What is mzXML good for?

Simon M Lin[†]*, Lihua Zhu, Andrew Q Winter, Maciek Sasinowski and Warren A Kibbe*

mzXML (extensible markup language) is one of the pioneering data formats for mass spectrometry-based proteomics data collection. It is an open data format that has benefited and evolved as a result of the input of many groups, and it continues to evolve. Due to its dynamic history, its structure, purpose and applicability have all changed with time, meaning that groups that have looked at the standard at different points during its evolution have differing impressions of the usefulness of mzXML. In discussing mzXML, it is important to understand what mzXML is not. First, mzXML does not capture the raw data. Second, mzXML is not sufficient for regulatory submission. Third, mzXML is not optimized for computation and, finally, mzXML does not capture the experiment design. In general, it is the authors' opinion that XML is not a panacea for bioinformatics or a substitute for good data representation, and groups that want to use mzXML (or other XML-based representations) directly for data storage or computation will encounter performance and scalability problems. With these limitations in mind, the authors conclude that mzXML is, nonetheless, an indispensable data exchange format for proteomics.

Expert Rev. Proteomics 5(6), 839–845 (2005)

CONTENTS

[†]*Author for correspondence*
Northwestern University,
Robert H Lurie Cancer Center,
Chicago, IL 60611, USA
Tel.: +1 312 695 1331
Fax: +1 312 695 1347
s-lin2@northwestern.edu

KEYWORDS:
experiment design, mass spectrometry, mzXML, proteomics, raw data, regulatory submission

What are mzXML & mzData?

mzXML (extensible markup language) is a coordinated effort by a group of institutions to introduce a common data format to represent mass spectrometry (MS)-based proteomics data [1]. The goal is to replace the many different proprietary formats that impede the exchange, analysis and comparison of measurement results with an interoperable and extensible standard.

This vision is shared by many different groups. The analytic chemistry community has explored this issue for a long time [2–3], but the focus has not only been on proteomics. Another effort by the Proteomics Standard Initiative (PSI) group of the Human Proteome Organisation (HUPO) resulted in XML representations of both protein–protein interaction data and MS data [4]. PSI proposed mzData to store peak lists, and mzIdent to store the mapping from peak lists to proteins. mzData and mzXML share many design principles. In fact, there are converters between mzXML and mzData. This review will highlight areas where mzData and mzXML differ, but use mzXML as the basis of the review.

An example mzXML file is illustrated in FIGURE 1. It utilizes the syntax of XML to encode data (between tags) and metadata (within tags). For proteomics, the data are mass-to-charge ratio (m/z) and intensity pairs, and all the rest of the measurement conditions are metadata.

mzXML is frequently discussed together with Minimum Information About a Proteomics Experiment (MIAPE). MIAPE is an emerging content description for MS or, more generally, proteomics experiments. It is a guideline on how to fully report a proteomics experiment. mzXML and mzData are nascent content representations for MS. Thus, the relationship between mzXML and MIAPE is similar to the relationship between microarray and gene expression (MAGE; XML representation) and Minimum Information About A Microarray Experiment (MIAME; content guideline) in microarrays.

What are the raw data of mass spectrometry-based proteomics?

Conceptually, a mass spectrum is a plot of intensity against m/z values (FIGURE 2). The m/z intensity pair might be directly observed as raw

10.1586/14789450.5.6.839

ISSN 1478-9450

data, or indirectly inferred from time measurements or time–frequency transformations, depending on the type of mass spectrometer.

For example, the time-of-flight (TOF) analyzer, used by Ciphergen Biosystems, Inc.'s surface-enhanced laser desorption/ionization (SELDI)-TOF machine, converts time into m/z data via a calibration process. Thus, Ciphergen's raw data only specify the intensities sampled at a regular time interval, and the time coordinate is converted to m/z data based on calibration data. The raw measurements are critical to evaluate the results, which can be transformed using the calibration data to m/z values [5].

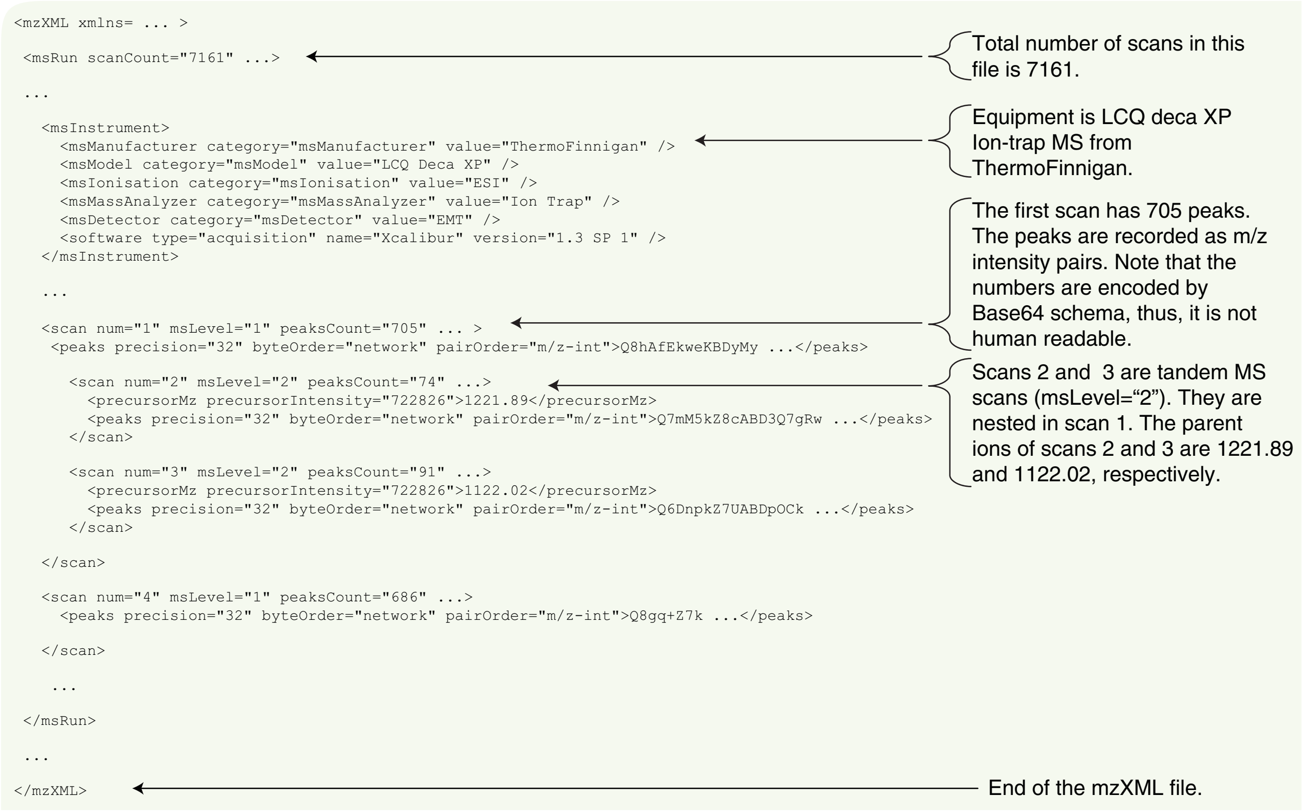
The conversion of raw data to m/z data by Fourier transform ion cyclotron resonance (FTICR)-MS is more complex. FTICR takes advantage of ICR to generate a resonance spectrum. Briefly, ions moving at their cyclotron frequency can absorb radiofrequency energy of that frequency. A pulse of radiofrequency excites the ions in the magnetic field. The ions then re-emit the radiation with a characteristic decay, and this free-induction decay (FID) signal can be Fourier transformed to produce the emitted frequencies. In turn, these frequencies are converted to give the masses of the ions present in the cyclotron. The FID signal in the time domain is transformed into frequency domain and, subsequently, into the mass domain via a calibration process.

New algorithms can further optimize the conversion process from raw data to m/z intensity pairs. With the available FTICR-MS raw data, Savitski and colleagues recently demonstrated that improved fast Fourier transform algorithms can increase the mass accuracy fourfold [6]. Increased mass accuracy can directly reduce the ambiguity of identifying peptides [7].

Raw data are also critical in identifying and evaluating the source of discrepancies between different data processing strategies. For example, without the raw data, it is impossible to prove that the difference in XPRESS quantification is a consequence of different algorithms used for charge state determination [1].

mzXML does not capture the raw data

As stated in the introduction, mzXML only captures processed data already transformed into m/z intensity pairs. This is also true for mzData, and pragmatically, this makes sense due to the dependence of the raw data on the technique used, the potential size of the raw data files and the fact that most current proteomics applications can only handle m/z intensity pairs as inputs. However, this limitation could hinder the development of novel statistical algorithms, because it will render the raw data unavailable in a machine-independent format.



```
<mzXML xmlns= ... >

 <msRun scanCount="7161" ...>

 ...

   <msInstrument>
     <msManufacturer category="msManufacturer" value="ThermoFinnigan" />
     <msModel category="msModel" value="LCQ Deca XP" />
     <msIonisation category="msIonisation" value="ESI" />
     <msMassAnalyzer category="msMassAnalyzer" value="Ion Trap" />
     <msDetector category="msDetector" value="EMT" />
     <software type="acquisition" name="Xcalibur" version="1.3 SP 1" />
   </msInstrument>

   ...

   <scan num="1" msLevel="1" peaksCount="705" ... >
    <peaks precision="32" byteOrder="network" pairOrder="m/z-int">Q8hAfEkweKBDyMy ...</peaks>

      <scan num="2" msLevel="2" peaksCount="74" ...>
        <precursorMz precursorIntensity="722826">1221.89</precursorMz>
        <peaks precision="32" byteOrder="network" pairOrder="m/z-int">Q7mM5kZ8cABD3Q7gRw ...</peaks>
      </scan>

      <scan num="3" msLevel="2" peaksCount="91" ...>
        <precursorMz precursorIntensity="722826">1122.02</precursorMz>
        <peaks precision="32" byteOrder="network" pairOrder="m/z-int">Q6DnpkZ7UABDpOCk ...</peaks>
      </scan>

   </scan>

   <scan num="4" msLevel="1" peaksCount="686" ...>
     <peaks precision="32" byteOrder="network" pairOrder="m/z-int">Q8gq+Z7k ...</peaks>

   </scan>

    ...

 </msRun>

 ...

</mzXML>
```

Total number of scans in this file is 7161.

Equipment is LCQ deca XP Ion-trap MS from ThermoFinnigan.

The first scan has 705 peaks. The peaks are recorded as m/z intensity pairs. Note that the numbers are encoded by Base64 schema, thus, it is not human readable.

Scans 2 and 3 are tandem MS scans (msLevel="2"). They are nested in scan 1. The parent ions of scans 2 and 3 are 1221.89 and 1122.02, respectively.

End of the mzXML file.

Figure 1. Example mzXML file. mzXML utilizes XML tagging syntax. All data are wrapped with metadata tags for explicit interpretation. The size of a single mzXML file can be in the range of mega- to gigabytes.
MS: Mass spectrometry; m/z: Mass-to-charge ratio.

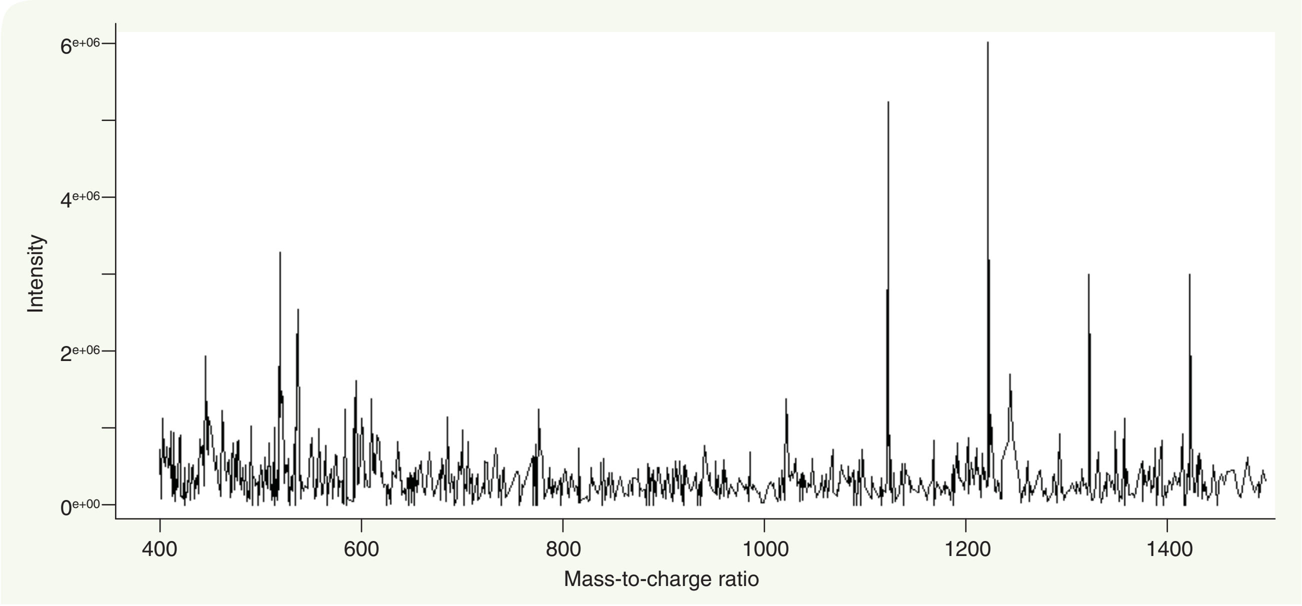


Figure 2. Example mass spectrum.

Fortunately, this problem can be readily addressed from a technical standpoint due to the open design of mzXML. Both mzXML and mzData adopted a strategy of separating the more stable structural relationships among data elements (data schema model) from the more evolving component details (domain-specific data ontology). Thus, the peaks object in the scan object can be expanded to describe raw data including the metadata necessary to track the instrument settings and information required to interpret the raw data. This change will not affect the structural integrity of the mzXML design.

Looking at the microarray community, the advocacy by the Microarray Gene Expression Data (MGED) Society to make microarray raw data available has resulted in a rapid proliferation and improvements in algorithms and software to handle raw data once obscured by built-in utilities bundled with instruments [8]. Similarly, by providing direct access to raw data from MS instruments to the bioinformatics community, the authors would expect the rapid emergence of new solutions to MS-based proteomics noise reduction, statistically valid dimension reduction, peak determination, charge state determination, and quantitative determination of differential labeling.

mzXML alone is not sufficient for regulatory submission

For applications of MS that will be submitted to the US Food and Drug Administration (FDA), Title 21 of the Code of Federal Regulations (CFR) Part 11 comes into effect and: dictates how electronic data must be kept; dictates how documentation and auditing of all data transformations must be held; and describes the chain of custody required to demonstrate the integrity of the data provided in support of an investigational new drug application (IND) or other FDA submission. (See the CFR for the actual text of the regulation.) For instance, in submitting clinical trial data, the raw data, the results of the statistical analysis and code used to perform the analysis are all submitted, along with a log of the chain of custody for the data.

As illustrated by the example in TABLE 1, critical information in an instrument-specific format can be lost when converting to mzXML. It should be noted that both formats fail to provide uncertainties of the m/z and intensity measurements, which are critical to interpret the results.

The authors of the original mzXML paper noted that mzXML should not be considered as a replacement for the native raw data, and that the raw data may need to be submitted for compliance with the US FDA regulations, including Title 21 CFR Part 11 [1]. Although current regulations indicate that the original data sets need to be maintained and submitted, it seems unrealistic that FDA examiners will be able to inspect the vast amounts of raw data in many native formats.

mzXML is not an appropriate data structure for computation

XML is designed to facilitate data exchange. All data are hierarchically surrounded by metadata tags for clarity during the transmission. However, this structure is inefficient for large-scale computations. After data exchange, a native data structure should be employed that is suited to the computation that needs to be performed.

For the purpose of signal processing, a vector representation is appropriate for a single spectrum, and a matrix representation for a collection of spectra. With these internal data structures, raw data can be efficiently manipulated. For example, a total ion current chromatogram from liquid chromatography (LC) MS can be easily constructed by summing the counts in each row of the matrix (FIGURE 3).

Thus, the authors can recommend mzXML for data exchange, but not for computation. Local computation should be done using data structures optimized for this purpose.

mzXML does not capture the overall experiment design

An MS data set is not a collection of spectra from disconnected MS scans. The spectra are related to each other according to investigational goals as specified by an experiment design. Capturing experiment design information is a critical step for efficient communication between biologists and statisticians. A miscommunication can result in the misinterpretation of large-scale data collection [9].

Based on the constraints and/or requirements of the instrumentation, which typically include a LC step to partially resolve the components in the sample, a set of MS spectra are often related by chromatography time (LC/MS). The mzXML community is considering a separation element to describe the LC details and allow the coupling of a set of spectra to a single original sample.

However, the biologic aspects of the experimental design are not captured by mzXML. For example, certain steps in the MS data collection can be designed to have biologic replicates, intra- or intermachine replicates [10]. Arguably, it might be inappropriate for mzXML to capture the experimental design. The attributes surrounding the experimental design are shared with other functional genomics experiments, including microarrays. To this end, the functional genomic experiment object model, which covers common design elements in both microarrays and proteomics, is an excellent candidate for standardization [11].

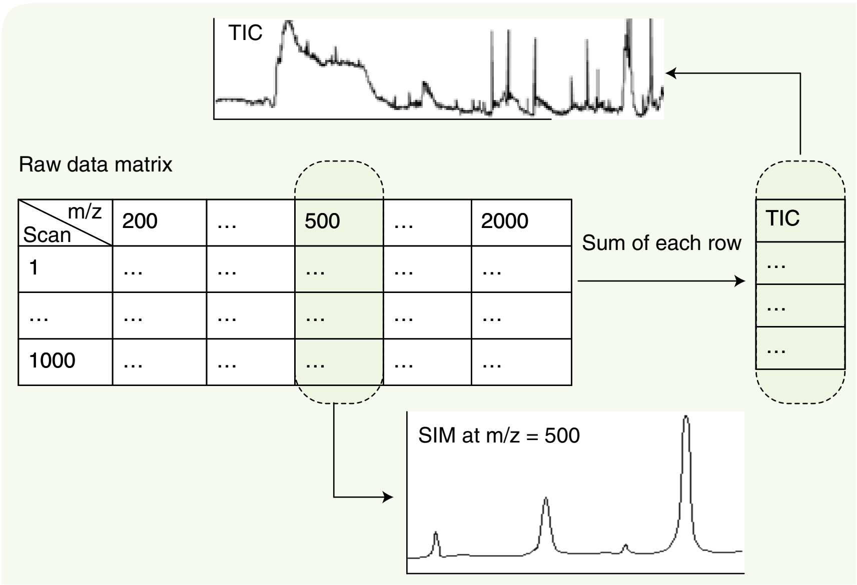
Table 1. Instrument-specific format, such as Ciphergen SeldiXML, can be mapped to mzXML.

SeldiXML (Ciphergen)	mzXML	SeldiXML (Ciphergen)	mzXML
<spectrum>	**<msRun>**	**<bioprocName>**	**<robotModel>**
<spectrumName>	**<scanCount>**	<sampleName>	XXXXXX
XXXXXX	<startTime>	<samplePatient/>	XXXXXX
XXXXXX	<endTime>	<sampleGroup/>	XXXXXX
...	...	...	...
XXXXXX (acquisition)	<software type>	XXXXXX	<dataProcessing>
XXXXXX (ProteinChipReader)	<name>	XXXXXX	<centroided>
<versionSoftware>	**<version>**	XXXXXX	<deIsotoped>
<userName>	**<operator>**	XXXXXX	<chargeDeconvoluted>
<samplePreparation>	**<separation>**	XXXXXX	<spotIntegration>
<bindAndWashWashCond>	XXXXXX	XXXXXX	<intensityCutoff>
<fractionName/>	XXXXXX	XXXXXX	<processingOperation>
<eamCompound/>	**<maldiMatrix>**	XXXXXX	<processingOperation>
<eamSolvent/>	XXXXXX	XXXXXX	<software>
<eamDilution/>	XXXXXX	...	...
<eamVolume/>	XXXXXX	<massCalibration>	XXXXXX
<arrayTypeName>	**<plateModel>**	<massCalibrationEquation>	XXXXXX
<arrayTypeSpotCount>	XXXXXX	<massCalibrationInfo>	XXXXXX
<arrayBarcode>	**<maldi plateID>**	<dateCalibrated>	XXXXXX
<arrayTypeFormat>	**<pattern>**	...	...
<spotIndex>	**<maldi spotId>**	<intensityUncertainty> (missing in both)	<intensityUncertainty> (missing in both)
XXXXXX (Ciphergen)	<robotManufacturer>	<positionUncertainty> (missing in both)	<positionUncertainty> (missing in both)

Note: Not all items in either format have an equivalent item in the other, and neither provides certain information critical for complete, valid analytic processing. Elements in bold are included in both formats. Missing elements are indicated by 'XXXXXX'.

XML itself is not a magic bullet

XML is a syntactic system to let different programs communicate with each other explicitly. As a result of its simplicity and flexibility, an explosion of XML standards or proposals has emerged in the past 2 years, from molecular descriptions to patient records [12–18]. These XML formats do not necessarily solve the data integration problem in bioinformatics [19]. For instance, XML dialects can create new communication barriers between applications by naming the same item differently in different XML schema. Without controlling definitions, vocabulary and semantic interpretation, XML itself does not improve semantic interoperability. However, XML does provide syntactic interoperability and provides standard methods for persistence and data sharing. Furthermore, extensible stylesheet language transformation (XSLT) provides a mechanism to transform from one XML schema to another. For this to be possible, proper design of the XML schema is critical.



m/z Scan	200	…	500	…	2000
1	…	…	…	…	…
…	…	…	…	…	…
1000	…	…	…	…	…

Figure 3. TIC and SIM chromatograms can be calculated from raw liquid chromatography (LC) mass spectrometry (MS) data. The scan number is arranged according to the retention time of the LC process. The TIC chromatogram is a plot of the total ion current in each MS scan against retention time. The SIM chromatogram is a plot of the ion current of a selected mass over time. The SIM plot is more specific than the full-scan TIC plot.
m/z: Mass-to-charge ratio; SIM: Selected ion monitoring; TIC: Total ion current.

An XML schema definition (XSD) specifies the tags and values allowed in the XML. The more defined the XSD becomes, the more defined, clear and/or valuable the marked up data are. At the same time, as an XSD becomes more specific, the language loses flexibility. For XML to become a useful communication language, it must overcome the flexibility restrictions by adequately defining every piece of data to be represented.

Since XML adds metadata to existing data, it also increases the size of the data file. The scale of proteomic data sets causes corresponding mzXML files to be much heavier than the original data. This increase in size can complicate the primary reason for marking up the data in the first place; larger files are more difficult to work with. Transport times, read times and searching, in particular, can easily approach levels that require alternate strategies such as indexing to make manipulations easier. As these tools become necessary, one of the primary benefits of markup languages such as mzXML, a simple format that everyone can leverage, is negated.

Expert commentary

This article has critically reviewed some problems of the current version of mzXML. These issues have not prevented the successful incorporation of mzXML into useful software.

mzXML acts as a bridge, allowing multiple input formats to be converted and incorporated into a common data analysis pipeline. New output formats generated by newly developed instruments can be integrated into a pre-existing analysis framework by converting the machine-specific native output to the mzXML format. The open structure of mzXML documents makes them suitable for data exchange, including data submission to a central repository to support the results presented in a publication. To encourage the adoption of the mzXML format and the further development of related analytic software, the mzXML toolbox includes the following utilities. InsilicosViewer (Insilicos LLC) provides various methods for visualization of data contained in mzXML files. *rap* offers random access to data stored in mzXML files. *File validator* can be used to validate mzXML files against an XML schema. The *mzXML2other* allows conversion of an mzXML formatted file to a tab-delimited text format and input files for proprietary software tools such as SEQUEST (.dta), ProteinLynx (.pkl) and Mascot (.mgf).

mzXML has played a pivotal role in several large-scale open-source proteomic data storage and analysis systems [20–22]. Desiere and his colleagues have developed a framework, systems biology experiment analysis management system (SBEAMS) for collecting, storing and accessing proteomics and microarray data [20]. Gärdén and coworkers have developed a system, Proteios, for storing, organizing, viewing, annotating and analyzing proteomics experiments [21]. Proteios is built on multi-layered and flexible architecture composed of one to several relational databases and one or many java clients. Craig and coworkers have developed an open source system, the global proteomics machine (GPM) for analyzing, storing and validating proteomics information derived from tandem MS [22]. GPM is composed of a data analysis server, a user interface and a relational database. Data are primarily stored in XML files, which, in turn, are stored as objects in the relational database, and the system provides interface to access the data via structured query language. The analysis server provides a standard interface to access the open source search engine X!TANDEM and generate tabular or graphical representations. These systems facilitate large-scale data collection at multiple institutes.

Five-year view

A successful data format needs community support from both the vendors and the users. Major instrument vendors and database searching vendors have already endorsed mzXML, mzData or both (a list can be found in the information resources section). A number of utilities are already available for users to visualize mzXML scans and to convert mzXML to a variety of other formats. mzXML has already been incorporated in several commercial systems. The authors are optimistic that mzXML and its variants will continue to become an accepted and useful standard in proteomics, although the authors note that there is a high turnover rate for data format standards in the analytic chemistry community.

In conclusion, mzXML and mzData have matured to the point that they should be used in any system integration strategy for the design of data-management and data-analysis systems in large-scale proteomics studies. The primary advantage of using a shared format like mzXML is that it will enable data exchange between instruments, data repositories and analysis software. However, until instrument and software vendors provide easy methods for exporting and importing data from mzXML, groups or individuals who anticipate interpreting and analyzing only a few spectra, mzXML is unlikely to provide enough immediate benefits to justify the conversion of data, from a proprietary format used by the software bundled with an instrument, to the mzXML (or mzData) format.

Information resources

- mzXML website
  http://sashimi.sourceforge.net/software_glossolalia.html
  (Viewed November 2005)
- Technical documentation of mzXML
  http://sashimi.sourceforge.net/schema_revision/mzXML_2.1/Doc/mzXML_2.1_tutorial.pdf
  (Viewed November 2005)
- HUPO PSI website, including mzData and mzIdent
  http://psidev.sourceforge.net
  (Viewed November 2005)
- Minimum Information About a Proteomics Experiment
  http://psidev.sourceforge.net/gps
  (Viewed November 2005)
- Vendors and software supporting mzXML or mzData
  http://psidev.sourceforge.net/ms
  (Viewed November 2005)
- Functional Genomics Experiment (FuGE) Data Model
  http://fuge.sourceforge.net
  (Viewed November 2005)

Key issues

- mzXML (extensible markup language) is a pioneering data exchange format to integrate various components in proteomics data analysis.
- mzXML does not capture all the raw data, leaving some data unavailable in a machine-independent format.
- mzXML is not sufficient for regulatory submission, and thus does not fulfill legal aspects of maintaining data.
- mzXML is not an appropriate data structure for computation, but is suitable for data exchange.
- mzXML does not capture the overall experiment design.

References

Papers of special note have been highlighted as:
• of interest
•• of considerable interest

1 Pedrioli PG, Eng JK, Hubley R *et al.* A common open representation of mass spectrometry data and its application to proteomics research. *Nature Biotechnol* 22, 1459–1466 (2004).
•• Primary reference of mzXML.

2 Baumbach JI, Davies AN, Lampen P, Schmidt H. JCAMP-DX. A standard format for the exchange of ion mobility spectrometry data – (IUPAC recommendations 2001). *Pure Appl. Chem.* 73, 1765–1782 (2001).

3 Kramer GW. ANIML: analytical information markup language for spectroscopy and chromatography data. *Abstracts Of Papers Of The American Chemical Society* 226, U304–U304 (2003).

4 Orchard S, Hermjakob H, Julian RK *et al.* Common interchange standards for proteomics data: public availability of tools and schema. *Proteomics* 4, 490–491 (2004).
• Standardization effort by Human Proteome Organisation Proteomics Standards Initiative.

5 Baggerly KA, Morris JS, Coombes KR. Reproducibility of SELDI-TOF protein patterns in serum: comparing data sets from different experiments. *Bioinformatics* 20, 777–785 (2004).

6 Savitski MM, Ivonin IA, Nielsen ML, Zubarev RA, Tsybin YO, Hakansson P. Shifted-basis technique improves accuracy of peak position determination in Fourier transform mass spectrometry. *J. Am. Soc. Mass Spectrom.* 15, 457–461 (2004).

7 Takach EJ, Hines WM, Patterson DH *et al.* Accurate mass measurements using MALDI-TOF with delayed extraction. *J. Protein Chem.* 16, 363–369 (1997).

8 Ball C, Brazma A, Causton H *et al.* An open letter on microarray data from the MGED Society. *Microbiology* 150, 3522–3524 (2004).

9 Liotta LA, Lowenthal M, Mehta A *et al.* Importance of communication between producers and consumers of publicly available experimental data. *J. Natl Cancer Inst.* 97, 310–314 (2005).

10 Anderson NL, Polanski M, Pieper R *et al.* The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol. Cell. Proteomics* 3, 311–326 (2004).

11 Jones A, Hunt E, Wastling JM, Pizarro A, Stoeckert CJ Jr. An object model and database for functional genomics. *Bioinformatics* 20, 1583–1590 (2004).

12 Leif RC, Leif SB, Leif SH. CytometryML, an XML format based on DICOM and FCS for analytical cytology data. *Cytometry A* 54, 56–65 (2003).

13 Westbrook J, Ito N, Nakamura H, Henrick K, Berman HM. PDBML: the representation of archival macromolecular structure data in XML. *Bioinformatics* 21, 988–992 (2005).

14 Kikuchi N, Kameyama A, Nakaya S *et al.* The carbohydrate sequence markup language (CabosML): an XML description of carbohydrate structures. *Bioinformatics* 21, 1717–1718 (2005).

15 Kasukawa T, Bono H, Hayashizaki Y, Okazaki Y, Matsuda H. MaXML: mouse annotation XML. *In Silico Biol.* 4, 7–15 (2004).

16 Garwood KL, Taylor CF, Runte KJ, Brass A, Oliver SG, Paton NW. Pedro: a configurable data entry tool for XML. *Bioinformatics* 20, 2463–2465 (2004).

17 Goldberg IG, Allan C, Burel JM *et al.* The Open Microscopy Environment (OME) Data Model and XML file: open tools for informatics and quantitative analysis in biological imaging. *Genome Biol.* 6, R47 (2005).

18 Barbera F, Ferri F, Ricci FL, Sottile PA. The Cadmio XML healthcare record. *Stud. Health Technol. Inform.* 90, 277–281 (2002).

19 Stein L. Creating a bioinformatics nation. *Nature* 417, 119–120 (2002).

20 Desiere F, Deutsch EW, Nesvizhskii AI *et al.* Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. *Genome Biol.* 6, R9 (2005).

- Application of mzXML in proteomics analytic system integration.

21 Gärdén P, Alm R, Häkkinen J. Proteios: an open source proteomics initiative. *Bioinformatics* 21, 2085–2087 (2005).

- Application of mzXML in proteomics analytic system integration.

22 Craig R, Cortens JP, Beavis RC. Open source system for analyzing, validating, and storing protein identification data. *J. Proteome Res.* 3, 1234–1242 (2004).

- Application of mzXML in proteomics analytic system integration.

Affiliations

- *Simon M Lin, MD*
  Northwestern University, Robert H Lurie Cancer Center, Chicago, IL 60611, USA
  Tel.: +1 312 695 1331
  Fax: +1 312 695 1347
  s-lin2@northwestern.edu
- *Lihua Zhu, PhD*
  Northwestern University, Robert H Lurie Cancer Center, Chicago, IL 60611, USA
  Tel.: +1 312 695 1333
  Fax: +1 312 695 1347
  l-zhu2@northwestern.edu
- *Andrew Q Winter*
  Northwestern University, Robert H Lurie Cancer Center, Chicago, IL 60611, USA
  Tel.: +1 312 695 1499
  Fax: +1 312 695 1347
- *Maciek Sasinowski, PhD*
  INCOGEN Inc., Williamsburg,
  VA 23185, USA
  Tel.: +1 757 221 0550
  Fax: +1 757 221 0117
  maciek@incogen.com
- *Warren A Kibbe, PhD*
  Northwestern University, Robert H Lurie Cancer Center, Chicago, IL 60611, USA
  Tel.: +1 312 695 1334
  Fax: +1 312 695 1347
  wakibbe@northwestern.edu