(19) DANMARK

(10) DK/EP 3092589 T3

(12) Oversættelse af europæisk patentskrift

Patent- og Varemærkestyrelsen

(51) Int.Cl.: *G 16 B 45/00 (2019.01)* *G 06 T 7/33 (2017.01)* *G 06 T 7/90 (2017.01)*
G 16 B 50/30 (2019.01)

(45) Oversættelsen bekendtgjort den: **2024-02-05**

(80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2023-11-08**

(86) Europæisk ansøgning nr.: **15701558.7**

(86) Europæisk indleveringsdag: **2015-01-09**

(87) Den europæiske ansøgnings publiceringsdag: **2016-11-16**

(86) International ansøgning nr.: **FR2015050051**

(87) Internationalt publikationsnr.: **WO2015104512**

(30) Prioritet: **2014-01-10 FR 1450176**

(84) Designerede stater: **AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR**

(73) Patenthaver: **Aliri France, 152, rue du Docteur Yersin, 59120 Loos, Frankrig**

(72) Opfinder: **PAMELARD, Fabien, 33 Rue Béranger, Appt. 7, F-59280 Armantieres, Frankrig**
STAUBER, Jonathan M., 16 Rue du Maréchal Leclerc, F-59320 Haubourdin, Frankrig

(74) Fuldmægtig i Danmark: **Plougmann Vingtoft A/S, Strandvejen 70, 2900 Hellerup, Danmark**

(54) Benævnelse: **FREMGANGSMÅDE TIL BEHANDLING AF MOLEKYLÆRE BILLEDDANNELSESDATA SAMT TILSVARENDE DATASERVER**

(56) Fremdragne publikationer:
WO-A1-2009/156897
IVO KLINKERT ET AL: "Methods for full resolution data exploration and visualization for large 2D and 3D mass spectrometry imaging datasets", INTERNATIONAL JOURNAL OF MASS SPECTROMETRY, vol. 362, 24 December 2013 (2013-12-24), pages 40-47, XP055477103, NL ISSN: 1387-3806, DOI: 10.1016/j.ijms.2013.12.012
Suits Frank ET AL: "Correlation Queries for Mass Spectrometry Imaging", Analytical Chemistry, vol. 85, no. 9, 7 May 2013 (2013-05-07), pages 4398-4404, XP055829220, US ISSN: 0003-2700, DOI: 10.1021/ac303658t Retrieved from the Internet: URL:https://pubs.acs.org/doi/pdf/10.1021/a c303658t>
RÜBEL OLIVER ET AL: "OpenMSI: A High-Performance Web-Based Platform for Mass Spectrometry Imaging", ANALYTICAL CHEMISTRY, vol. 85, no. 21, 5 November 2013 (2013-11-05), pages 10354-10361, XP055829223, US ISSN: 0003-2700, DOI: 10.1021/ac402540a Retrieved from the Internet: URL:https://pubs.acs.org/doi/pdf/10.1021/a c402540a>
NORRIS ET AL: "Processing MALDI mass spectra to improve mass spectral direct tissue analysis", INTERNATIONAL JOURNAL OF MASS SPECTROMETRY, ELSEVIER SCIENCE PUBLISHERS , AMSTERDAM, NL, vol. 260, no. 2-3, 4 January 2007 (2007-01-04), pages 212-221, XP005827570, ISSN: 1387-3806

Description

TECHNICAL FIELD OF THE INVENTION

The present invention relates to a method for processing molecular imaging data based on the use of a high-volume database structure, as well as the corresponding data server. The molecular imaging generally encompasses all imaging techniques essentially making it possible to observe the molecular functioning of organs and/or organisms ex vivo or in vivo by means that are as least invasive as possible or disrupt the observed biological or biophysical systems as little as possible.

TECHNOLOGICAL BACKGROUND

The molecular imaging is experiencing considerable growth in the evolution of the number of technologies, particularly in mass spectrometry imaging (MSI), magnetic resonance imaging (MRI), or again Raman spectrometry. The use of these technologies makes it possible to carry out biodistribution studies of targeted endogenous or exogenous compounds such as pesticides, drugs, proteins or lipids in order to study their roles in biological systems. The applications of these technologies with the help of software also make it possible to search for potential biomarkers in an unsupervised manner.

Nonetheless, the increase in data to be analyzed and interpreted in molecular imaging constitutes a limitation for synthetic analyzes on several datasets each associated with an image. The data then becomes difficult to use for synthetic representations or for statistical analysis.

Indeed, the high volume of information in datasets is often greater than the size of the random access memory of classic computers (as opposed to super computers). In mass spectrometry imaging for example, images of 50,000 positions no longer allow all of the raw information to be stored in memory and therefore require calculations to be carried out to reduce the impact in memory and allow the exploitation of these images. The calculations reduce the data and therefore induce bias by not taking into account all the available information. Additionally, current analysis tools provide no means of normalizing multiple datasets with each other.

There are many storage formats linked mainly to the manufacturers of dataset acquisition machines. Nonetheless, none of these formats is adapted and optimized to be queried in a detailed and rapid manner. Some areas of molecular imaging have developed storage formats such as Analysis 7.5 which are found in MRI (Magnetic Resonance

Imaging) and mass spectrometry imaging. In mass spectrometry imaging, it is also found the iMZML format which, nonetheless, does not constitute a storage system for handling imaging data.

Formats of the HDF5 type (for "Hierarchical Data Format 5") based on a hierarchical organization of data have recently been used in the storage of high volume data in mass spectrometry imaging. These formats make it possible to build remote query interfaces via Internet browsers. They allow faster statistical calculations than with non-hierarchical files. Nonetheless, these formats do not make it possible to obtain a global and searchable comparative analysis of the data in studies of several datasets simultaneously.

There is therefore a need for a method allowing to analyze, compare, and interrogate several datasets in relation to each other without loss of information in order to avoid introducing analysis biases.

AIM OF THE INVENTION

The present invention aims to respond effectively to this need by proposing a method for computer processing molecular imaging data as defined in claim 1.

More generally, the invention proposes to process datasets, advantageously acquired in an automated manner, by cutting into sections the molecular spectra associated with each of the spatial positions of the sets of spatial positions composing said datasets. Each spectrum thus corresponds to a succession of spectrum sections, together comprising the set of the molecular information of the associated spectrum. According to the invention, the step of cutting into sections is carried out in an automated manner, during the acquisition of the spectral data or during the analysis of the spectral data.

In a first embodiment, the successive sections do not overlap and are simply adjacent, so that no molecular information of a considered spectrum is lost or redundant. In another embodiment, some of the sections of a considered spectrum may partially overlap, so that certain molecular information can be contained in two successive spectrum sections of the same spectrum.

The invention thus makes it possible to substantially reduce the loading time of the data to be analyzed, insofar as it is possible to recover, by selection of sections, the useful information in the database. It is also possible to section the spectral data as soon as the data is recorded, during their acquisition, so that the sections of spectral data are integrated directly during the analysis of the sample by a molecular imaging method.

Due to the division into sections of the signal, without reduction of the spectrum, the invention also makes it possible to work without reduction of information in the context of combined studies of pharmacokinetics, pharmacodynamics and quantification of endogenous or exogenous molecules in a same analysis in order to observe physiological events as a whole such as screening methods (comparison of distribution of several drug candidates), quantification methods, methods for studying receptor occupancy or tissue screening, metabolism study or 3D reconstruction. The invention also makes it possible to optimize the use of memory necessary for statistical processing and representations of molecular imaging data on several datasets.

According to one implementation, the data processing or recording method of the invention includes the step of visualizing the molecular information of interest in the form of a set of data maps each corresponding to a dataset.

Advantageously, the molecular spectra are cut according to a determined pitch, which may be identical or different from one spectrum to another.

According to one implementation, the data processing or recording method of the invention includes the following steps:

- cutting the set of molecular information into spectrum sections following at least one predetermined pitch,
- inserting a reference axis establishing a correspondence between point indexes of the different spectrum sections and the corresponding molecular information, and
- selecting the spectrum section containing the molecular information of interest according to the reference axis and the predetermined pitch.

According to one implementation, before or after the insertion of the sections into the database, said method includes at least one pre-processing step consisting of a spectral alignment and/or background noise subtraction and/or intra-dataset normalization.

According to one implementation, the data processing or recording method of the invention includes the step of resizing the spatial positions of the datasets having different spatial sizes so as to be aligned with the dataset having the finest spatial size.

According to one implementation, the data processing or recording method of the invention further includes the step of extracting a set of data, such as a maximum

intensity, an average intensity, an area under a peak, a signal-to-noise ratio of each peak of interest in the set of molecular information.

According to one implementation, the extraction is carried out according to peak selection criteria defined by a quality criterion of a signal-to-noise ratio and/or a spectral resolution.

According to one implementation, the data processing or recording method of the invention includes the step of normalizing the sets of molecular information of the plurality of datasets with each other so as to be able to compare the different datasets with each other.

According to one implementation, to carry out the standardization step, said method may include the following steps:

- selecting a molecule or several endogenous or exogenous reference molecules common to all the datasets,
- calculating, for each dataset, a normalization factor according to this or these reference molecules, and
- correcting the molecular information of interest with this normalization factor to obtain a set of normalized datasets.

According to one implementation, the normalization factor is defined from at least one reference molecule present on a set of samples corresponding to the datasets or in a particular reference zone of the samples corresponding to the datasets.

According to one implementation, the data processing or recording method of the invention includes the step of visualizing, comparing, and analyzing, in the form of a plurality of data maps each coming from a standardized dataset, the molecular information of interest with a color scale common to all maps.

In one implementation, the common color scale is based on a smallest and largest intensity of all datasets for the molecule of interest.

According to one implementation, said datasets are obtained by a mass spectrometry method.

According to one implementation, said datasets are obtained by an Imaging method such as Positron Emission Tomography (PET) or Magnetic Resonance Imaging (MRI).

The invention also relates to a data server as defined in claim 14.

BRIEF DESCRIPTION OF THE FIGURES

The invention will be better understood on reading the following description and examining the accompanying figures. These figures are given only for illustrative purposes but in no way limitative of the invention.

Figure 1 schematically represents the different steps of a mass spectroscopy method for obtaining datasets used by the method according to the present invention;

Figure 2 is a block diagram of the integrated database management system allowing the implementation of the data processing method according to the present invention;

Figure 3 is a schematic representation of the step making it possible to obtain the sections of the spectra associated with each position of a dataset;

Figure 4 represents the steps making it possible to establish the correspondence between a molecular information and the corresponding spectrum section;

Figure 5 schematically shows the inter-image normalization steps making it possible to precisely compare several images with each other.

The identical, similar, or analogous elements retain the same references from one figure to another.

DESCRIPTION OF AN IMPLEMENTATION EXAMPLE OF THE INVENTION

The implementation example of the method according to the invention which follows is described for a MALDI type mass spectrometry method. Of course, another imaging device than MALDI, such as for example the sources: SIMS, DESI, LAESI, DIOS, ICP, MALDI Microscope, SNOM, SMALDI, LA-ICP, ESI (extraction liquid on fabric), MILDI, JEDI, ELDI etc. could be used in a substantially identical manner. The method is also applicable to any other method making it possible to produce data usable by a molecular imaging method such as for example PET ("Positron Emission Tomography") or MRI (Magnetic Resonance Imaging).

Figure 1 schematically shows the different steps implemented by a MALDI mass spectrometry method. A first step El consists of cutting, generally by cryo-section, a slice of tissue 1 which is then placed on a blade 2. It is then deposited in a step E2 a thin uniform layer of ionizing matrix 3 on the slice of tissue 1. The acquisition step E3 consists of generating automated laser pulses 4 having a predefined size, for example in the range of 100 micrometers per acquisition zone, in order to ionize the molecules of the matrix 3. The molecules thus ionized are analyzed in a manner known per se by a mass spectrometer. We can refer to document FR2973112 for more details on the implementation of the method.

It is then obtained a dataset Jk defined by a set of spatial positions (Xi, Yj) to each of which a set of molecular information is associated. The spatial positions are defined according to a reference frame of X and Y axes. The set of molecular information is formed in this case in particular by a two-dimensional spectrum S indicating an intensity as a function of a molecular mass. For example, the obtained dataset Jk presents 50,000 intensities (namely 50,000 points) per spectrum over 20,000 spatial positions Xi, Yj. The method according to the invention is based on the processing of a plurality of datasets J1-Jn.

As shown in Figure 3, for each dataset Jk capable of being represented in the form of an image, the molecular spectrum S(Xi, Yj) associated with each position (Xi, Yj) is divided into several spectra sections T1-Tm. The sections T1-Tm are cut following a predetermined pitch P'. For example, for each spectrum of 50,000 points per spatial position, each spectrum could be cut according to a pitch P' of 10,000 points. In other words, the spectrum S(Xi, Yj) is divided into five sections T1-T5 of 10,000 points each. Alternatively, the used pitch P' may be variable.

It is also defined a reference axis Aref, shown in Figure 4, establishing a correspondence between point indexes of the different sections T1-Tm and the corresponding molecular mass.

Preferably, before insertion into a BDD database, the method implements at least one step of pre-processing the spectra S (Xi, Yj) via a module 100. This pre-processing step may consist of spectral alignment. For this purpose, reference peaks are considered and the spectrum is shifted so as to position the corresponding peak values on these reference peaks. It will also be possible to carry out a subtraction of the background noise by realigning the spectrum according to a difference between the minimum value of the spectrum and a reference value to which the minimum value of the spectrum should correspond.

Furthermore, different types of signal normalization (internal normalization) can be carried out (by an internal standard, by using the total intensities, etc.). In other words, an intra-dataset normalization is carried out here, that is to say a normalization for each dataset independently of the other datasets, which corresponds to steps known per se to those skilled in the art. This intra-dataset normalization is carried out in addition to the subsequent normalization steps between the datasets J1-Jn detailed below (inter-dataset

normalization) which will make it possible to compare, precisely and reliably, the datasets J1-Jn with each other.

Moreover, the method includes the step of extracting a set of data, such as a maximum intensity Pk1-Pkl, an average intensity, an area under a peak, a signal-to-noise ratio of each peak of interest of the spectra S (Xi, Yj). It is thus possible to extract the maximum intensity of each peak Pk1-Pkl from the spectrum S(Xi, Yj) following a method called "PeakPicking". The extraction is carried out according to peak selection criteria defined by a quality criterion of the signal-to-noise ratio and/or the spectral resolution.

The obtained sections T1-Tm for each position (Xi, Yj) of each dataset J1-Jn as well as the reference axis Aref and the peaks Pk1-PKl previously extracted are inserted into the database BDD as shown in Figure 2. Thus, at each position (Xi, Yj) of each dataset J1-Jn are associated a set of indexed sections T1-TM and a set of peaks Pk1-PKl and the aforementioned complementary data (average intensities, areas of each peak).

The database BDD is a high-volume database of the NoSQL (for "Not only Structured Query Language") or SQL type allowing the visualization, management, interpretation and statistical analysis of the inserted data. The database management system can for example be selected from the following non-exhaustive systems: Hypertable, Haddoop, Cassandra, MongoDB, PolyBase, Hadoop on Azure, Hive, Pig. The database system ensures the consistency, reliability, and relevance of the contained data. Replication and backup techniques can guarantee the reliability of the information.

The steps allowing the selection of relevant information in the database BDD are described below. Following a request relating to a molecular information of interest, the section(s) of the positions of the sets of datasets J1-Jn containing the requested molecular information are selected. To this end, for each position Xi, Yj of each dataset J1-Jn, the section Tp containing the requested molecular information is selected according to the reference axis Aref and the requested molecular information.

Thus, in Figure 4, it is illustrated a section of a spectrum following a pitch P' of five points to facilitate the understanding of this step. As is often the case for spectra from a mass spectrometry method, the scale of molecular masses is not linear. In the case where the user requires information relating to a molecule having for example a molecular mass of a value 3 m/z, the reference axis Aref makes it possible to establish that the index of the point corresponding to this molecular mass value is 9, so that it can be deduced using

the pitch P' that the section Tp containing the intensity of the molecular mass of interest is the second section T2.

Alternatively, it can be used a table which stores the minimum and maximum limits of the molecular masses and the indexing of the points of each corresponding section T1-Tm. In this case, for a requested molecular information, the identifier of the section T1-Tm is selected according to these minimum and maximum limits.

Within each selected section T1-Tm for each of the points of all the datasets J1-Jn, the corresponding intensity of the requested molecular mass is then selected. It then becomes possible to visualize, in the form of a set of data maps C1-CN each corresponding to a dataset, the distribution of the intensity of presence of the molecule of interest on the different maps C1-CN. A color is assigned based on the intensity of the molecular information on a color scale. Alternatively, or in a complementary manner, an extraction of the peaks Pk1-PKl makes it possible to quickly produce images providing an initial analysis of the distribution of the intensity of a molecule of interest.

Preferably, the method includes the step of normalizing the molecular information of the datasets J1-Jn in order to be able to compare several datasets with each other (see module 101). For this purpose, a reference molecule Mref common to all the datasets J1-Jn is selected. Alternatively, several endogenous or exogenous molecules common to all the datasets J1-Jn is selected. It is possible to take into account the presence of the reference molecule(s) Mref in the complete samples or more specifically in a reference zone of the corresponding samples.

More precisely, for each dataset J1-Jn, a normalization factor Fnor is calculated based on this or these reference molecules Mref. The molecular information of interest is then corrected with this normalization factor Fnor in order to obtain a set of normalized datasets J1-Jn.

For example, for three datasets corresponding to the maps C1-C3 shown in Figure 5, a reference molecule Mref=*500 m/z* as a normalizing molecule (normalizing standard) is selected. This molecule Mref has respectively a value of 5, 10 and 15 in the zone Z of the maps C1-C3. For each map C1-C3, a normalization factor Fnor on the basis of each of these values of Mref is calculated:

Fnor= Average of the values Mref for C1-C3 / Mref value of the considered map. In the example shown in Figure 5, a normalization factor Fnor of 2, 1, and 10/15 respectively for the maps C1, C2 and C3 is obtained. This coefficient is then applied to all the intensity

values of the spectra associated with the different positions of all the datasets J1-Jn and in particular to the intensity values of a molecule of interest Mi=*300 m /z* at the position Xi; Yj. Normalized maps C1'-C3', which can be directly compared with each other are then obtained.

The normalization thus makes it possible to correct the intensities linked to changes in experimental conditions due to the preparation of the sample and its analysis. Indeed, in the previous example, it is noted that it would have been possible to arrive at an erroneous conclusion of a stronger presence of the molecule of interest Mi=*300 m/z* in the sample corresponding to the map C3, while after application of the normalization factor Fnor, the study reveals a homogeneous presence of the molecule of interest Mi in the three samples with an intensity of 200 at the position Xi; Yj.

Following this normalization, it is then possible to visualize, compare and analyze, in the form of standardized data maps C1'-C3', the molecular information with a color scale common to all maps C1'-C3' (see module 102). The obtained scale is based on the smallest and largest intensity determined on all datasets for the molecule of interest Mi.

Of course, it is possible to produce more than three datasets and corresponding maps to analyze by statistical grouping several datasets J1-Jn in a transverse manner without needing to load them all into memory. The imaging data can then be compared to identify potential markers. It is also possible to carry out position groupings with common behavior in order to allow a new interpretation of the data.

In other words, the invention thus makes it possible, in a single tool, to compare and query the data to obtain more relevant, standardized information whose quality is quantifiable. Several datasets can be loaded with quality controls inserted into each of the images, thus allowing the data to be normalized with each other.

Furthermore, the method may, if necessary, make it possible to resize the positions of datasets J1 and J2 of the same biological sample having different spatial sizes so as to be aligned with the dataset having the finest spatial size. It is then possible to combine the processing of a map from the mass spectrometer (MALDI imaging) with the processing of a map from a PET (Positron Emission Tomography) or MRI (Magnetic Resonance Imaging) type system or MRI (Magnetic Resonance Imaging).

It is also possible to represent several co-distributed datasets J1-Jn in a single image. It is then possible to carry out queries of the type "select all the values v, w of all

the positions x and y (of the same spatial size) for the field z and a of the current map". The fields z and a represent the values of two different imaging technologies.

It is also possible to carry out temporary statistical processing which can be added by insertion requests into the database BDD. Thus, the method according to the invention provides great flexibility to the extent that it is possible to add information on each position of the image at any time during the analysis method.

The method is based on a client/server architecture. The database BDD can be searched via a single interface. The data is stored and queried in a single-station architecture or in a cluster of servers accessible locally or via a network (internal or external). The client software interface can be installed on the workstation or in web interface version. The invention also relates to the data server 10 including a memory storing software instructions for the implementation of at least part of the steps of the method for processing a plurality of datasets J1-Jn.

In addition, the quantity of processed data can grow easily, so that the volume of data to be analyzed is no longer a limitation to the analysis. The insertion data is efficient regardless of the size of the database BDD.

Of course, the preceding description has been given by way of example only and does not limit the field of the invention from which one would not escape by replacing the execution details with any other equivalents.

Patentkrav

1. Fremgangsmåde til computerbehandling af molekylære billeddannelsesdata omfattende et stort antal spektraldatasæt (J1-Jn), hvor hvert sæt spektrale data (J1-Jn) er defineret ved en samling rumlige positioner (Xi, Yj), til hvilke der hver især er tilordnet et molekylspektrum med mindst to dimensioner, der indeholder en samling molekylære informationer (S(Xi, Yj)), **kendetegnet ved, at** den omfatter følgende trin:

- for hvert datasæt (J1-Jn), automatisk opdeling af det til hver position (Xi, Yj) tilordnede molekylspektrum i flere spektralafsnit (T1-Tm), der hvert især indeholder en del af spektrets molekylære informationer, hvor spektralafsnittene tilsammen omfatter samlingen af det dertil tilordnede molekylspektrums molekylære informationer,
- indsætning af de for hver position (Xi, Yj) i hvert datasæt (J1-Jn) opnåede afsnit (T1-Tm) i en database (BDD), således at der til hver position (Xi, Yj) i hvert datasæt (J1-Jn) tilordnes en samling indekserede spektralafsnit (T1-Tm),
- udvælgelse i databasen (BDD) efter en anmodning vedrørende en molekylær information af interesse af det eller de afsnit (T1-Tm), der indeholder den molekylære information af interesse, og
- udvælgelse af den molekylære information af interesse inden for hvert afsnit (T1-Tm).

2. Fremgangsmåde ifølge krav 1, **kendetegnet ved, at** den omfatter et trin til visualisering af den molekylære information af interesse i form af en samling datakort (C1-Cn), der hvert især svarer til et datasæt (J1-Jn).

3. Fremgangsmåde ifølge et hvilket som helst af kravene 1 til 2, **kendetegnet ved, at** den omfatter følgende trin:

- opdeling af de molekylære spektre (S(Xi, Yj)) i spektralafsnit (T1-Tm) ifølge mindst et forudbestemt skridt (P'),
- indsætning af en referenceakse (Aref), der etablerer en korrespondance mellem de forskellige afsnits (T1-Tm) punktindices og den korresponderende molekylære information, og

- udvælgelse af det afsnit (T1-Tm), der indeholder den molekylære information af interesse, som funktion af referenceaksen (Aref) og det forudbestemte skridt (P').

4. Fremgangsmåde ifølge et hvilket som helst af kravene 1 til 3, **kendetegnet ved, at** fremgangsmåden før eller efter indsætningen af afsnittene (T1-Tm) i databasen (BDD) omfatter mindst et forbehandlingstrin, der består i en spektral opstilling og/eller en subtraktion af baggrundsstøj og/eller en intra-datasæt-normalisering (J1-Jn).

5. Fremgangsmåde ifølge et hvilket som helst af kravene 1 til 4, **kendetegnet ved, at** den omfatter et trin til at redimensionere datasættenes (J1-Jn) rumlige positioner (Xi, Yj) med forskellige rumlige størrelser for at tilpasses til datasættet (J1-Jn) med den mindste rumlige størrelse.

6. Fremgangsmåde ifølge et hvilket som helst af kravene 1 til 5, **kendetegnet ved, at** den desuden omfatter et trin til udtrækning af en samling data, såsom en maksimal intensitet (Pk1-Pkl), en middelintensitet, et areal under en spids, et signal/støj-forhold for hver spids af interesse i samlingen af de molekylære spektre, og eventuelt **kendetegnet ved, at** udtrækningen udføres som funktion af spids-udvælgelseskriterier, der er defineret ved et kvalitetskriterium for et signal/støj-forhold og/eller en spektral opløsning.

7. Fremgangsmåde ifølge et hvilket som helst af kravene 1 til 6, **kendetegnet ved, at** den omfatter et trin til at normalisere det store antal datasæts (J1-Jn) molekylære spektre i forhold til hinanden for at kunne sammenligne de forskellige datasæt (J1-Jn) med hinanden.

8. Fremgangsmåde ifølge krav 7, **kendetegnet ved, at** fremgangsmåden for at udføre normaliseringstrinnet omfatter følgende trin:

- valg af et eller flere endogene eller exogene referencemolekyler (Mref), der er fælles for samlingen af datasæt (J1-Jn),
- for hvert datasæt (J1-Jn) beregning af en normaliseringsfaktor (Fnor) som funktion af dette eller disse referencemolekyler (Mref) og

- korrektion af den molekylære information af interesse med denne normaliseringsfaktor for at opnå en samling normaliserede datasæt (J1-Jn).

9. Fremgangsmåde ifølge krav 8, **kendetegnet ved, at** normaliseringsfaktoren (Fnor) er defineret ud fra mindst et referencemolekyle (Mref) der findes i en samling prøver, der svarer til datasættene (J1-Jn) eller i en særlig referencezone (Z) for prøver, der svarer til datasættene (J1-Jn).

10. Fremgangsmåde ifølge et hvilket som helst af kravene 7 til 9, **kendetegnet ved, at** den omfatter et trin til at visualisere, sammenligne og analysere, i form af et stort antal datakort (C1'-Cn'), der hvert især stammer fra et normaliseret datasæt (J1-Jn), den molekylære information af interesse med en farveskala, der er fælles for samlingen af kort.

11. Fremgangsmåde ifølge krav 10, **kendetegnet ved, at** den fælles farveskala er baseret på en mindste og en største intensitet for alle datasættene (J1-Jn) for molekylet af interesse.

12. Fremgangsmåde ifølge et hvilket som helst af kravene 1 til 11, **kendetegnet ved, at** datasættene (J1-Jn) er opnået ved en massespektrometrisk fremgangsmåde.

13. Fremgangsmåde ifølge et hvilket som helst af kravene 1 til 11, **kendetegnet ved, at** datasættene (J1-Jn) er opnået ved en billeddannelsesfemgangsmåde af typen positronemissionstomografi (PET) eller billeddannelse ved magnetisk resonans (MR).

14. Dataserver (10) med en hukommelse, der lagrer programinstruktioner til udførelse af fremgangsmådetrinnene til behandling af et stort antal datasæt (J1-Jn) som defineret i et hvilket som helst af de foregående krav.

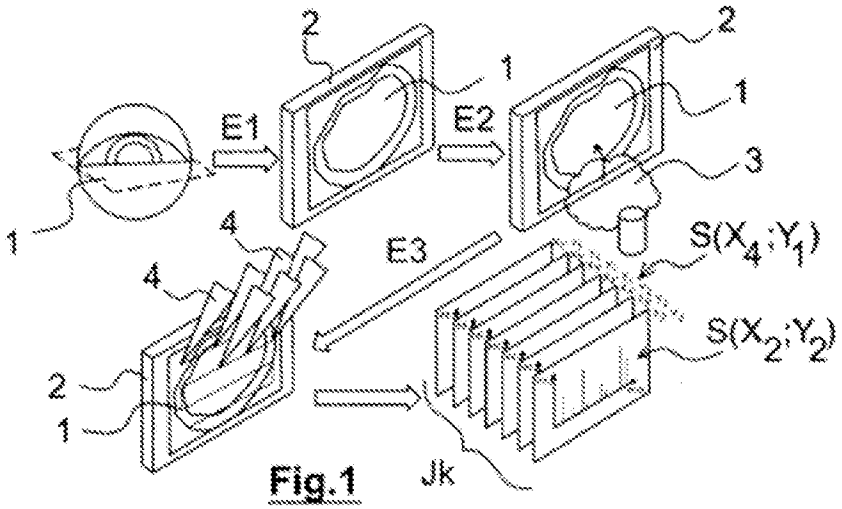

2
1
E1
E2
2
1
3
1
4
4
E3
$S(X_4;Y_1)$
$S(X_2;Y_2)$
2
1
Jk

Fig.1

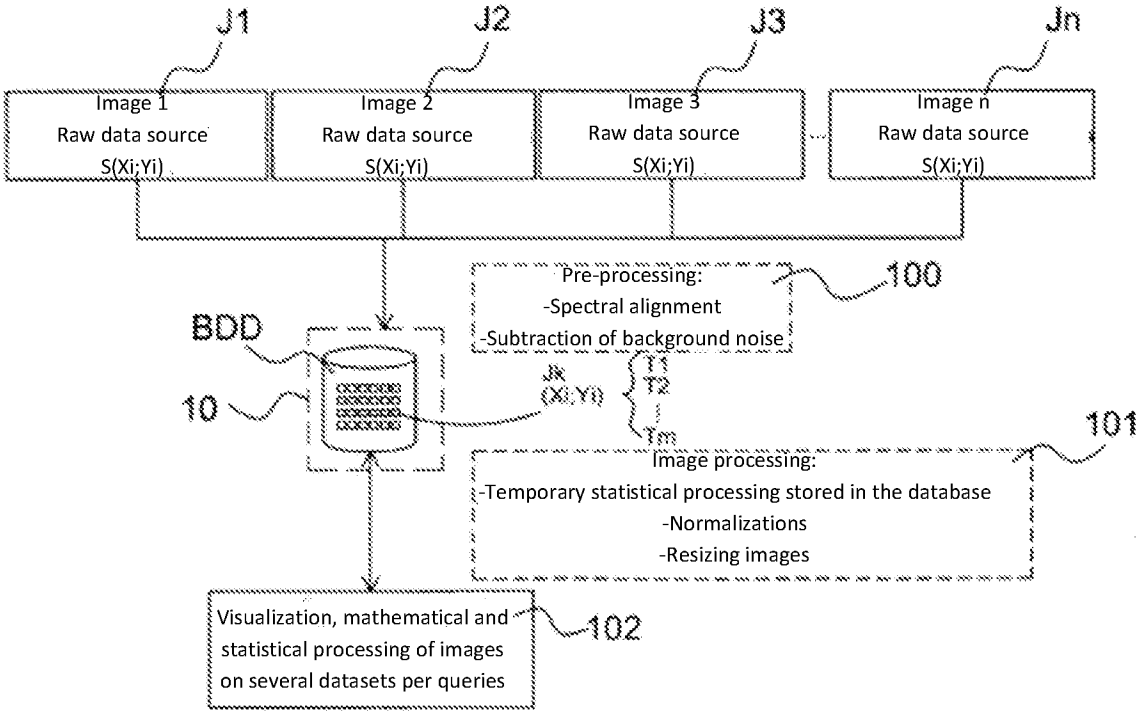

J1
J2
J3
Jn
Image 1
Raw data source
S(Xi;Yi)
Image 2
Raw data source
S(Xi;Yi)
Image 3
Raw data source
S(Xi;Yi)
Image n
Raw data source
S(Xi;Yi)
Pre-processing:
-Spectral alignment
-Subtraction of background noise
100
BDD
10
Jk
(Xi;Yi)
T1
T2
Tm
101
Image processing:
-Temporary statistical processing stored in the database
-Normalizations
-Resizing images
Visualization, mathematical and
statistical processing of images
on several datasets per queries
102

Fig.2

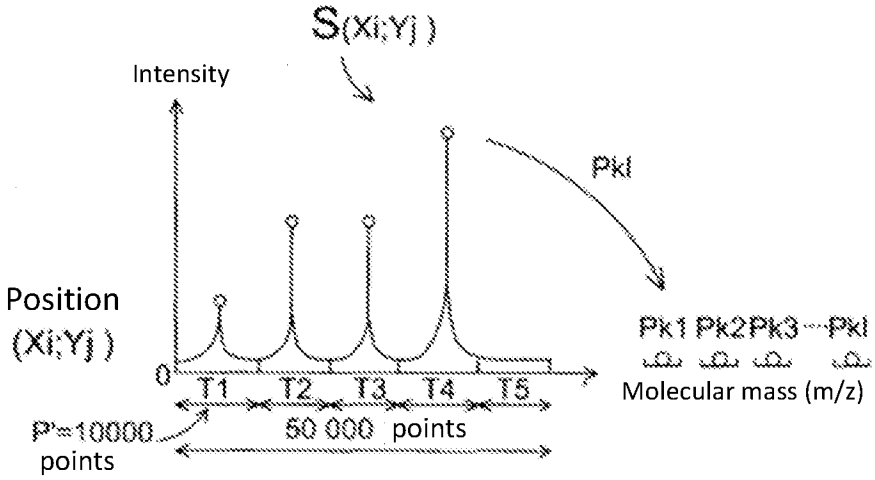

S(Xi;Yj)
Intensity
Pkl
Position
(Xi;Yj)
Pk1 Pk2 Pk3 ··· Pkl
Molecular mass (m/z)
0
T1
T2
T3
T4
T5
P'=10000
points
50 000 points

Fig.3

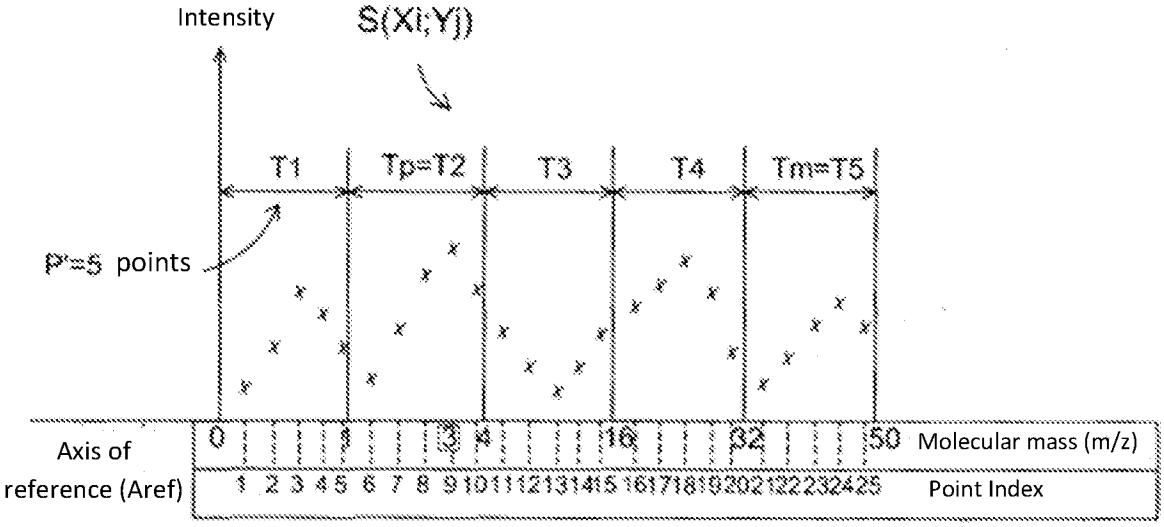

Intensity
S(Xi;Yj)
T1
Tp=T2
T3
T4
Tm=T5
P'=5 points
Axis of
reference (Aref)
0
8
16
32
50
Molecular mass (m/z)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Point Index

Fig.4

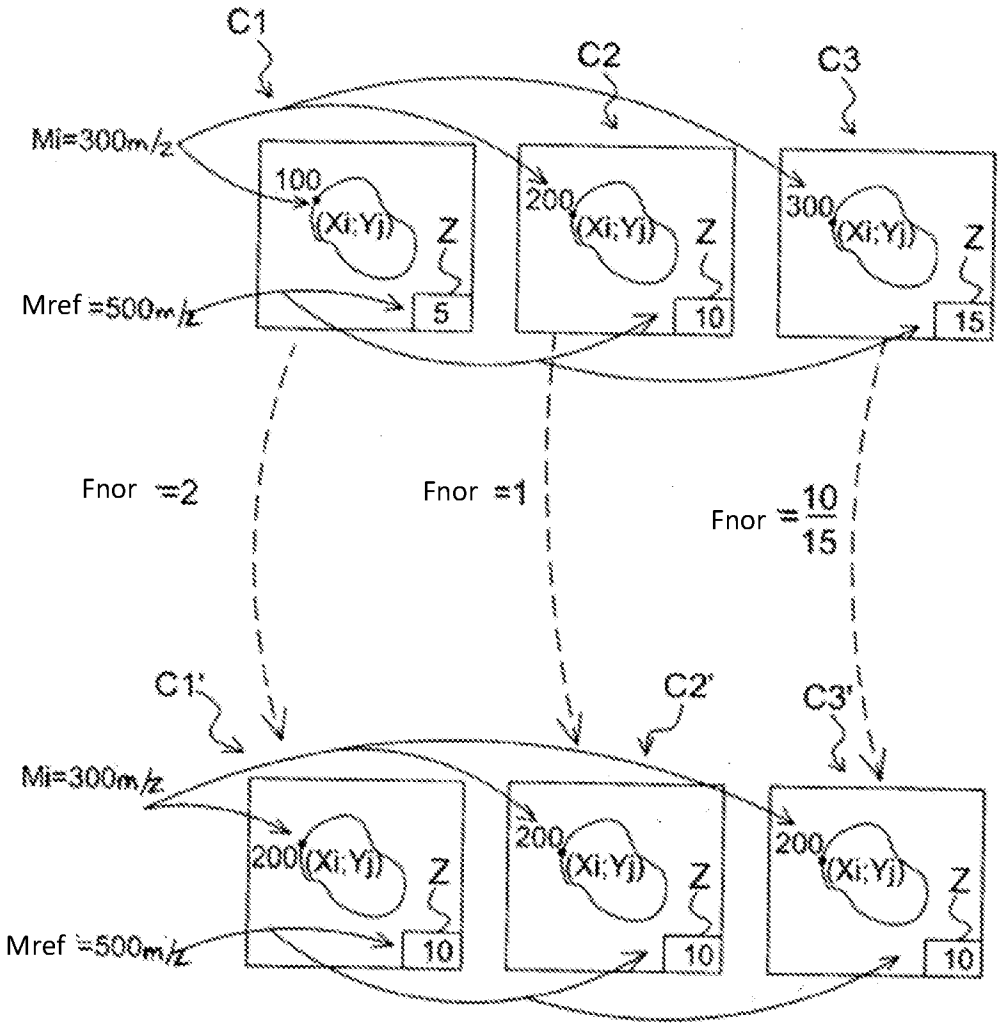

C1
C2
C3
Mi=300m/z
100
(Xi;Yj)
Z
5
200
(Xi;Yj)
Z
10
300
(Xi;Yj)
Z
15
Mref =500m/z
Fnor =2
Fnor =1
Fnor =10/15
C1'
C2'
C3'
Mi=300m/z
200
(Xi;Yj)
Z
10
200
(Xi;Yj)
Z
10
200
(Xi;Yj)
Z
10
Mref =500m/z

Fig.5