International Journal of Mass Spectrometry 362 (2014) 40–47

Contents lists available at ScienceDirect

International Journal of Mass Spectrometry

journal homepage: www.elsevier.com/locate/ijms

Methods for full resolution data exploration and visualization for large 2D and 3D mass spectrometry imaging datasets

Ivo Klinkert, Kamila Chughtai, Shane R. Ellis, Ron M.A. Heeren*

FOM Institute AMOLF, Science Park 104, 1098 XG Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 27 September 2013
Received in revised form
16 December 2013
Accepted 16 December 2013
Available online 24 December 2013

Keywords:
Mass spectrometry imaging
Visualization
Datacube Explorer
imzML

ABSTRACT

Mass spectrometry imaging (MSI) produces such large amounts of high-resolution data that fast visualization of full data sets in both high spatial and spectral resolution is often problematic. Instrument specific software tools are available, but often struggle with the size and the complexity of the MSI data sets. We describe new methods to improve the handling of these large MSI data sets by means of innovative data structures and visualization strategies developed specifically for MSI. Two new software instruments implement these new methods for rapid data exploration and visualization of both 2D and 3D data sets in full spatial and spectral resolution, and can handle existing MSI data formats, including imzML and BioMap.

1. Introduction

As mass spectrometry imaging (MSI) evolves over the years, the amount of high-resolution data generated by its experiments increases. This increase in data volume requires special-purpose software tools, methods and computing infrastructure to manage the data-intensive nature of the MSI experiments. MSI distinguishes itself from regular mass spectrometry (MS) by the ability to spatially localize compounds across surfaces in the case of 2D imaging [1,2]. In the case of 3D imaging volumes are analyzed by dividing the 3D volume into multiple 2D images [3]. With the introduction of ion mobility separation, even 4D datasets appear, with the ion drift time as the fourth dimension.

MSI visualization software can be classified into three main groups, (1) commercially available software from vendors of MSI instruments for control, acquisition and analysis like WinCadence (Physical Electronics, Inc., Chanhassen, MN), FlexImaging (Bruker Corporation, Billerica, MA) or High Definition Imaging (Waters Inc., Milford, MA), (2) MSI generic tools written in e.g. ImageJ or Matlab like OmniSpect [4] or MSiReader [5], and (3) specially developed stand-alone MSI visualization tools like BioMap (Novartis, Basel, Switzerland) or Mirion [6]. The same kind of classification holds for MSI data formats, (1) vendor-specific, often proprietary un-disclosed data formats, (2) generic, non-MSI data formats like netCDF (http://www.unidata.ucar.edu/netcdf), or (3) dedicated MSI data formats independently developed such as imzML [7], or a biomedical imaging data format like ANALYZE 7.5 (Mayo Foundation), also known as the BioMap format.

An important development toward a uniform data exchange format is the development of imzML [8], based on the HUPO Proteomics Standards Initiative standard mzML for mass spectrometry. mzML is a combination of two different data format standards, mzData and mzXML, where the best of both standards was used to define mzML [9].

Some of the currently available visualization software is limited in its capability to manage the enormous data sizes generated by 2D and 3D MS experiments. Limitations exist in the form of long data loading times for the entire dataset or for switching between different parts of the data sets, visualization at non-maximum resolution, non-optimal data organization, or undisclosed data formats that cannot be accessed by software other than the vendor-specific analysis tools or costly third party software. In the case of visualization software that is not in-house developed, adding new or specific features to the software is most of the time impossible. The development of personalized visualization software can be time-consuming, but it does allow full control over the functionality and can be fully optimized for specific needs. This is one of the reasons we have initiated the development of the Datacube Explorer software.

We developed new methods and strategies for the visualization of MSI data for many different experimental setups, and to fulfill method specific software requirements. These methods take into account the above mentioned issues about visualization software and data formats into account. All methods and strategies described

* Corresponding author. Tel.: +31 20 754 7100; fax: +31 20 7547290.
E-mail address: heeren@amolf.nl (R.M.A. Heeren).

1387-3806/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijms.2013.12.012

here have been analyzed and tested by means of two concrete software applications. These include the Datacube Explorer (DCE), a software visualization and analysis tool for 2D MSI datasets, and the Volume Explorer (VE), a software tool for visualizing data from even larger 3D MSI experiments.

The DCE is able to visualize data stored in the BioMap data format, imzML and the newly developed datacube data format described in this article. The DCE enables researchers to analyze their large MSI data sets, without dealing with computer programming. In research that involves multiple, even experimental MSI equipment, and where a mutually comparable way of visualizing the results is required, a tool like the DCE is desired. This even holds when multiple ionization techniques like MALDI or SIMS, often acquired on separate instruments with their own data structure, are combined, or likewise for different mass spectrometric techniques like time-of-flight MS or Fourier-transform MS are used in concert. Because of the open data formats supported by the DCE, data originating from any instrument or technique can be converted into a common comparable standard that makes a mutual comparison possible.

The Volume Explorer software tool is capable of visualizing 3D datasets through the selected combination of several 2D datasets. Volumes that reveal the 3D distribution of any three mass ranges can be reconstructed within a few seconds without any additional programming efforts, which makes scrolling through 3D data in almost real time possible. For 3D visualization, performance issues are even more important because of the large data sizes and the graphical computer reconstruction [10].

2. Experimental

2.1. 2D and 3D mass spectrometry imaging experiments

The MSI datasets employed to evaluate and test the methods described in this paper were acquired from samples taken from a xenograft breast tumor model as described in [3]. A MDA-MB-231 breast cancer cell line was purchased from the American Type Culture Collection (ATCC) and genetically modified to express a red fluorescent protein (tdTomato) under control of hypoxia response elements as described in [11,12]. The cells were injected into the upper thoracic mammary fat pad of athymic nude mice ($2\times$ 106 cells/injection) and tumor growth was monitored with standard calipers. When the tumors reached a volume of approximately 500 mm^3, the mice were sacrificed and tumors were removed. Each tumor was embedded into a gelatin block and cresyl violet fiducial markers [3] were injected at three different positions inside the block next to the tumor. The block was sectioned into serial 2-mm thick fresh tumor sections which were snap-frozen immediately.

From each 2-mm thick section, 10-μm thick sections were cut at −16 °C using a Microm HM550 cryo-microtome (Microm International GmbH, Walldorf, Germany) and were mounted onto Indium Tin Oxide (ITO) coated slides (Delta Technologies, USA). Consecutive 10-μm thick sections were selected for the 3D-reconstruction, such that the inter-section distance was 0.5 mm.

Before MSI analysis, the tissue sections were briefly washed by immersion in 70% and 90% ethanol and dried in a vacuum desiccator for 10 min. Trypsin was resuspended in water at a concentration of 0.05 μg/μL, and 5 nL per spot in a 150-μm × 150-μm raster was deposited by CHIP (Shimadzu, Japan). CHCA matrix was prepared at a concentration of 10 mg/ml in 1:1 ACN:H_2O/0.1% TFA and was applied by an ImagePrep (Bruker, Germany) application system. Samples were analyzed on a MALDI-Q-TOF (Synapt HDMS, Waters, UK) instrument in time-of-flight (TOF) mode detecting the positive ions. The images were acquired with a raster that employed a 150 μm × 150 μm spacing between individual acquisition points.

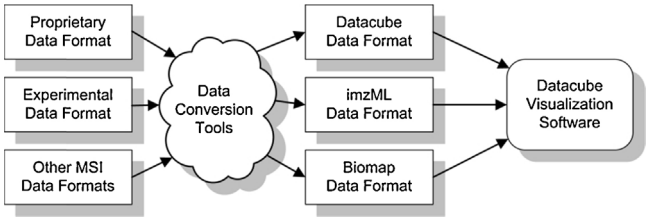

Fig. 1. Data processing workflow. Mass spectrometry imaging data processing workflow for the datacube-based visualization tools.

A pre-release version of Waters' High Definition Imaging (HDI) (Waters Inc., Milford, MA) software was used to convert the Synapt HDMS data into imzML datasets, which was used for additional comparative processing and visualization.

2.2. Software development platform

Both the Datacube Explorer and the Volume Explorer software applications were written in the C# programming language version 4.0 of the .Net platform version 4.0 (Microsoft, Redmond, WA, USA) with Microsoft Visual Studio 2010 as the software development environment. Subversion (http://subversion.apache.org/, The Apache Software Foundation) was used for the central multi-user version control during the software development, with VisualSVN (VisualSVN Limited) and Tortoise (http://tortoisesvn.tigris.org/) as the client side interfaces. Resharper (Jetbrains s.r.o., Prague, Czech Republic) was used as a developer productivity tool for Microsoft Visual Studio.

In addition, The Visualization Toolkit (VTK, Kitware, Inc., Clifton Park, NY) version 4.5 was used for all 3D visualization in the Volume Explorer. VTK is an open-source, freely available software system for 3D computer graphics, image processing and visualization [13].

3. Methods

3.1. MSI visualization

A large variety of data formats for storing mass spectrometry imaging datasets exist and most of these datasets can only be visualized using the corresponding proprietary commercial software tools. Several commercial software tools exist that allow data visualization in other tools by offering data export capabilities to more generic data formats. The 2D and 3D visualization methods presented here use a newly developed 'datacube' data format [14], with import capabilities for both the recently developed imzML and the widely used BioMap data formats.

Fig. 1 shows the principle dataflow used to create a generic approach for visualizing MSI datasets. MSI datasets can be stored in proprietary data formats, experimental data formats, or another open data format. In order to be able to use these data sets they need to be converted into one of the three more generic three datasets types which can be read by the visualization tools presented here. Data conversion tools play a key role in the establishment of a generic MSI visualization approach. These conversion tools can be included in existing (commercial) software tools as data export option, or provided as separate conversion tools.

3.2. Datacube data format

An optimal data structure is the key factor to optimize performance in mass spectrometry imaging (MSI) for handling Gigabytes of data while still delivering high performance in the analysis by the users. Without the correct data organization and processing approach, the analysis of data sets can become very

time consuming, which can be observed in some commercially available software. MSI datasets are commonly stored spectrum-by-spectrum, since the nature of the imaging experiments is to scan per pixel and collect spectra for each of these pixels. This spectra-based sequence is inefficient for image-based visualization, and therefore the datacube data format is image-based where the data is stored image-by-image.

The datacube data format is developed in order to have a dataset format which provides a simple and uniform way to visualize MSI data from several different custom-built or commercial experimental setups. A datacube contains 2D MSI data organized using three dimensions. The first two dimensions are the x and y-coordinates of the measured sample surface, and the third dimension is the m/z axis where the mass spectral information is stored. All positions in the datacube contain ion intensity information (e.g. ion counts), which can be considered the fourth dimension. The actual x and y coordinates are the orthogonal measurement positions, where every position contains a mass spectrum. These positions are called pixels in the datacube, as they are shown as pixels in the selected ion images. The spectral data of a dataset is stored in mass bins in a datacube, where all ion intensities in a very small mass range (the mass bin) are summed together. Therefore, the coordinate in the m/z dimension is the mass bin index, as the m/z axis is organized into fixed-sized mass bins.

The data file itself stores all intensity values of the MSI datacube sequentially, as a series of consecutive images, with the images stored sequentially line by line. Per data file the intensity values can be stored as 8, 16 or 32 bit integer values, as well as 32 bit floating point values. The option of using different data types for the intensity values enables the use of the most optimal data type for different data sources with respect to the total data size of the datacube, which gives the smallest possible dataset with the fastest reading possible by the visualization software, without unintended loss of precision. The Datacube Explorer software application has a built-in function to analyze the current data type and propose a more efficient data type. The entire datacube dataset consists of two corresponding similarly named files, a text-based meta-data file (file extension .ini), and a binary data file (file extension .dat), with an optional text-based configuration file (file extension .cfg). This separation into different file types makes it possible to have the most optimal data reading and processing methods in the software, text – based methods for the small metadata and configuration, and binary methods for the large amount of MSI data. The metadata file is a basic text file which contains a set of properties with their values that describe the dimensions of the data set (x, y and m/z range), the mass resolution of the data and the data type of the intensity values.

The datacube format is highly comparable to the FITS data format [15], but more simplified and optimized for MSI. A full description of the syntax of the open datacube data format can be found in the freely available user manual [14].

3.3. Mosaic data structures

An MSI datacube can be visualized by means of the Datacube Explorer, as long as an entire dataset on disk can be loaded into the memory of the computer at once. Datasets up to 700 Mbytes can be loaded depending on typical mainstream PCs, with a maximum accesable memory limit of 2 GBytes on 64-bit Windows operating systems, due to the maximum array size in the Microsoft .Net platform version 4.0. This limitation does not hold from .Net version 4.5 and up on 64-bit Windows, but to have these requirements rules out users of somewhat older PCs. A solution to this memory limitation, called the mosaic datacube architecture, was developed to handle this problem that can arise with larger datasets, such as those acquired with FTICR-MSI, without a compromise to the original resolution of the dataset [16]. The mosaic architecture splits a dataset into separate adjacent datacubes, which together represent the original full datacube. Re-combining these separate datacubes, together with the functionality of reading any volume-of-interest (VOI) of the full dataset into the DCE, solves the memory problem in a pragmatic way. It makes real-time scrolling to virtually any image at any available spectral resolution of a part of the dataset possible, as illustrated in Fig. 2. Fig. 2a shows a single high-resolution datacube which cannot be visualized due to its data size. Fig. 2b shows the datacube with the same size and high-resolution as Fig. 2b, but with the dataset decomposed into separate smaller datacubes that form the mosaic structure. By virtually recomposing the full data set as shown in Fig. 2c, any VOI out of this large data cube can be visualized at the full high-resolution. Fig. 2d shows the compared required data volumes. The conventional dataset on the left does not fit into memory, where the manageable sub datacubes at full resolution in the middle have no memory problems and together create the entire dataset where a manageable VOI can be extracted, as shown on the right. It is important to realize that the separate datacubes of the 3D mosaic and the VOI are completely independent concepts. The separate datacubes physically store a part of a dataset, where a VOI is an on-demand user selected datacube. The number of datacubes in the 3D mosaic is only limited by the required amount of available disk space for these cubes together, where visualization of a VOI is limited by the internal memory constraints mentioned above.

3.4. 2D visualization

The Datacube Explorer (DCE) performs all 2D visualization independent of the type of the source dataset (Datacube, imzML, or BioMap file). In the DCE, the internal data format in memory is similar to the data format on disk, hence the name Datacube Explorer. In memory, all ion intensity data of the MSI dataset is loaded into a virtual datacube at once, a 1D data array with all images of the mass bins sequentially stored each image line by line. This memory representation enables an extremely fast search and retrieval of image data due to the fixed mass bin size of every image.

Visualization in the DCE is performed both image-based as well as spectrum-based. For the image-based visualization, an ion image is constructed by summing all ion intensities of the mass bins inside a requested mass window per individual pixel of the image. Currently available m/z scrolling steps range from 0.001 to 10 m/z, smaller steps can be implemented into future versions of the DCE when required.

Images are visualized in 8-bit grayscale and by default the pixel intensities of the generated images are scaled up to maximum contrast, the minimum pixel (ion) intensity as black and the maximum pixel (ion) intensity as white. This auto-scaling delivers maximum contrast to the images, additionally, a fixed manual intensity scaling can be applied in order to compare images of different masses on an identical intensity scale. In cases where the square-sized images contain pixels without actual measured spectra, and consequently no ion intensity in these pixels, the auto-scaling algorithm is supplied with an option to ignore these zero-intensity pixels and by that increasing the image contrast. Four types of inter-pixel interpolation are available, based upon the available Microsoft interpolation modes (InterpolationMode Enumeration) in the .Net framework, including bilinear and bicubical interpolation. For spectrum-based visualization, the ion intensities over the entire mass range of the dataset can be visualized for both the summed spectra of all image pixels, and for the three user-defined XY regions-of-interest (ROIs). Spectra visualization of smaller mass ranges is possible by means of a user-selectable mass range, where the center of this user-selected range changes with the selected mass window of the displayed image. Because of the image based

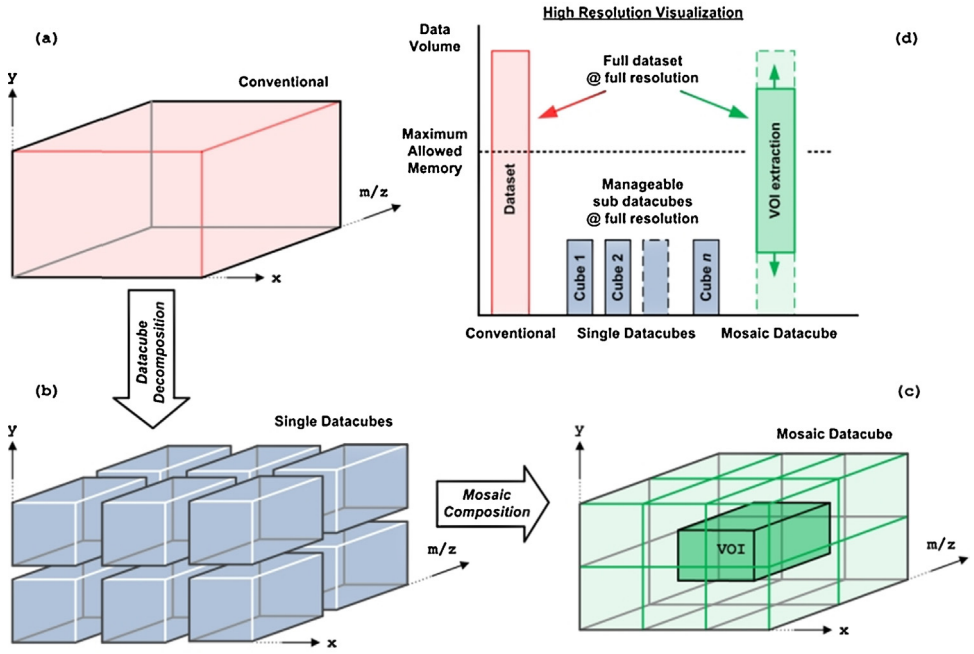

Fig. 2. Principle mosaic datacube structure. Mosaic datacube data structure. (a) Conventional data structure, (b) decomposition of data into sub single datacubes, (c) VOI selection out of the mosaic composed datacubes and (d) data volume comparison between conventional, single datacubes and the mosaic datacube structure.

data organization, preprocessing of spectral data is performed to achieve maximum visualization performance, where for changing ROI definitions a new preprocessing run is performed.

The region-of-interest (ROI) function implements the ability to determine and analyze spectra of selected spatial regions of the sample. For three user-selectable ROIs the corresponding summed spectra for each ROI are shown, as shown in Fig. 3. The three ROI spectra are displayed in different colors (red, green and blue) to distinguish the spectra of the ROI from the summed spectrum of the entire dataset, The intensities of the ROI spectra can be normalized with respect to the selected area of individual ROIs, resulting in comparable spectrum intensities even when the ROIs do not have equal surface areas. The spatial location of the ROIs are displayed in the same colors as overlays on the selected ion image, where the colors of overlapping ROIs are the mixed RGB color channels of the ROIs, with a pixel intensity of the original pixel intensity of the selected ion intensity of the image. The rationale behind the 8-bit grayscale used for the visualization of the individual ion images instead of a color scale, is that a color scale would compromise the use of mixed RGB color channels used to visualize three separate, but possibly overlapping, ROIs.

One of the many possibilities of statistical analysis on datacube datasets is shown with the implemented self-organizing map feature. Self organizing maps, also known as Kohonen networks [17], are unsupervised competitive learning neural networks. The datacube is divided into a set of images of mass resolution-sized mass bins, which represents the dataset spatially to perform this analysis. A two-dimensional Kohonen layer configuration was used to map all these images into the network. Modification of weight vectors of the 2D network was used to separate the most common input images from the rare ones, which results in the clustering of the most abundant images of the MSI dataset. The analysis reveals the characteristic spatial distributions of a dataset which can be used as an input to create ROIs, to analyze their spectral manifestation. These distributions should be entered to the DCE as ROIs to link these characteristic spatial distributions back to spectra. A more detailed description of this feature can also be found in the freely available user manual [14].

3.5. 3D datacube dataset

Typical MSI experiments deal with the spatial, two-dimensional, examination of samples, where mass spectral information from the surface can be obtained, as MSI is a surface measurement technique. In the case that MSI information must be obtained in three dimensions, like for MRI, the 3D sample must be physically segmented in order to perform measurement in the third physical dimension. Among different techniques, one method to be able to perform measurements in the third dimension is to cut the sample in several parallel slices and perform standard MSI measurements on the individual sample slices, which is the case for the measurement results presented in this article.

The 3D datacube dataset was developed to store the data of this type of 3D MSI measurements in a uniform manner. The data is stored on disk in a set of datacubes, where each datacube contains the 2D data of one sample slice. An additional 3D configuration file holds the metadata of this set of datacubes, including the sequence of cubes in the 3D space and the mutual 2D orientations.

3.6. 3D reconstruction and visualization

The successively measured 2D sample data must artificially be reconstructed by software to visualize the molecular distributions in three dimensions. The Volume Explorer (VE) software application implements this reconstruction process. In order to create a correct reconstruction, two image co-registration actions must be performed. First, the mutual spatial alignment of the individual slices must be arranged, to correct for the orientation differences of the slices relative to the positions in which the samples were measured. Second, the normalization of ion intensities of the individual slices must take place.

For the mutual spatial alignment of the individual slices a novel fiducial marker technique was developed [3]. The principle on which spatial alignment can be performed is the presence of detectable markers at known positions outside the tissue sample. These detectable markers are used to mutually align the individual slices by (manual entered) translations of the parallel slices.

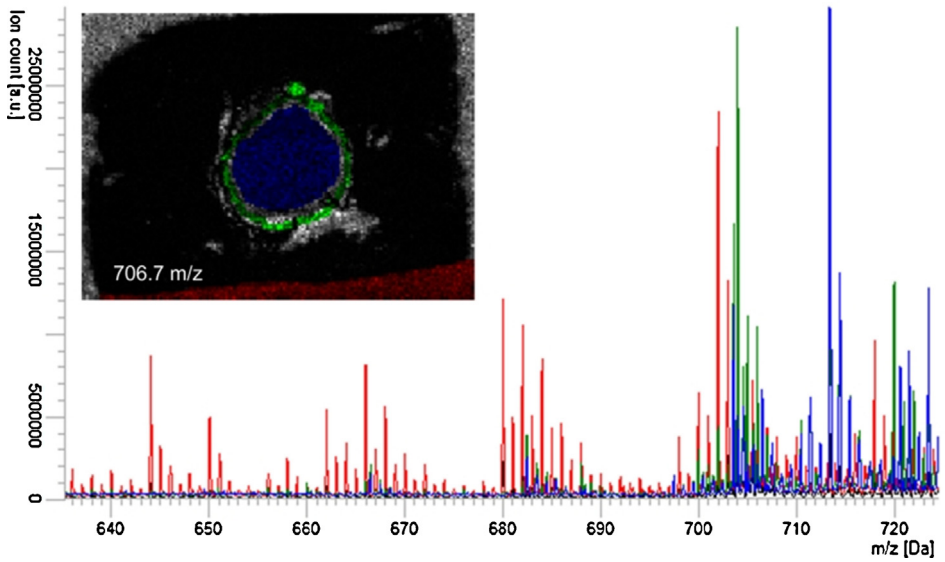

Fig. 3. Datacube Explorer 2D visualization. Visualization results of the Datacube Explorer, including the region-of-interest (ROI) function. The mass spectrum (ion intensity versus m/z) of the dataset is shown in the main image, with the image of ions at m/z 706.7 displayed in the upper left inset. Three user-selectable ROIs are shown in red, green and blue in the upper left inset, as well as their corresponding spectra.

Although not implemented yet in the VE, additional rotation and spatial scaling based on these markers can also easily be performed automatically. The ion intensity normalization is required because a realistic 3D reconstruction is only possible with individual sample slices that have comparable ion intensity levels. The differences in ion intensity levels are mainly caused by the differences in ionization efficiency in the different slices that constitute the 3D dataset. A scaling factor is applied to the ion pixel intensities per slice, where the scaling factors can be determined by analysis of the ion intensities of known compounds [3]. In addition to this realistic 3D reconstruction using scaling factors, an engineering scaling mode was developed where each individual slice is scaled up to the maximum contrast. This produces high contrast images that result in a clearly visible 3D reconstruction. This mode is particularly useful for the alignment of the fiducial markers. The additional 3D configuration file holds both the spatial alignment settings as well as the intensity scaling values.

The enormous amount of data of a 3D dataset cannot be loaded into computer memory fully, contrary to the DCE where the datasets to be visualized are loaded into memory at once. For the 3D visualization the required data will be read from file on demand. Acceptable read time performance of 1–5 s is achieved by both leaving the individual data files open in between several reads, and the linear organization of data in the random-access files, in the case of scrolling through different masses of the dataset.

The 3D visualization of the localization of ions is performed by means of volume extraction, where an ion intensity threshold level determines the presence of the voxels in the volume, as shown in Fig. 4B. The following visualization pipeline in the Visualization Toolkit (VTK) software was used to perform this volume extraction, with the internal VTK functions used mentioned between brackets). Data for the user-selected mass ranges is loaded into an equally spaced 3D data grid (vtkImageData) with optional spatial binning of the voxels of the data grid (vtkImageShrink3D). Low-intensity voxels in the entire datagrid can be filtered out by a user-selected intensity threshold level (vtkImageThreshold) and from the remaining voxels a volume extraction is created (vtkContourFilter), smoothed (vtkSmoothPolyDataFilter and vtkPolyDataNormals) and displayed (vtkPolyDataMapper). The parameters of the individual steps in this pipeline can be set in the VE, and are stored in the 3D configuration file to be available every time the dataset is loaded into the VE software. Three different mass windows are visualized in red, green and blue to be able to visualize multiple m/z ranges at once, as shown in Fig. 4A.

In the ideal situation, ion intensities will appear inside the tissue sample only, which results in a clear view of the 3D distribution of the mass of interest. In practice ion intensities also appear outside the tissue sample of interest as a result of chemical noise in the data to be visualized in 3D. High amounts of chemical noise can obstruct a clear view of the 3D molecular distribution (Fig. 4C). A noise suppression method was developed and implemented in the VE to remove these noise artifacts, by using the region-of-interest (ROI) information defined in the DCE. The ROIs defined per datacube are used as the area of the sample to suppress ion intensities in the VE volume extraction (Fig. 4B and C). The ROIs for this noise suppression method can be manually defined in the DCE, or externally generated and imported into the DCE when automatic background discovery is used, as described in reference [18].

3.7. Software development method

The software described in this article was developed by means of an iterative and incremental method, were functionality was added step by step over several years, more generally known as agile software development [19]. Scientific research environments are typical environments where software functionality evolves over time, and where new insights will develop in during research. Agile software development allows this iterative and incremental way of development.

A common pitfall in developing software is that in the course of expanding and adjusting software it becomes less and less maintainable, and adding new functionality will cause existing functionality to fail. Software design patterns [20] were used for the development over time to minimize this problem here, using the Open-Closed principle. This principle states that software should be open for additions, but closed for changes. Particularly the bridge pattern [21] for decoupling the handling of different dataset formats from different visualization functionalities was very efficient for implementing the different dataset formats in the DCE (imzML, BioMap and the different datacube formats).

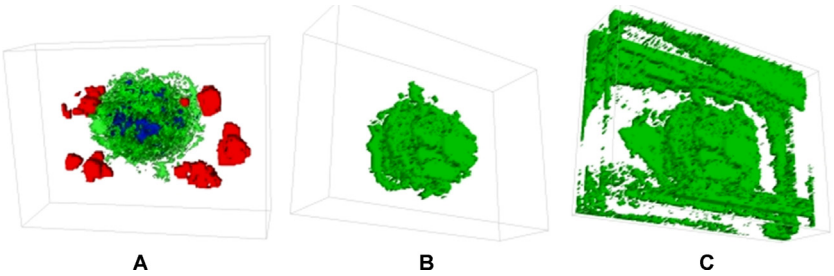

Fig. 4. Volume Explorer 3D Visualization. Visualization results of the Volume Explorer. In A: the volume extraction of three different masses of interest are shown in red (m/z 262.0), green (m/z 560.3) and blue (m/z 760.5). (B and C) shows the effect of the background filtering methods for m/z 560.3, where in B is the filtered visualization of the unfiltered visualization in C. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Results and discussion

4.1. 2D-visualization with the Datacube Explorer

The Datacube Explorer (DCE) enables a uniform way of full resolution visualization and data exploration of large MSI datasets. Images of any mass range of interest can be displayed, ranging from the minimum mass range of a dataset up to the total-ion-image of the entire mass range. Images can be displayed with the intensity scaling range set from the minimum to the maximum ion intensity for every image individually, or any other user definable scaling range, which makes absolute mutual ion intensity comparison between images of different masses possible. In addition to the images, the mass spectrum of the dataset as a combined spectrum for all pixels is presented, where both the full mass range can be selected as well as a user-selectable zoomed mass range. Finally, in the case an imzML dataset is loaded, all metadata included in the imzML dataset is hierarchically shown in the user interface of the DCE.

In order to handle MSI datasets with varying mass resolution, navigation through the entire mass range of the dataset is possible by means of different desired m/z step sizes. The visualization methods used in the DCE are able to present new mass selection images almost instantaneously, which is experienced as very convenient by many users. Based on this m/z selection response time, this puts the DCE among the fastest MSI visualization software of today. The initial loading of the dataset presented here (20086 spectra with 6700 m/z samples each) into the DCE takes 9 s (Intel® Core™ i7 CPU 2.8 GHz, 8 GByte, Windows 7 64-bit).

The region-of-interest (ROI) visualization methods are demonstrated in Fig. 3, where for three selected ROIs the corresponding spectra are shown in red, green and blue, on top of the combined spectrum for all pixels (in black). The ROIs selection is shown as an inset in the upper left of Fig. 3. On top of this image of molecular ion at m/z 706.7 shown in grayscale, an area outside of the gelatin block was selected as the first ROI (in red), the tissue rim was selected as the second ROI (in green), and the inner part of the tissue was marked as the third ROI (in blue). The ROI spectra show here that in the lower mass region (below m/z 700) MALDI matrix-related ions are abundant, while in the mass above m/z 700 tissue-related lipid signals are mostly present. This provides a rapid comparison between the spectral signals associated with different regions of the sample.

The results of the unsupervised clustering algorithm of the self-organizing map functionality are shown in Fig. 5 where six typical distributions from a 7 by 7 Kohonen layer configuration are shown. Tissue center-related distributions are presented in A and B, tissue rim and markers are shown in C and D, E presents the image of the glass slide background, while F highlights markers and tissue rim.

4.2. 3D-visualization with the Volume Explorer

Fig. 4 shows the results of the methods used for the visualization of 3D MSI datasets, implemented by the Volume Explorer. Fig. 4A shows the 3D distribution of three selected masses: in red the fiducial markers at m/z 262.0, in green a lysophospholipid ion (LPC) 18:1/0:0 at m/z 560.3, and in blue phosphatidylcholine (PC) 16:0/18:1 at m/z 760.5. Blue voxels have been set to less opacity

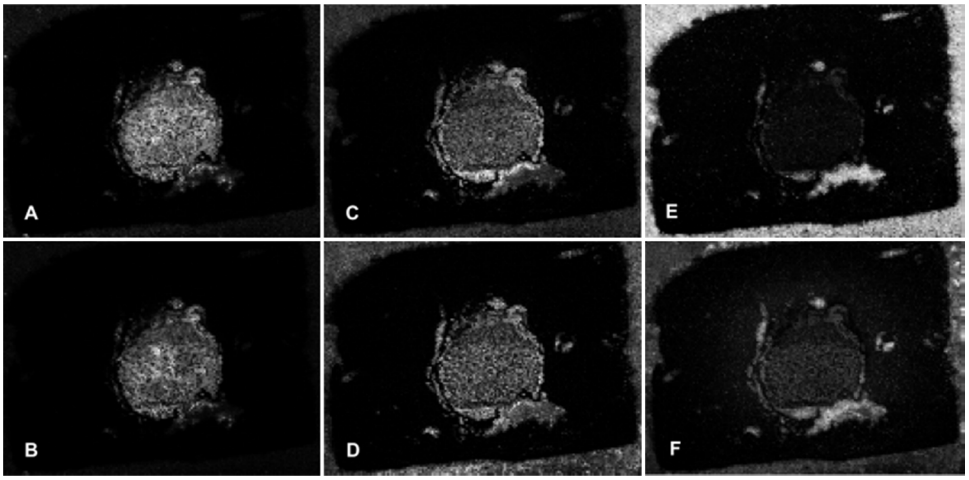

Fig. 5. Self-organizing map analysis. Results of the unsupervised clustering algorithm of the self-organizing map: six typical distributions from a 7 by 7 Kohonen layer configuration, with tissue center-related distributions in A and B, tissue rim and markers in C and D, the image of the glass slide background in E, and markers and tissue rim in F.

than the green voxels, so at positions where both co-localize a green voxel is presented or displayed.

Fig. 4B and C shows the effect of the background filtering method, here for mass m/z 560.3. Without the use of this imaging background filtering method, the unwanted imaged area blocks the view of areas of interest as shown in Fig. 4C. The filtering method applied in Fig. 4B hides the non-tissue area (a glass slide and gelatin block), defined as a region-of-interest in the Datacube Explorer for every slice of the 3D dataset, and thus aids visualization of the voxels of interest.

Results of the Volume Explorer were already used in different publications [3,22], and 3D datasets up to 41 slices [22] have been successfully handled by the Volume Explorer.

4.3. 2D and 3D integration

The results presented in the previous sections show the integration between the 2D visualization by the Datacube Explorer (DCE) and the 3D visualization by the Volume Explorer (VE). This integration enables a scientist to handle both the 2D and 3D data in a uniform way without additional programming, and thus make the DCE and the VE together an integrated visualization platform for MSI.

4.4. Mosaic data structure

The mosaic datacube composition is unique as it enables the visualization of extremely large MSI datasets (e.g., those acquired with FTICR-MSI) at the highest available (experimentally determined) resolution in both the spatial and the spectral dimension. In contrast, memory limitations typically result in the need to visualize such datasets at a lower resolution then that achieved during the experiment. Although the mosaic data structure separates the data into several parts, the division into these individual parts does not affect or limit the visualization of parts of the dataset at full resolution in the DCE. Splitting-up a dataset into separate smaller datacubes also opens the possibility of utilizing the power of parallel processing for generating datacubes out of the raw measurement data, which largely reduces the data processing time and partly eliminates computer memory problems.

Although this transformation is very data intensive, it only has to be performed once. After this data transformation, the optimized data structure is present every time the data set is visualized.

The initial loading time of a mosaic dataset into the DCE is proportional to the loading time mentioned in Section 4.1 and the number of datacubes that the volume-of-interest selection comprises (see also Fig. 2(c)). This is also true for new volume-of-interest selections. The speed of selecting different m/z values of a loaded volume-of-interest is also almost instantaneously, like with a single datacube.

4.5. Datacube versus imzML

Many factors determine what the best data format to store MSI data is, which is dependent on many circumstances. The two main data formats described in this article, the datacube and imzML, are near the extreme edges in terms of simplicity and completeness.

On one side, the datacube data format is extremely simple and straightforward. Only basic metadata is provided and the data itself is stored sequentially (however performance-wise it is optimized for MSI data retrieval). The plain and simple open datacube data format has enabled many successful scientific collaborations regarding the reuse of MSI data. Because of its relative simplicity, reading in the datacube data is straightforward for software programmers, and generating datacube datasets that can be visualized by the DCE can be performed instantly. These custom data conversions have been performed with simple software scripts written in for example Matlab (MathWorks, Natick, United States) or LINQPad (Joseph Albahari, North Beach, Australia). The downside of this simplicity is that the metadata of the dataset should be stored separately elsewhere.

On the other side, the self-describing, fully specified, and open data format imzML supports the global exchange of MSI data. The use of a MSI domain specific controlled vocabulary [7] makes the content of this type of MSI data set unambiguous, and the XML-based structure can contain different data storage configurations. The imzML data structure is optimized for performance and scalability for MSI data, which is not generally the case for XML-based formats [23]. The downside of this sophisticated data format is the considerable amount of effort that is required to read or generate data.

5. Conclusion

This article describes new methods for data exploration and visualization for large 2D and 3D mass spectrometry imaging (MSI) datasets where both the 2D and 3D visualization is integrated which creates a flexible software platform for MSI experimenters. We focused on how to efficiently handle large data sizes and still maintain the full resolution, both spatially and spectrally, of the data.

The Datacube Explorer (DCE) application is one of the many applications capable of visualizing MSI data, but is currently one of the few applications able to visualize imzML datasets without additional required programming. It is unique in its ability to visualize (1) BioMap datasets, (2) imzML datasets including the underlying metadata, and (3) mosaic datacubes that enables the visualization of virtually any size or resolution and can be split up in manageable chunks of data on disk. This renders the DCE a generic, publicly available, light-weight software application for fast image-based visualization of large imaging mass spectrometry datasets. The DCE offers a generic uniform visualization method for MSI data produced by different instruments and different detection techniques, which makes MSI data visualization vendor independent. Although the DCE provides a generic way of visualizing the data, the underlying data representation remains flexible. This flexibility enables the integration of data from different instruments and measurement techniques, despite the fact that the data has different formats and ranges. The DCE is full-functioning software that enables scientists to visualize MSI data without any software programming effort, has already proven its value in many scientific publications [2,16,24–34], and is be a good addition to the many other valuable and important software applications in the field of MSI. When imzML is available as a data format, the DCE is an often selected visualization tool [35].

The Volume Explorer (VE) implements the visualization of 3D measurements based on the datacube data format and closely integrated with the DCE. The retrieval of mass resolved data out of the enormous data sets is optimized for 3D MSI visualization, resulting in a fast visualization of user-selectable mass-ranges.

The methods and tools described here support the MSI community to move forward with the exploration and visualization of their MSI data, and have already been applied in multiple experiments.

The Datacube Explorer software can be freely downloaded from http://www.amolf.nl/download/datacubeexplorer and can be used after a free license has been obtained. The Volume Explorer software can be obtained upon special request by contacting AMOLF via msi-software@amolf.nl.

Acknowledgements

This work is part of the research program of the "Stichting voor Fundamenteel Onderzoek der Materie (FOM)," which is financially supported by the "Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO)". This publication was also supported by the Dutch national program COMMIT. The work on the implementation of imzML is part of the Computis Program, 6th European Framework Program for Research and Technological Development (FP6), project no. LSHGCT-2005-5181194. We gratefully acknowledge financial support from NIH grant R01 CA134695 to obtain the mass spectrometry data sets used in this article.

We would specially like to thank Thorsten Schramm, Alfons Hester, and Andreas Roempp from Justus Liebig University in Giessen for their effective collaboration on the development on imzML which helped in the development of the DCE. We would like to thank Marco Konijnenburg and Marco Seynen from our Software Engineering department for their software development support, and current and past members of our Biomolecular Imaging Mass Spectrometry group, especially Don Smith, for their continuous constructive feedback on the MSI software presented here.

References

[1] L.A. McDonnell, R.M.A. Heeren, Imaging mass spectrometry, Mass Spectrometry Reviews 26 (4) (2007) 606–643.
[2] E.R.A. van Hove, D.F. Smith, R.M.A. Heeren, A concise review of mass spectrometry imaging, Journal of Chromatography A 1217 (25) (2010) 3946–3954.
[3] K. Chughtai, L. Jiang, T.R. Greenwood, I. Klinkert, E.R.A. van Hove, R.M.A. Heeren, K. Glunde, Fiducial markers for combined 3-dimensional mass spectrometric and optical tissue imaging, Analytical Chemistry 84 (4) (2012) 1817–1823.
[4] R.M. Parry, A.S. Galhena, C.M. Gamage, R.V. Bennett, M.D. Wang, F.M. Fernandez, OmniSpect: an open MATLAB-based tool for visualization and analysis of matrix-assisted laser desorption/ionization and desorption electrospray ionization mass spectrometry images, Journal of the American Society for Mass Spectrometry 24 (4) (2013) 646–649.
[5] G. Robichaud, K.P. Garrard, J.A. Barry, D.C. Muddiman, MSiReader: an open-source interface to view and analyze high resolving power MS imaging files on Matlab platform, Journal of the American Society for Mass Spectrometry 24 (5) (2013) 718–721.
[6] C. Paschke, A. Leisner, A. Hester, K. Maass, S. Guenther, W. Bouschen, B. Spengler, Mirion—a software package for automatic processing of mass spectrometric images, Journal of the American Society for Mass Spectrometry 24 (8) (2013) 1296–1306.
[7] MALDI-MSI Interest Group, imzML – A Common Data Format for MS Imaging, 2011, Available from: http://www.imzml.org
[8] M. Hamacher, M. Eisenacher, C. Stephan, A. Römpp, T. Schramm, A. Hester, I. Klinkert, J.-P. Both, R.M.A. Heeren, M. Stöckli, B. Spengler, imzML: imaging mass spectrometry markup language: a common data format for mass spectrometry imaging, in: J.M. Walker (Ed.), Data Mining in Proteomics, Humana Press, New York, 2011, pp. 205–224.
[9] HUPO Proteomics Standards Initiative. mzML 1.1.0 Specification. Available from: http://www.psidev.info/index.php?q=node/257
[10] X. Xiong, W. Xu, L.S. Eberlin, J.M. Wiseman, X. Fang, Y. Jiang, Z. Huang, Y. Zhang, R.G. Cooks, Z. Ouyang, Data processing for 3D mass spectrometry imaging, Journal of the American Society for Mass Spectrometry 23 (2012) 1147–1156.
[11] V. Raman, D. Artemov, A.P. Pathak, P.T. Winnard Jr., S. McNutt, A. Yudina, A. Bogdanov Jr., Z.M. Bhujwalla, Characterizing vascular parameters in hypoxic regions: a combined magnetic resonance and optical imaging study of a human prostate cancer model, Cancer Research 66 (20) (2006) 9929–9936.
[12] K. Glunde, T. Shah, P.T. Winnard Jr., V. Raman, T. Takagi, F. Vesuna, D. Artemov, Z.M. Bhujwalla, Hypoxia regulates choline kinase expression through hypoxia-inducible factor-1 alpha signaling in a human prostate cancer model, Cancer Research 68 (1) (2008) 172–180.
[13] Kitware Inc. The Visualization Toolkit (VTK). Available from: http://www.vtk.org
[14] FOM Institute AMOLF. FOM Institute AMOLF Download Area. Available from: http://www.amolf.nl/download/datacubeexplorer/
[15] W.D. Pence, L. Chiappetti, C.G. Page, R.A. Shaw, E. Stobie, Definition of the flexible image transport system (FITS), version 3.0, Astronomy and Astrophysics 524 (2010) A42.
[16] D.F. Smith, A. Kharchenko, M. Konijnenburg, I. Klinkert, L. Pasa-Tolic, R.M.A. Heeren, Advanced mass calibration and visualization for FT-ICR mass spectrometry imaging, Journal of the American Society for Mass Spectrometry 23 (11) (2012) 1865–1872.
[17] T. Kohonen, The self-organizing map, Proceedings of the IEEE 78 (9) (1990) 1464–1480.
[18] G.B. Eijkel, B. Kükrer Kaletaş, I.M. van der Wiel, J.M. Kros, T.M. Luider, R.M.A. Heeren, Correlating MALDI and SIMS imaging mass spectrometric datasets of biological tissue surfaces, Surface and Interface Analysis 41 (8) (2009) 675–768.
[19] T. Dybå, T. Dingsøyr, Empirical studies of agile software development: a systematic review, Information and Software Technology 50 (9/10) (2008) 833–859.
[20] E. Gamma, R. Helm, R. Johnson, J. Vlissides, Design Patterns: Elements of Reusable Object-Oriented Software, Addisson-Wesley, Toronto, Ontario, Canada, 1995.
[21] A. Shalloway, J. Trott, Design Patterns Explained: A New Perspective on Object-Oriented Design (2nd Edition) (Software Patterns Series), Addison-Wesley Professional, Boston, 2004.
[22] L. Fornai, A. Angelini, I. Klinkert, F. Giskes, A. Kiss, G. Eijkel, E.A. Amstalden-van Hove, L.A. Klerk, M. Fedrigo, G. Pieraccini, G. Moneti, M. Valente, G. Thiene, R.M.A. Heeren, Three-dimensional molecular reconstruction of rat heart with mass spectrometry imaging, Analytical and Bioanalytical Chemistry 404 (10) (2012) 2927–2938.
[23] S.M. Lin, L.H. Zhu, A.Q. Winter, M. Sasinowski, W.A. Kibbe, What is mzXML good for? Expert Review of Proteomics 2 (6) (2005) 839–845.
[24] E.R.A. van Hove, D.F. Smith, L. Fornai, K. Glunde, R.M.A. Heeren, An alternative paper based tissue washing method for mass spectrometry imaging: localized washing and fragile tissue analysis, Journal of the American Society for Mass Spectrometry 22 (10) (2011) 1885–1890.
[25] K. Aizikov, D.F. Smith, D.A. Chargin, S. Ivanov, T.Y. Lin, R.M.A. Heeren, P.B. O'Connor, Vacuum compatible sample positioning device for matrix assisted laser desorption/ionization Fourier transform ion cyclotron resonance mass spectrometry imaging, Review of Scientific Instruments 82 (5) (2011) 054102 1–054102 8.
[26] D.F. Smith, K. Aizikov, M.C. Duursma, F. Giskes, D.J. Spaanderman, L.A. McDonnell, P.B. O'Connor, R.M.A. Heeren, An external matrix-assisted laser desorption ionization source for flexible FT-ICR mass spectrometry imaging with internal calibration on adjacent samples, Journal of the American Society for Mass Spectrometry 22 (1) (2011) 130–137.
[27] E.R.A. van Hove, T.R. Blackwell, I. Klinkert, G.B. Eijkel, R.M.A. Heeren, K. Glunde, Multimodal mass spectrometric imaging of small molecules reveals distinct spatio-molecular signatures in differentially metastatic breast tumor models, Cancer Research 70 (22) (2010) 9012–9021.
[28] K. Chughtai, R.M.A. Heeren, Mass spectrometric imaging for biomedical tissue analysis, Chemical Reviews 110 (5) (2010) 3237–3277.
[29] R.M.A. Heeren, D.F. Smith, J. Stauber, B. Kukrer-Kaletas, L. MacAleese, Imaging mass spectrometry: hype or hope? Journal of the American Society for Mass Spectrometry 20 (6) (2009) 1006–1014.
[30] I.M. Taban, A.F.M. Altelaar, J. Fuchser, Y.E.M. van der Burgt, L.A. McDonnell, G. Baykut, R.M.A. Heeren, Imaging of peptides in the rat brain using MALDI-FTICR mass spectrometry, Journal of the American Society for Mass Spectrometry 18 (1) (2007) 145–151.
[31] A.F. Maarten Altelaar, S.L. Luxembourg, L.A. McDonnell, S.R. Piersma, R.M.A. Heeren, Imaging mass spectrometry at cellular length scales, Nature Protocols 2 (5) (2007) 1185–1196.
[32] L. MacAleese, M.C. Duursma, L.A. Klerk, G. Fisher, R.M.A. Heeren, Protein identification with liquid chromatography and matrix enhanced secondary ion mass spectrometry (LC-ME-SIMS), Journal of Proteomics 74 (7) (2011) 993–1001.
[33] D.F. Smith, A. Kiss, F.E. Leach III, E.W. Robinson, L. Paša-Tolić, R.M.A. Heeren, High mass accuracy and high mass resolving power FT-ICR secondary ion mass spectrometry for biological tissue imaging, Analytical and Bioanalytical Chemistry 405 (18) (2013) 6069–6076.
[34] D.F. Smith, E.W. Robinson, A.V. Tolmachev, R.M.A. Heeren, L. Paša-Tolić, C-60 secondary ion Fourier transform ion cyclotron resonance mass spectrometry, Analytical Chemistry 83 (24) (2011) 9552–9556.
[35] J. Thunig, S.H. Hansen, C. Janfelt, Analysis of secondary plant metabolites by indirect desorption electrospray ionization imaging mass spectrometry, Analytical Chemistry 83 (9) (2011) 3256–3259.