Towards Comprehensive, Reliable and Accurate Mass Data Repositories

Eva Duchoslav[1]; Lyle Burton[1]; Emmanuel Varesio[2]; Ron Bonner[1]; Gérard Hopfgartner[2]

[1]AB SCIEX, 71 Four Valley Drive, Concord, ON, L4K 4V8 Canada; [2]Life Sciences Mass Spectrometry, School of Pharmaceutical Sciences, University of Geneva, University of Lausanne, Geneva, Switzerland

ABSTRACT

A research tool for automated assessment, enhancement and annotation of spectral repositories has been created. This tool was used to annotate several spectral repositories collected with different mass spectrometry platforms. The research tool was instrumental in annotating MS/MS fragments with elemental compositions and likely substructure portions of parent molecule as well as tracking fragments that could not be assigned.
By studying chemical space and quality of annotation of spectral repositories as a whole, we could improve the accuracy and quality of the spectral repositories and better understand the relationship between the chemical structures and fragments in the repositories.

INTRODUCTION

High resolution tandem accurate mass spectrometry has been adopted for use in qualitative applications in many different areas. Spectral repositories are often used in unattended routine data processing workflows; for these high throughput solutions, it is critical to employ reference spectral sources of the highest possible quality[1]. Here we present a research tool for automated assessment, enhancement and annotation of spectral repositories. We also present strategies for their potential use in small molecules qualitative applications for both screening and for the identification of unknowns.

MATERIALS AND METHODS

Sample Preparation:

Several high resolution MS/MS data sets in MS Access format were used in this study:

1/ internal spectral libraries of pesticides, antibiotics, and forensics collected on an accurate mass system,
2/ an internal spectral library of drugs, metabolites and environmental contaminants containing 378 compounds collected on a nominal mass system,
3/ a subset of the compounds of the Human Metabolome Library (http://www.hmdb.ca/hml) collected on the TripleTOF® 5600+ system.

Fragment and neutral loss annotations (with formula, mass error and potential substructure) were added to the databases using a research prototype tool. SQL queries as well as R scripts were used to inspect the quality of the library and to investigate common features in the spectral datasets as well as their relationships to the parent molecules to yield the information needed to annotate unknown spectra.

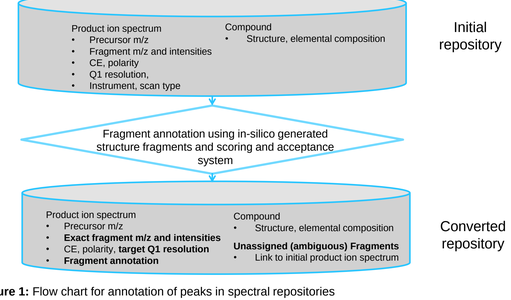


Figure 1: Flow chart for annotation of peaks in spectral repositories

RESULTS AND DISCUSSION

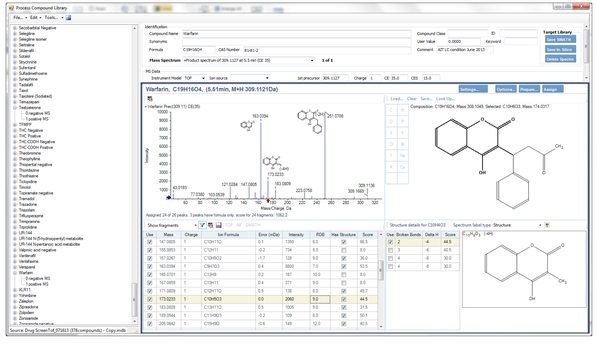


Figure 2: User interface of a prototype Populate Compound Library application. This tool was used for conversion of the spectral libraries into higher accuracy ones where fragments in product ion spectra have their exact masses.

Fragment assignment is done by finding an elemental composition for each measured m/z, assuming that the element counts range from 0 to the number in the precursor ion of the studied compound. The elemental composition assignment is scored based on m/z error and change in DBE. To assign potential substructure pieces to the fragment, the compound structure undergoes in-silico fragmentation and the elemental compositions of substructure parts are aligned with elemental compositions of the spectrum fragments. In order to determine the most likely substructure piece, the structure fragment annotation scores are based on the number of broken bonds, the type of broken bonds (in relation to the CE used), the type of internal bonds, evidence of cascading fragmentation, hydrogen migration, rearrangement and evidence / agreement with fragments in product ion spectra from compounds of similar structures.

Library Name	Accurate mass	Number of compounds	Number of spectra	Annotated fragments	Ambiguous fragments	Ambiguous % TIC
pesticides	yes	544	1216	22416	831 (3.6%)	2.5%
antibiotics	yes	244	259	6701	199 (2.9%)	1.7%
forensics	yes	371	517	8841	344 (3.7%)	1.7%
Drug screen	no	1254	3478	82617	NA	
HMDB positive	yes	472	472	23766	2443	
HMDB positive subset*	yes	398	398	17939	1010(5.3%)	3.6%
HMDB negative	yes	465	465	6914	320 (4.6%)	3.1%

Table 1: Annotation of fragments in spectral repositories. Spectral peaks for annotation were selected down to 1% of the base peak. Up to 4 single or aromatic bonds were broken within in-silico fragmentation. The mass tolerance for elemental composition assignment was set to 5mDa for accurate mass data and 100mDa for nominal mass data. Fragments that could not be assigned any elemental composition were considered to be ambiguous.
* Compounds with MH^+ precursor intensity of less than 1000cps were removed from the dataset.

The main cause of unassigned fragments was interfering material in the Q1 isolation window. This effect was multiplied in cases where ionization efficiency of the putative precursor $M+H^+$ or $M\text{-}H^-$ was not strong. Overall, the median relative intensity of unassigned ions was 1.7% of the base peak.

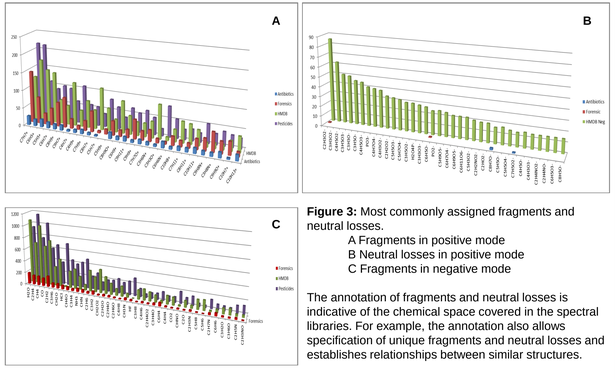
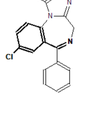


Figure 3: Most commonly assigned fragments and neutral losses.
A Fragments in positive mode
B Neutral losses in positive mode
C Fragments in negative mode

The annotation of fragments and neutral losses is indicative of the chemical space covered in the spectral libraries. For example, the annotation also allows specification of unique fragments and neutral losses and establishes relationships between similar structures.

Product Ion Data Type (m/z tolerance)			**A:** Accurate Mass TOF MS/MS (2mDa)				**B:** Accurate Mass TOF MS/MS (100mDa)			**C:** Nominal Mass EPI (100mDa)		
Mass	Charge	Ion Formula	Error (mDa)	Intensity	Has Structure	Score	Candidate Compositions	Compositions with Structures	Rank of Proposed Result	Error (mDa)	Score	Rank
138.0098	1	$C7H5NCl+$	-0.7	1669	TRUE	40.5	12	7	1	72.9	40.5	1
165.0204	1	$C8H6N2Cl+$	-1	2380	TRUE	34.5	12	6	1	62	34.5	1
205.0753	1	$C14H9N2+$	-0.7	2557	TRUE	30.5	11	6	2	81	30.5	1
224.0255	1	$C14H7NCl+$	-0.7	1370	TRUE	43	9	6	1	97.7	38	1
241.0515	1	$C14H10N2Cl+$	-1.2	2724	TRUE	55.6	8	4	1	56.2	50.6	1
251.0367	1	$C15H8N2Cl+$	-0.4	1583	TRUE	55	5	3	1	35.2	49	1
255.0679	1	$C15H12N2Cl+$	-0.5	1515	TRUE	53.6	6	5	1	58.2	61.6	1
274.1199	1	$C17H14N4+$	-1.4	4359	TRUE	48.5	2	1	1	53.5	48.5	1
281.0698	1	$C16H12N3Cl+$	-1.6	10000	TRUE	30.5	3	2	1	12.6	38.6	1
309.0888	1	$C17H14N4Cl+$	-1.4	4229	TRUE	55	1	1	1	-10.5	55	1

Table 2: Annotation of TOF MS/MS of Alprazolam (C17H13N4Cl) with A: mass tolerance of 2mDa, which is compatible with accurate mass data. B: The same MSMS with a mass tolerance of 100mDa where the selection of the best fragment is more affected by the fragment substructure and corresponding enhanced product ion spectrum. Nine out of ten fragments ranked as number 1 with the 100mDa tolerance and one entry ranked as number 2. C: All fragments were correctly assigned to nominal mass data.



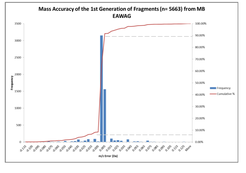


Figure 5. Annotation of curated accurate mass data of MassBank EAWAG library[2] with target mass accuracy of 100mDa.

We used a curated accurate mass dataset for proof of concept of the automated conversion of nominal mass data. Assuming fragment mass accuracy in the curated dataset, the unattended accurate mass conversion was correct for more than 82% of fragments. No provisions were made for fragments that are not the direct product of the parent compound or fragments which scored close to the best result.

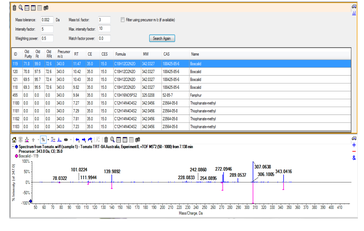


Figure 6. Tomato extract screening using an exact mass pesticide spectral library. Converted spectral libraries can be used directly from the Analyst® or PeakView® Software[3].

A peak at 7.1 minutes was confirmed to be boscalid. The narrow mass tolerance eliminated all less likely candidates.

CONCLUSIONS

The prototype Populate Spectral Library tool was used to confirm the quality of the accurate mass spectral libraries collected on the Triple TOF® 5600+ platform, to identify and filter out signal that is not attributed to parent molecules and to convert the spectral libraries into exact mass ones. The in-silico converted libraries allowed for more generic parameter library search settings and more confident search results.

We have developed a scoring system that takes into account elemental composition as well as potential fragment structure. By expanding the fragment scoring system we were able to annotate nominal mass product ion spectra with an accuracy better than 80%.

REFERENCES

1. G. Hopfgartner, E. Varesio, Abhinandan K.R., E. Duchoslav, L. Burton, R. Bonner: Towards a Workflow for the Automated Structural Elucidation of Unknowns from MS/MS Spectra, 61th ASMS conference, 2013
2. MassBank: A public repository for sharing mass spectral data for life sciences", H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa. Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Yokota Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito and T. Nishioka, J. Mass Spectrom., **45**, 703-714 (2010)
3. G. Hopfgartner, E. Varesio, L. Burton, E. Duchoslav, R. Bonner: SWATH Libraries and Common Fragment Libraries for Metabolites Identification in Urine, 62th ASMS conference, 2014, MP 629

TRADEMARKS/LICENSING