification using automated comparison of high-resolution multistage mass spectral trees. Anal. Chem. **84**, 5524–5534 (2012)
21. Zhang, H., Singh, S., Reinhold, V.N.: Congruent strategies for carbohydrate sequencing. 2. FragLib: an MSn Spectral Library. Anal. Chem. **77**, 6263–6270 (2005)
22. Ashline, D.J., Lapadula, A.J., Liu, Y.-H., Lin, M., Grace, M., Pramanik, B., Reinhold, V.N.: Carbohydrate structural isomers analyzed by sequential mass spectrometry. Anal. Chem. **79**, 3830–3842 (2007)
23. Lapadula, A.J., Hatcher, P.J., Hanneman, A.J., Ashline, D.J., Zhang, H., Reinhold, V.N.: Congruent strategies for carbohydrate sequencing. 3. OSCAR: an algorithm for assigning oligosaccharide topology from MSn Data. Anal. Chem. **77**, 6271–6279 (2005)
24. Kameyama, A., Kikuchi, N., Nakaya, S., Ito, H., Sato, T., Shikanai, T., Takahashi, Y., Takahashi, K., Narimatsu, H.: A Strategy for identification of oligosaccharide structures using observational multistage mass spectral library. Anal. Chem. **77**, 4719–4725 (2005)
25. Lynn, K.S., Cheng, M.L., Chen, Y.R., Hsu, C., Chen, A., Lih, T.M., Chang, H.Y., Huang, C.J., Shiao, M.S., Pan, W.H., Sung, T.Y., Hsu, W.L.: Metabolite identification for mass spectrometry-based metabolomics using multiple types of correlated ion information. Anal. Chem. **87**, 2143–2151 (2015)
26. Kim, J.S., Monroe, M.E., Camp II, D.G., Smith, R.D., Qian, W.J.: In-source fragmentation and the sources of partially tryptic peptides in shotgun proteomics. J. Proteome Res. **12**, 910–916 (2013)
27. Xu, Y.F., Lu, W., Rabinowitz, J.D.: Avoiding misannotation of in-source fragmentation products as cellular metabolites in liquid chromatography-mass spectrometry-based metabolomics. Anal. Chem. **87**, 2273–2281 (2015)
28. Han, J., Kamber, M.: Data mining: concepts and techniques. Morgan Kaufmann Publishers, San Francisco (2001)