analytical chemistry

Technical Note

A Space Efficient Direct Access Data Compression Approach for Mass Spectrometry Imaging

Patrik Källback, Anna Nilsson, Mohammadreza Shariatgorji, and Per E. Andren

Anal. Chem., **Just Accepted Manuscript** • DOI: 10.1021/acs.analchem.7b03188 • Publication Date (Web): 23 Feb 2018

Downloaded from http://pubs.acs.org on February 23, 2018

ACS Publications

Analytical Chemistry is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

A Space Efficient Direct Access Data Compression Approach for Mass Spectrometry Imaging

(Technical Note)

Patrik Källback, Anna Nilsson, Mohammadreza Shariatgorji, and Per E. Andrén*

Biomolecular Mass Spectrometry Imaging, National Resource for Mass Spectrometry Imaging, Science for Life Laboratory, Department of Pharmaceutical Biosciences, Uppsala University, Box 591 BMC, 75124, Uppsala, Sweden.

* Corresponding author: Tel: +46-70-167 9334; E-mail address: per.andren@farmbio.uu.se

Abstract

Advances in mass spectrometry imaging that improve both spatial and mass resolution are resulting in increasingly larger data files that are difficult to handle with current software. We have developed a novel near-lossless compression method with data entropy reduction that reduces the file size significantly. The reduction in data size can be set at four different levels (coarse, medium, fine, and superfine) prior to running the data compression. This can be applied to spectra, spectrum-by-spectrum, or on the transpose arrays, array-by-array, to efficiently read the data without decompressing the whole dataset. The results show that a compression ratio of up to 5.9:1 was achieved for data from commercial mass spectrometry software programs and 55:1 for data from our in-house developed msIQuant program. Comparing the average signals from regions of interest, the maximum deviation was 0.2% between compressed and uncompressed datasets with coarse accuracy for the data entropy reduction. In addition, when accessing the compressed data by selecting a random *m/z* value using msIQuant, the time to update an image on the computer screen was only slightly increased from 92 (±32) ms (uncompressed) to 114 (±13) ms (compressed). Furthermore, the compressed data can be stored on readily accessible servers for data evaluation without further data reprocessing. We have developed a space efficient, direct access data compression algorithm for mass spectrometry imaging, which can be used for various data-demanding mass spectrometry imaging applications.

Introduction

Mass spectrometry imaging (MSI) is a powerful technique for studying the spatial distribution of biomolecules [1] such as lipids [2,3], proteins [1,4], peptides [5], endogenous metabolites and neurotransmitters [6,7], as well as pharmaceutical substances [8-10] directly in tissue sections. Powerful MS techniques, such as Fourier transform ion cyclotron resonance (FTICR), present new opportunities for the imaging of molecules because of the high mass resolution capabilities that enable the extraction of thousands of ion images during a single MSI experiment. Furthermore, the spatial resolution of matrix-assisted laser desorption ionization (MALDI) MSI is currently <5 µm [11] and, for high resolution nanoscale secondary ion mass spectrometry (nanoSIMS), it is <100 nm [12]. This means that MSI instruments that are capable of high mass resolution and high spatial resolution may produce several hundred gigabytes (GB) of data in an experiment on a large tissue section (>100,000 pixels) [13].

It is a significant challenge to handle such large amounts of data and to extract data for visualization efficiently [13,14]. Several strategies to reduce the amount of data have been developed by removing zero intensity data in the mass spectra [15], by converting profile mode spectra into centroid mode data [16] or by reducing the dimensionality of the data by compressed PCA process [17], or by wavelet transform [18]. It can also be compressed using lossless compression to minimize storage size [14]. It is also reported that intensity values from mass spectra is typecast to integer values before compression to reduce the datasize [19]. Normally, the size of MSI data is also reduced using different strategies to fit the data in the internal memory of a computer. Strategies include re-binning profile data [20], creating centroid data [16] or by reducing the number of displayed pixels [20]. Other strategies involve transposing the MSI data, which allows efficient access without performing any data reduction [13] or to divide the MSI data into a number of chunks with limited mass range [14].

Our in-house developed software msIQuant [13] stores data, non-reduced and non-compressed, spectrum-by-spectrum; their transposes are utilized for rapid visualization in real-time. The drawback of this data structure is that a large storage space is needed. Here we report a method that efficiently compresses MSI data and their transposes, enabling fast visualization of images in real-time and space efficient data archiving. Each spectrum in the data is compressed using our compression method. To store data with a high degree of compression, low data entropy is needed [21]. Our algorithm reduces the data entropy before compression by reducing the number of significant figures in the intensity values [22], while leaving the number of *m/z* channels and the precision of the *m/z* values in the spectra unaffected. This strategy is generic and can be implemented on many types of spectral data. The advantage of our method is that both baseline noise and high intensity values are maintained.

Material and Methods

MSI datasets

Four datasets were used for this study (Table 1), with datasets 1 and 2 acquired from mouse brain tissue sections (unpublished data). Datasets 3 and 4 are from our previously published work [6,23]. All animal studies were carried out in accordance with the European Communities Council Directives of 1986 (86/609/EEC) and 2010 (2010/63/EU) and were approved by the local ethics committees on animal experiments. The MSI experiments used MALDI time-of-flight (TOF) (Ultraflextreme, Bruker Daltonics, Bremen, Germany; dataset 1), MALDI quadrupole-TOF (Q-TOF) (Synapt G2si, Waters Corp., Manchester, UK; dataset 2), desorption electrospray ionization (DESI)-Q-TOF (Synapt G2si, Waters Corp.; dataset 3), and MALDI FTICR instruments (Solarix XR, 12T, Bruker Daltonics; dataset 4).

Data Processing

All datasets were converted to imzML format [15] with the software used in the MSI experiment (see Table 1). The imzML converter used for dataset 1 and 4 was the software flexImaging (Bruker Daltonics) which generated double precision floating-point for both *m/z* and intensity values. Dataset 2 and 3 were generated by the imzML converter available in Masslynx (Waters Corp) which produced imzML data with single precision for both *m/z* and intensity values. Each imzML file was converted into the msIQuant format (uncompressed) where the *m/z* values are presented as double precision and the intensities as single precision.

Algorithms for data entropy reduction and compression

Next, the data entropy was reduced and compressed datasets were created with the msIQuant data compression algorithm. The algorithms for data entropy reduction and compression are described in detail in the Supporting Information and are divided into five different algorithms. The data entropy reduction can be carried out at four different levels: coarse accuracy (10 bits), medium accuracy (13 bits), fine accuracy (16 bits), and superfine accuracy (20 bits). The uncompressed accuracy is defined using a single precision floating point value.

Compression and Decompression Method

The data compression and decompression was derived from the lossless compression method Lempel-Ziv-Markov chain algorithm (LZMA) Software Development Kit [24], which is based on LZ77 compression [25].

Test Software

Two versions of msIQuant were used to investigate the changes in performance following data compression: msIQuant (version 2.0.1.14) [13] and a modified version of msIQuant, adapted for the compressed data structure.

Test Procedure

The four datasets were used to evaluate the performance and quality of the data entropy reduction and the compression algorithm. The following evaluations were carried out: a) the compression ratio was compared for both raw data from commercial MS software and msIQuant uncompressed data, and b) the relative deviation between an uncompressed and a compressed spectrum, which was evaluated by calculating the root mean square (RMS). The spectrum with the highest total ion current (TIC) was selected from each dataset because they contain the highest degree of information. RMS was used instead of average values since the relative deviation was close to zero and produced both positive and negative values. The relative deviation between the uncompressed and compressed spectrum was calculated as $\frac{(I_c[i]-I_u[i])}{I_u[i]}$, where $I_c[i]$ is the intensity from the compressed dataset and $I_u[i]$ is the intensity from the uncompressed dataset. Thus, the resulting equation for calculating the RMS for the relative deviation was $\mathrm{RMS} = \sqrt{\frac{1}{n}\sum_{i=0}^{n-1}\left(\frac{(I_c[i]-I_u[i])}{I_u[i]}\right)^2}$. A further evaluation was carried out: the image of an ion of interest in datasets 3 and 4 was selected and compared between the compressed and uncompressed datasets. The similarity ratio between the compressed and uncompressed dataset was calculated as $\left(1-\frac{RMS}{256}\right)$, and the RMS was defined as the pixel value difference in a grayscale image, which has 256 discrete levels (Supporting Information, Figure S1). The ions of interest, dipalmitoylphosphatidylcholine (DPPC) $[M+K]^+$ (*m/z* 772.5) and derivatized dopamine (DA-DPP) (*m/z* 368.165) (Figure 1) were selected, since both have important biological functions in the brain. More specifically, the average signals from annotated ROIs of the tissue were compared between the compressed and uncompressed datasets. Finally, the compressed dataset 4, stored on a server at Uppsala University, was accessed using a virtual private network (VPN) client from a remote location and the time to read data was measured using a timer function from the msIQuant software, after selecting a random m/z value. The broadband speed at the remote location was 14 Mbit/s across the VPN connection.

Results

To estimate the robustness of the data entropy reduction and compression method, it was tested on various MSI datasets acquired using different ionization methods, mass analyzers, and software (Table 1). Compression ratios using the four data entropy reduction levels used were estimated from both commercial and the msIQuant software. Results from MSI experiments were investigated qualitatively and quantitatively. Qualitative comparison was achieved by estimating a similarity ratio between a compressed and an uncompressed image, and quantitative comparison was achieved by annotating different ROIs in a tissue section and comparing the ROI average ion intensity between the compressed and uncompressed dataset. Finally, individual spectra from compressed and uncompressed datasets were compared to investigate the ion intensity deviation.

Data compression ratios and compression time

Our results showed that after carrying out the data entropy reduction and compression, the compression ratio (using coarse accuracy) ranged from 4.2:1 to 55:1 (Table 1) compared to the uncompressed msIQuant dataset, and 0.9:1 to 5.9:1 compared with raw data not converted to msIQuant. The compression time varies based on the spectra size, data entropy and the compression method (Table 1). The coarse, medium and fine accuracy compression have similar compression times for each data set. Compression times using superfine accuracy and 'LZMA only' are twice as long. The 'zlib only' method has the fastest compression times but also the lowest compression ratios compared to the other methods. When selecting an *m/z* value for image visualization, the time to update the image on the computer screen increased by 22 ms (24%) for the coarse accuracy compression, from an average of 92 (±32) to 114 (±13) ms (measured using the timer in msIQuant).

Intensity deviation using data compression

When comparing the relative deviation of the intensity in a spectrum with the high TIC intensity from each dataset, the RMS values ranged from 2,894 to 4,652 ppm for the coarse accuracy, 278 – 580 ppm for the medium accuracy, 27 – 75 ppm for the fine accuracy and 1.4 – 4.6 ppm for the superfine accuracy (Table 1). The relative unit ppm is used instead of percentage since the relative deviation is low (i.e. 1,000 ppm is equal to 0.1%).

Qualitative and quantitative effects of data compression

The images generated from the uncompressed and compressed datasets are compared by calculating a similarity ratio which is described in Supporting Information, Figure S1. The similarity ratio between an uncompressed and a compressed ion image of DPPC (Figures 1A, B) ranged from 99.61% (coarse accuracy) to 99.99 % (superfine accuracy) (Supporting Information, Figure S1).

Three brain structures, cortex (CTX), caudate-putamen (CPu) and anterior commissure (aca), were selected in an image acquired using DESI-Q-TOF MSI and the average intensities were evaluated for uncompressed and compressed (coarse accuracy) data (Figure 1C). The uncompressed dataset provided the following average intensities (% of maximum intensity): CTX 64.959, CPu 51.181, and aca 25.584. The coarse accuracy compression produced these average intensities (% of maximum intensity): CTX 64.962, CPu 51.182, and aca 25.561. This resulted in a deviation of the coarse from the uncompressed dataset by CTX 0.005%, CPu 0.002%, and aca -0.089%. The deviation between the average intensities for the three brain structures was less than 0.1% (absolute value).

Comparison of an uncompressed and compressed (course accuracy) MALDI FTICR ion image of dopamine (Figures 1D, E) showed that the similarity ratio was 99.77%, and the image quality increased with lower compression rates (medium accuracy, 99.92%, fine accuracy 99.98%, and superfine accuracy 100.00%, see Supporting Information, Figure S-1).

The ion intensities for dopamine were compared for three brain structures in the MALDI FTICR experiment (Figure 1F). The uncompressed dataset provided average intensities (%) of: CPu 66.332, aca 13.290 and nucleus accumbens shell (AcbSh) 59.616. The coarse accuracy compression level provided average intensities (%) of: CPu 66.330, aca 13.261 and AcbSh 59.622. This resulted in a deviation of the coarse from the uncompressed dataset by CPu -0.004%, aca -0.223%, and AcbSh 0.010%. The deviation in this dataset was less than 0.2% (absolute value).

To investigate how the ion intensity is affected by our compression method, spectra from three different pixels in dataset 3 were compared with the uncompressed and the decompressed course accuracy compression (Figure 2). The result showed that the uncompressed and decompressed spectra were consistent through the whole m/z range and from low to high intensity values. For example, the average relative deviation was 4,642 ppm. Additional result (Supporting Information, Figure S2) showed that the uncompressed and decompressed coarse accuracy compression spectra had 99.56% similarity. When compiling a quantile-quantile (Q-Q) plot of the two spectra, the slope was 1.000850, intercept 0.000984 and R2 0.999961, which verified that the two datasets have a common m/z distribution.

Access of compressed dataset housed on a remote server

A compressed data array (coarse accuracy level) on the server at Uppsala University was accessed by a computer running msIQuant at a remote location, for visualization on the computer screen. After selecting a random *m/z* value from dataset 4, the time to update the image on the computer screen was between 94 – 1563 ms, with an average time of 279 ms (measured using the timer function in msIQuant). These values were, as expected, slightly higher compared to the time it takes to visualize the same m/z ratio directly from the computer (92 ms, uncompressed data and 114 ms, coarse accuracy compression).

Discussion

Data files from MSI experiments are rapidly increasing in size, hence our aim was to develop an effective way of compressing the data using a near-lossless [19] compression method. Modern methods for lossless compression are based on the Lempel–Ziv 1977 algorithm (LZ77) compression method [25]. The compression method implemented in our msIQuant software uses the Lempel–Ziv–Markov chain algorithm (LZMA) [24]. The main attribute for a lossless compression algorithm is that it cannot carry out any compression of a dataset with true random numbers because of the Shannon limit [21]. That is, in a mass spectrum, there are sequences that repeat themselves, and hence a lossless compression can be used. However, when a spectrum contains noise, the obtained compression rate is lower and the data entropy for such a spectrum is relatively high [21]. This was demonstrated by processing using a 5-point Savitsky-Golay smoothing filter [26] which was enough to increase the compression ratios, which had the highest effect when coarse accuracy precision was implemented (Supporting Information, Table S1). The exception was the spectrum from dataset 4 where the compression ratio was decreased. This was due to the smoothing filter that introduces more non-zero data in the spectrum for datasets with zero baseline, which increases the data entropy.

The software used to generate MSI raw data in this study (Fleximaging and Masslynx) both use lossless compression methods, but are undocumented since the software structures are proprietary information. The standardized imzML data format [15] can be compressed with a method called zlib [27], but the software OpenMSI [14] can also reduce data with gzip compression [28]. Both compression methods are based on a technique called DEFLATE [29], which in turn is based on the LZ77 algorithm [25].

Our way of reducing data entropy is to reduce the number of significant figures, e.g., changing from representing a floating point value as double precision to a single precision or event to half precision [22] (defined in standard IEEE 754 [30]). This method reduces the number of significant figures from approximately 16 decimal digits to 7 or 3 decimal digits. A mass array with *m/z* values should always be expressed with double precision as the mass accuracy of modern mass spectrometers is increasing. An intensity array, however, may be expressed with less significant figures since an intensity signal may vary from zero to $10^7 – 10^9$ counts. As an example, if an ion intensity has the value of 9.813854×10^7 and this value is represented by a binary value, the exponent can be expressed with 3 bits, while the mantissa needs to be expressed with 24 bits, a total of 27 bits. If the acceptable accuracy of the same ion intensity can be expressed as 9.81×10^7, or even as 9.8×10^7, then the mantissa can be expressed with 7 bits and the exponent with 3 bits, so the total size is now 10 bits. This process of reducing the number of bits of a binary value from 27 to 10 bits is the same as reducing the data entropy. Details about the algorithms for data entropy reduction and compression and implementation are described in Supporting Information (Algorithms including pseudo-code). The algorithms were written in C++ to show the pseudo-code as programming code.

There are other compression methods for MSI data available, e.g., ‘randomized approximation compressed PCA process’ and ‘numerical compression schemes’. The compressed PCA process reported by Palmer et al. [17] reduces the dimensionality of the MSI data and has a compression ratio > 170:1, which is much higher than our compression ratio (~5:1). The difference between the two methods is that the compressed PCA process is lossy. The method can recreate spectrum peaks but cannot recreate the noise and the fine structure in the spectra.

The numerical compression schemes for MS data reported by Teleman et al. [19], converts the intensities to integer values before compressing the data with gzip or zlib compression methods. The two main algorithms described in the paper are MS Numpress positive integer compression (numPic), and MS Numpress short logged float compression (numSlof). The numPic method truncates the intensity to nearest integer, while numSlof method takes the natural logarithm of the intensity multiplied by a scaling factor before truncating the resulting value to nearest integer. The numSlof method is similar to our algorithm but with the difference that our method use four different ranges for the integer that optimizes compression. The compression method that is used in our algorithm is LZMA which give higher compression ratio than gzip or zlib.

Our results showed that compressing data at the coarse accuracy level produced a compression ratio up to 5.9:1 when compared to data obtained using commercial software (Table 1). However, for dataset 1, the compression ratio was only 0.9:1 for the msIQuant software, meaning that the compressed data was larger. The reason for this is that the compressed msIQuant data contain both spectra and transposed data [13] but the commercial software raw data contain only spectral data. When comparing just the spectrum data size, the file size from commercial software was 3.53 GB and the msIQuant coarse accuracy compressed spectra data file size was 2.21 GB (not listed in Table 1), giving a compression ratio of 1.6:1. The maximum deviation was 0.2% when comparing the average ion intensity from annotated ROIs between uncompressed and coarse accuracy compression (Figure 1).

To investigate the effectiveness of the data entropy reduction and compression algorithm, datasets compressed with LZMA and zlib methods without any reduction of the precision were compared with our compression method. LZMA compression only produced 1 – 42% larger file sizes compared with superfine accuracy, while zlib compression produced 39 – 52% larger files. The benefit with zlib compared with LZMA is that the method is 100 – 350 % faster, but with the cost of less effective compression ratio.

Images compared using image differential analysis to show differences (Figures 1A to 1B and Figures 1D to 1E and Supporting Information, Figure S-1) resulted in an image similarity ratio of 99.61% and 99.77% between uncompressed and coarse accuracy compression. To compare the technical variation with our compression algorithm, three consecutive coronal rat brain tissue sections were investigated (Supporting Information, Figure S3A-F). The average intensities of the lipid PC(32:0) $[M+K]^+$ (*m/z* 772.5 ± 0.1) from two different brain areas (striatum and cortex) were measured. The standard deviations from intensity measurement of the technical replicates were 0.519 and 0.821, while the average absolute deviations between the uncompressed and the decompressed coarse accuracy compression were 0.000466 and 0.000387 respectively. This demonstrates that the loss in accuracy using our compression method is 1,000-fold lower than the technical variation in such experiment. Further, analyzing the ion intensity of individual spectra from each dataset showed that the relative deviation decreased by one order of magnitude for each degree of accuracy i.e. the medium level has a higher degree of accuracy than the coarse level (Table 1, Relative deviation from uncompressed spectrum). Apart from the reduced data size, another important feature of the compression algorithm is that the data can be directly accessed without the need for uncompressing or reprocessing the dataset. This means that the data can be stored on a server while still being accessible without moving them to a local computer for evaluation.

The data entropy reduction and compression algorithm described in this study is not only valid for compressing MSI spectra, but could also be implemented in all various scientific methodologies where spectra are obtained, such as Raman spectroscopy [31], multi-modal imaging analysis [32], and hyperspectral imaging surveillance by satellites [33]. A similar approach to reducing the number of significant figures has been described for computerized calculations of simulating material stress, which generates large amount of data [34].

In summary, we have developed a space efficient, direct access data compression algorithm for mass spectrometry imaging, which could also be used for various other applications where spectra are generated. We have tested both the efficiency and the accuracy of the compression, which can be set to four different levels: coarse, medium, fine and superfine. We have also proven that the compressed datasets can be stored on a server and still be accessed for data analysis without moving them to a local computer for data processing.

Associated Content

Supporting Information

The Supporting Information is available free of charge on the publication's website at DOI:.

Author Information

Corresponding Author

Email: per.andren@farmbio.uu.se

ORCID

Per E. Andrén: 0000-0002-4062-7743

Author Contribution

The study and manuscript was carried out and written with contributions from all the authors. All authors approved the final version of the manuscript.

Notes

The authors declare no competing financial interest.

Acknowledgments

We thank Daniil Sarkisyan, Dept. of Pharmaceutical Biosciences, Uppsala University, for consultation about pseudocode annotation. This work was supported by the Swedish Research Council (Medicine and Health #2013-3105, Natural and Engineering Science #2014-6215), the Swedish Brain Foundation, the Swedish Foundation for Strategic Research #RIF14-0078, and a Science for Life Laboratory (SciLifeLab) grant.

Figures

Figure 1. Quality comparisons between images extracted from the uncompressed dataset and images from the coarse accuracy (10 bit) dataset of a coronal rat brain tissue section. Tissue sections showing dipalmitoylphosphatidylcholine (DPPC) $[M+K]^+$, *m/z* 772.5 from (A) uncompressed and (B) coarse accuracy compressed data acquired at a spatial distribution of 150 µm. The images were generated using DESI coupled to a Q-TOF instrument. (C) Three brain regions (cerebral cortex (CTX), caudate-putamen (CPu), and the anterior part of the anterior commissure (aca) were selected to measure the average intensity of DPPC $[M+K]^+$. The average intensity deviations were compared between the uncompressed and the coarse accuracy level compression (0.005%, 0.002% and -0.089% for CTX, CPu and aca, respectively). Dopamine (DA) derivatized with DPP-TFB, m/z 368.165 is shown from the (D) uncompressed dataset and (E) coarse accuracy compressed data acquired at a spatial distribution of 150 µm using a MALDI FTICR mass spectrometer. (F) Three brain regions (CPu, aca and nucleus accumbens shell (AcbSh) were selected to measure the average intensity of derivatized DA, and the average intensity deviations were compared between the uncompressed and the coarse accuracy level compression (-0.004%, -0.223% and -0.010% for CPu, aca and AcbSh respectively). The blue-red rainbow color scale represents the intensity from 0 – 100 %.

320

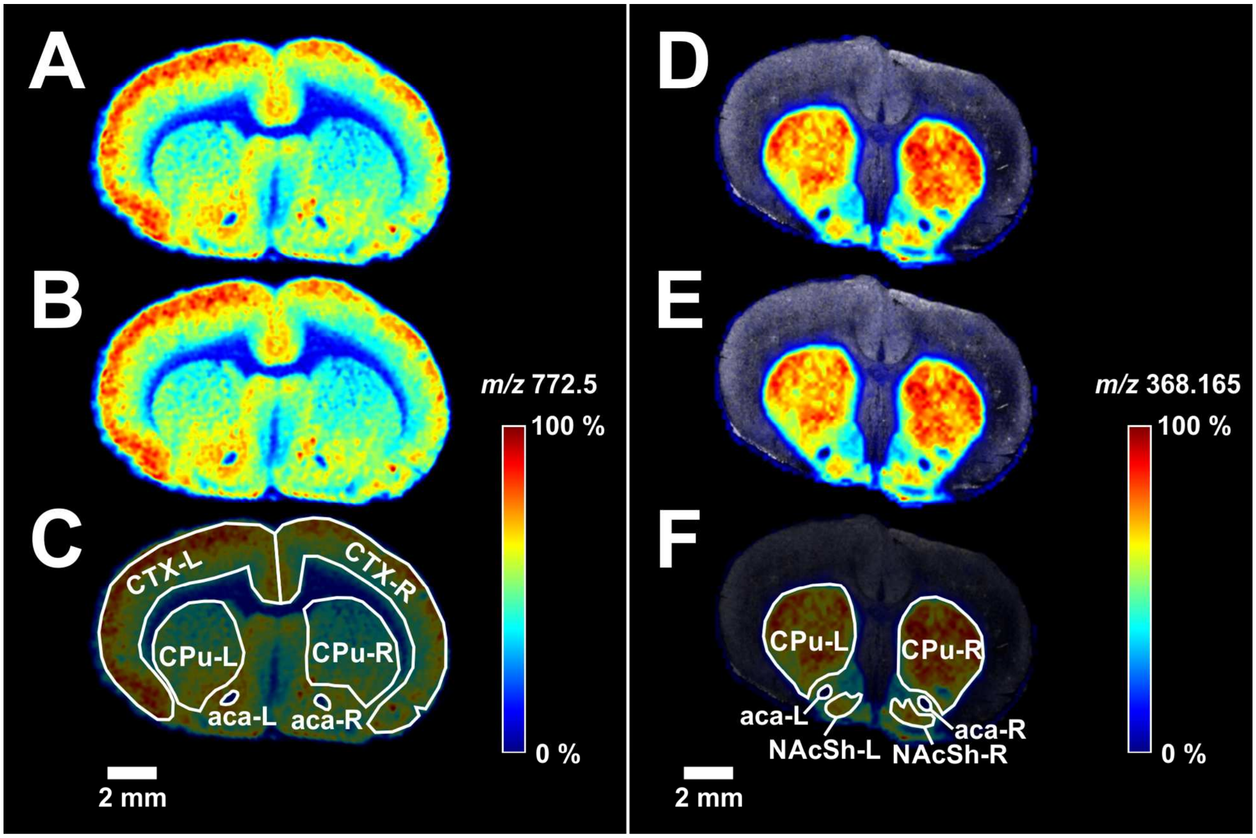
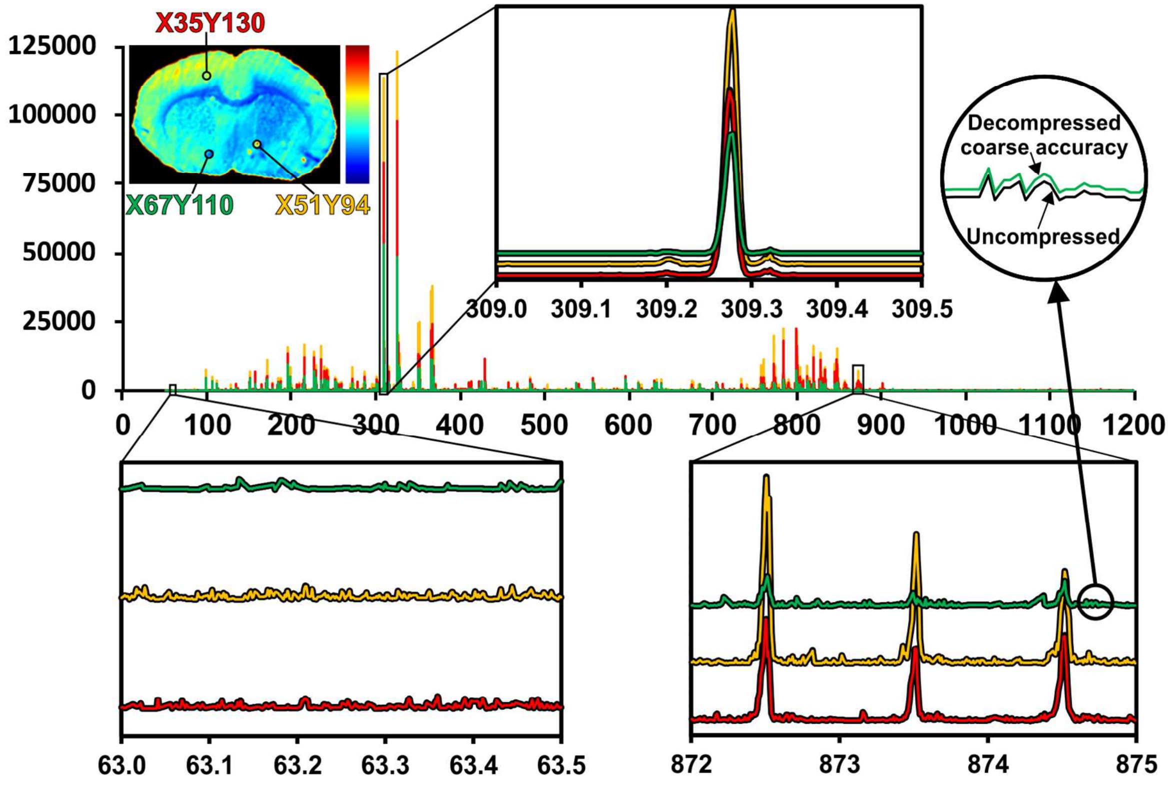

A
B
C
CTX-L
CTX-R
CPu-L
CPu-R
aca-L
aca-R
2 mm
m/z 772.5
100 %
0 %
D
E
F
CPu-L
CPu-R
aca-L
aca-R
NAcSh-L
NAcSh-R
2 mm
m/z 368.165
100 %
0 %

Figure 2. Comparison of mass spectra from three pixels in a brain tissue section before and after compression using the coarse accuracy level compression level.

Overlay of uncompressed (solid black) and decompressed coarse accuracy (red, yellow and green) spectra from three pixels shown in the inset tissue image, displaying TIC-normalization-factor distribution [35]. Three rectangular enlargements, from different parts of the *m/z* axis show that decompressed coarse accuracy spectra maintain the same fine structure as the uncompressed spectra. The spectra from the three pixels are displayed with intensity offset to distinguish them from each other. Each uncompressed spectrum is displayed with a thicker line than the decompressed coarse accuracy spectrum. The circular enlargement displays a small part of the uncompressed and the decompressed spectrum with a small intensity offset.



Table

Table 1. Properties of four MSI datasets acquired from different MSI platforms.

The original sizes of the data files are shown together with the compressed files at four different levels of accuracy. The compression ratio in relation to the instrument-specific software used to acquire the data is shown in brackets, while the compression ratio in relation to the msIQuant files is shown in bold in brackets. The compression times of msIQuant files are shown as hours and minutes (hr:min) and the measurements were performed on an HP Z440 workstation. The relative deviation of ion intensity between compressed and uncompressed data for a single spectrum is shown as RMS (in ppm) at the four different levels of accuracy.

	Dataset 1	**Dataset 2**	**Dataset 3**	**Dataset 4**
Ion Source	MALDI	MALDI	DESI	MALDI
Mass Analyzer	TOF	Q-TOF	Q-TOF	FTICR
Spatial resolution [µm]	50	100	150	150
Number of pixels	54,808	19,370	28,718	12,411
Number of m/z channels	84,821	110,376	310,734	372,516
imzML converter	Fleximaging	Masslynx	Masslynx	Fleximaging
Raw data size [GB]*	4.17	11.28	37.35	3.7
msIQuant data size [GB]	***34.67***	***15.94***	***66.51***	***34.46***
Coarse data size [GB]	4.57 (0.9:1) ***(7.6:1)***	3.82 (3.0:1) ***(4.2:1)***	14.49 (2.6:1) ***(4.6:1)***	0.63 (5.9:1) ***(54.7:1)***
Medium data size [GB]	4.71 (0.9:1) ***(7.4:1)***	5.36 (2.1:1) ***(3.0:1)***	22.11 (1.7:1) ***(3.0:1)***	0.79 (4.7:1) ***(43.6:1)***
Fine data size [GB]	4.79 (0.9:1) ***(7.2:1)***	6.87 (1.6:1) ***(2.3:1)***	29.05 (1.3:1) ***(2.3:1)***	0.93 (4.0:1) ***(37.1:1)***
Superfine data size [GB]	5.05 (0.8:1) ***(6.9:1)***	8.80 (1.3:1) ***(1.8:1)***	38.51 (1.0:1) ***(1.7:1)***	1.24 (3.0:1) ***(27.8:1)***
LZMA only [GB]	5.08 (0.8:1) ***(6.8:1)***	11.79 (1.0:1) ***(1.4:1)***	54.54 (0.7:1) ***(1.2:1)***	1.63 (2.3:1) ***(21.1:1)***
Zlib only [GB]	7.01 (0.6:1) ***(4.9:1)***	12.90 (0.9:1) ***(1.2:1)***	58.25 (0.6:1) ***(1.1:1)***	1.72 (2.2:1) ***(20.0:1)***
Coarse comp. time [hh:mm]	01:29	00:29	01:44	00:19
Medium comp. time [hh:mm]	01:19	00:30	01:52	00:20
Fine comp. time [hh:mm]	01:15	00:28	01:43	00:20
Superfine comp. time [hh:mm]	02:32	00:57	03:51	00:30
LZMA only comp. time	02:46	00:54	03:27	00:28
Zlib only comp. time	00:37	00:22	01:41	00:06
Coarse, RMS [ppm]	3,222	4,637	4,652	2,894
Medium, RMS [ppm]	529	580	580	278

Fine, RMS [ppm]	75	72	72	27
Superfine, RMS [ppm]	1.4	4.5	4.6	1.4

** The raw data size refers to vendors' proprietary data.*

References

(1) Caprioli, R. M.; Farmer, T. B.; Gile, J. *Anal Chem* **1997**, *69*, 4751-4760.

(2) Touboul, D.; Piednoel, H.; Voisin, V.; De La Porte, S.; Brunelle, A.; Halgand, F.; Laprevote, O. *European journal of mass spectrometry* **2004**, *10*, 657-664.

(3) Woods, A. S.; Jackson, S. N. *The AAPS journal* **2006**, *8*, E391-395.

(4) Stoeckli, M.; Chaurand, P.; Hallahan, D. E.; Caprioli, R. M. *Nature medicine* **2001**, *7*, 493-496.

(5) Skold, K.; Svensson, M.; Nilsson, A.; Zhang, X.; Nydahl, K.; Caprioli, R. M.; Svenningsson, P.; Andren, P. E. *J Proteome Res* **2006**, *5*, 262-269.

(6) Shariatgorji, M.; Nilsson, A.; Goodwin, R. J.; Kallback, P.; Schintu, N.; Zhang, X.; Crossman, A. R.; Bezard, E.; Svenningsson, P.; Andren, P. E. *Neuron* **2014**, *84*, 697-707.

(7) Sugiura, Y.; Setou, M. *J Neuroimmune Pharm* **2010**, *5*, 31-43.

(8) Nilsson, A.; Fehniger, T. E.; Gustavsson, L.; Andersson, M.; Kenne, K.; Marko-Varga, G.; Andren, P. E. *PLoS One* **2010**, *5*, e11411.

(9) Nilsson, A.; Goodwin, R. J.; Shariatgorji, M.; Vallianatou, T.; Webborn, P. J.; Andren, P. E. *Anal Chem* **2015**, *87*, 1437-1455.

(10) Reyzer, M. L.; Hsieh, Y.; Ng, K.; Korfmacher, W. A.; Caprioli, R. M. *J Mass Spectrom* **2003**, *38*, 1081-1092.

(11) Guenther, S.; Rompp, A.; Kummer, W.; Spengler, B. *Int J Mass Spectrom* **2011**, *305*, 228-237.

(12) Herrmann, A. M.; Ritz, K.; Nunan, N.; Clode, P. L.; Pett-Ridge, J.; Kilburn, M. R.; Murphy, D. V.; O'Donnell, A. G.; Stockdale, E. A. *Soil Biol Biochem* **2007**, *39*, 1835-1850.

(13) Kallback, P.; Nilsson, A.; Shariatgorji, M.; Andren, P. E. *Anal Chem* **2016**, *88*, 4346-4353.

(14) Rubel, O.; Greiner, A.; Cholia, S.; Louie, K.; Bethel, E. W.; Northen, T. R.; Bowen, B. P. *Anal Chem* **2013**, *85*, 10354-10361.

(15) Rompp, A.; Schramm, T.; Hester, A.; Klinkert, I.; Both, J. P.; Heeren, R. M.; Stockli, M.; Spengler, B. *Methods Mol Biol* **2011**, *696*, 205-224.

(16) Bielow, C.; Gropl, C.; Kohlbacher, O.; Reinert, K. *Methods Mol Biol* **2011**, *719*, 331-349.

(17) Palmer, A. D.; Bunch, J.; Styles, I. B. *Anal Chem* **2013**, *85*, 5078-5086.

(18) Barclay, V. J.; Bonner, R. F.; Hamilton, I. P. *Anal Chem* **1997**, *69*, 78-90.

(19) Teleman, J.; Dowsey, A. W.; Gonzalez-Galarza, F. F.; Perkins, S.; Pratt, B.; Rost, H. L.; Malmstrom, L.; Malmstrom, J.; Jones, A. R.; Deutsch, E. W.; Levander, F. *Mol Cell Proteomics* **2014**, *13*, 1537-1542.

(20) Klinkert, I.; Chughtai, K.; Ellis, S. R.; Heeren, R. M. A. *Int J Mass Spectrom* **2014**, *362*, 40-47.

(21) Shannon, C. E. *Bell Syst Tech J* **1948**, *27*, 379-423, DOI: 10.1002/j.1538-7305.1948.tb01338.x.

(22) Maass, C.; Baer, M.; Kachelriess, M. *Med Phys* **2011**, *38 Suppl 1*, S95.

(23) Shariatgorji, M.; Strittmatter, N.; Nilsson, A.; Kallback, P.; Alvarsson, A.; Zhang, X.; Vallianatou, T.; Svenningsson, P.; Goodwin, R. J.; Andren, P. E. *Neuroimage* **2016**, *136*, 129-138.

(24) Pavlov, I. *LZMA SDK (Software Development Kit)* **1998**, http://www.7-zip.org/.

(25) Ziv, J.; Lempel, A. *IEEE Trans Inf Theory* **1977**, *23*, 337-343.

(26) Savitzky, A.; Golay, M. J. E. *Anal Chem* **1964**, *36*, 1627-1639.

(27) Gailly, J. l.; Adler, M. *zlib* **1995**, https://zlib.net/.

(28) Gailly, J. l.; Adler, M. *gzip* **1992**, https://www.gnu.org/software/gzip/.

(29) Katz, P. W., U.S.A., *Patent: US5051745 A* **1991**.

(30) IEEE Computer Society, IEEE Std 754-2008, *IEEE* **2008**, p 70.

(31) Freudiger, C. W.; Min, W.; Saar, B. G.; Lu, S.; Holtom, G. R.; He, C.; Tsai, J. C.; Kang, J. X.; Xie, X. S. *Science* **2008**, *322*, 1857-1861.

(32) Van de Plas, R.; Yang, J.; Spraggins, J.; Caprioli, R. M. *Nat Methods* **2015**, *12*, 366-372.

(33) Cudahy, T. J.; Hewson, R.; Huntington, J. F.; Quigley, M. A.; Barry, P. S. *Int Geosci Remote Se* **2001**, 314-316.

(34) Thole, C. A. *5th European LSDYNA Conference* **2005**, 1-6.

(35) Deininger, S. O.; Cornett, D. S.; Paape, R.; Becker, M.; Pineau, C.; Rauser, S.; Walch, A.; Wolski, E. *Analytical and Bioanalytical Chemistry* **2011**, *401*, 167-181.

For TOC only