(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété Intellectuelle
Bureau international

(43) Date de la publication internationale
16 juillet 2015 (16.07.2015)

WIPO | PCT

(10) Numéro de publication internationale
WO 2015/104512 A1

(51) Classification internationale des brevets :
G06F 19/26 (2011.01) *G06F 19/28* (2011.01)

(21) Numéro de la demande internationale :
PCT/FR2015/050051

(22) Date de dépôt international :
9 janvier 2015 (09.01.2015)

(25) Langue de dépôt : français

(26) Langue de publication : français

(30) Données relatives à la priorité :
1450176 10 janvier 2014 (10.01.2014) FR

(71) Déposant : **IMABIOTECH** [FR/FR]; 885, avenue Eugène Avinée, F-59120 Loos (FR).

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

(74) Mandataires : **LEBRETTE, Camille** et al.; 25, rue Louis le Grand, F-75002 Paris (FR).

(81) États désignés *(sauf indication contraire, pour tout titre de protection nationale disponible)* : AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) États désignés *(sauf indication contraire, pour tout titre de protection régionale disponible)* : ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), eurasien (AM, AZ, BY, KG, KZ, RU, TJ, TM), européen (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,

[Suite sur la page suivante]

(54) Title : METHOD FOR PROCESSING MOLECULAR IMAGING DATA AND CORRESPONDING DATA SERVER

(54) Titre : PROCEDE DE TRAITEMENT DE DONNEES D'IMAGERIE MOLECULAIRE ET SERVEUR DE DONNEES CORRESPONDANT

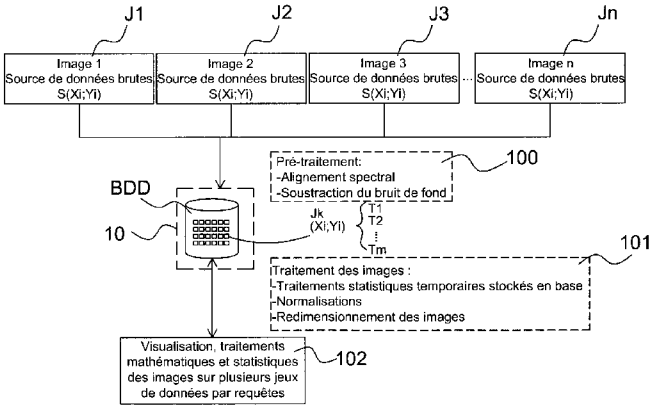

Fig.2

J1-Jn Raw data source
100 Pre-processing

- Spectral alignment
- Removal of background noise

101 Image processing:
- Temporary statistical processing stored in base
- Normalisations
- Image resizing

102 Viewing, mathematical and statistical processing of images on a plurality of datasets by requests

(57) **Abstract** : The invention relates mainly to a method for processing a plurality of spectral datasets (J1-Jn) intended for being used by a molecular imaging method or a method for recording a plurality of spectral datasets (J1-Jn), each spectral dataset (J1-Jn) being defined by a set of spatial positions (Xi, Yj) each of which is associated with a molecular spectrum with at least two dimensions containing a set of molecular information (S(Xi, Yj)), characterised in that it comprises in particular the following steps: for each dataset (J1-Jn), cutting the molecular spectrum associated with each position (Xi, Yj) into a plurality of spectrum segments (T1-Tm); inserting the segments (T1-Tm) obtained for each position (Xi, Yj) of each dataset (J1-Jn) into a database (BDD); selecting in the database (BDD), following a request relating to molecular information of interest, the one or more segments (T1-Tm) containing the molecular information of interest; and selecting, within each segment (T1-Tm), said molecular information of interest.

(57) **Abrégé** :

[Suite sur la page suivante]

WO 2015/104512 A1

SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Publiée :

— *avec rapport de recherche internationale (Art. 21(3))*

L'invention porte principalement sur un procédé de traitement d'une pluralité de jeux de données spectrales (J1-Jn) destinés à être exploités par un procédé d'imagerie moléculaire ou un procédé d'enregistrement d'une pluralité de jeux de données spectrales (J1-Jn), chaque jeu de données spectrales (J1-Jn) étant défini par un ensemble de positions spatiales (Xi, Yj) à chacune desquelles est associée un spectre moléculaire à au moins deux dimensions contenant un ensemble d'informations moléculaires (S(Xi, Yj)), caractérisé en ce qu'il comporte notamment les étapes suivantes : - pour chaque jeu de données (J1-Jn), découper le spectre moléculaire associé à chaque position (Xi, Yj) en plusieurs tronçons de spectre (T1-Tm), - insérer les tronçons (T1-Tm) obtenus pour chaque position (Xi, Yj) de chaque jeu de données (J1-Jn) dans une base de données (BDD), - sélectionner dans la base de données (BDD), suite à une requête relative à une information moléculaire d'intérêt, le ou les tronçons (T1-Tm) contenant l'information moléculaire d'intérêt, et - sélectionner, au sein de chaque tronçon (T1-Tm), ladite information moléculaire d'intérêt.

PROCEDE DE TRAITEMENT DE DONNEES D'IMAGERIE MOLECULAIRE ET SERVEUR DE DONNEES CORRESPONDANT

DOMAINE TECHNIQUE DE L'INVENTION

La présente invention porte sur un procédé de traitement de données d'imagerie moléculaire basé sur l'utilisation d'une structure de base de données à haute volumétrie, ainsi que sur le serveur de données correspondant. L'imagerie moléculaire englobe de façon générale toutes les techniques d'imagerie permettant essentiellement d'observer le fonctionnement moléculaire des organes et/ou organismes ex vivo ou in vivo par des moyens les moins invasifs possibles ou perturbants le moins possible les systèmes biologiques ou biophysiques observés.

ARRIERE PLAN TECHNOLOGIQUE

L'imagerie moléculaire connaît un essor considérable dans l'évolution du nombre de technologies notamment en imagerie par spectrométrie de masse (MSI pour "Mass Spectrometry Imaging" en anglais), imagerie par résonnance magnétique (MRI pour "Magnetic Resonance Imaging"), ou encore la spectrométrie Raman. L'utilisation de ces technologies permet de réaliser des études de bio-distribution de composés ciblés endogènes ou exogènes tels que les pesticides, les médicaments, les protéines ou les lipides afin d'étudier leurs rôles dans des systèmes biologiques. Les applications de ces technologies avec le concours de logiciels permettent également de rechercher de potentiels bio-marqueurs de manière non supervisée.

Toutefois, l'augmentation des données à analyser et à interpréter en imagerie moléculaire constitue une limitation pour les analyses synthétiques sur plusieurs jeux de données associés chacun à une image. Les données deviennent alors difficilement exploitables pour des représentations synthétiques ou analysables de façon statistique.

En effet, la haute volumétrie des informations des jeux de données est souvent supérieure à la taille de la mémoire vive des ordinateurs classiques (par opposition

aux supers calculateurs). En imagerie par spectrométrie de masse par exemple, des images de 50 000 positions ne permettent plus de stocker l'intégralité de l'information brute en mémoire et nécessitent donc la réalisation de calculs pour réduire l'impact en mémoire et permettre l'exploitation de ces images. Les calculs réduisent les données et par conséquent induisent des biais en ne prenant pas en compte toute l'information disponible. De plus, les outils d'analyse actuels ne présentent aucun moyen de normalisation de plusieurs jeux de données entre eux.

Il existe de nombreux formats de stockage liés principalement aux constructeurs des automates d'acquisition des jeux de données. Toutefois, aucun de ces formats n'est adapté et optimisé pour être interrogé de façon fine et rapide. Certains domaines de l'imagerie moléculaire ont mis au point des formats de stockage tels que l'Analyse 7.5 que l'on retrouve en IRM (Imagerie par Résonance Magnétique) et en imagerie par spectrométrie de masse. Dans l'imagerie par spectrométrie de masse, on trouve également le format iMZML qui ne constitue toutefois pas un système de stockage pour manipuler des données d'imagerie.

Des formats du type HDF5 (pour "Hierarchical Data Format 5" en anglais) basés sur une organisation hiérarchique des données ont récemment été utilisés dans le stockage des données à haute volumétrie en imagerie par spectrométrie de masse. Ces formats permettent de construire des interfaces d'interrogation à distance via des navigateurs Internet. Ils permettent des calculs statistiques plus rapides qu'avec des fichiers non hiérarchisés. Ces formats ne permettent toutefois pas d'obtenir une analyse comparative globale et interrogeable des données dans des études de plusieurs jeux de données simultanément.

Il existe donc le besoin d'un procédé permettant d'analyser, de comparer, et d'interroger plusieurs jeux de données les uns par rapport aux autres sans perte d'information afin d'éviter d'introduire des biais d'analyse.

OBJET DE L'INVENTION

La présente invention vise à répondre efficacement à ce besoin en proposant un procédé de traitement d'une pluralité de jeux de données spectrales, destinés notamment à être exploités par un procédé d'imagerie moléculaire, chaque jeu de

données étant défini par un ensemble de positions spatiales à chacune desquelles est associé un spectre moléculaire à au moins deux dimensions contenant un ensemble d'informations moléculaires, caractérisé en ce qu'il comporte notamment les étapes suivantes :
- pour chaque jeu de données, découper le spectre moléculaires associé à chaque position en plusieurs tronçons d'informations moléculaires, ou tronçons de spectre, contenant chacun un ensemble réduit d'informations moléculaires,
- insérer les tronçons obtenus pour chaque position de chaque jeu de données dans une base de données, en sorte qu'à chaque position de chaque jeu de données est associé un ensemble de tronçons de spectre indexés,
- sélectionner dans la base de données, suite à une requête relative à une information moléculaire d'intérêt, le ou les tronçons contenant l'information moléculaire d'intérêt, et
- sélectionner, au sein de chaque tronçon, ladite information moléculaire d'intérêt.

L'invention concerne également procédé d'enregistrement d'une pluralité de jeux de données spectrales lors d'un procédé d'imagerie moléculaire, chaque jeu de données étant défini par un ensemble de positions spatiales à chacune desquelles est associé un spectre moléculaire à au moins deux dimensions contenant un ensemble d'informations moléculaires, caractérisé en ce qu'il comporte les étapes suivantes :
- lors de l'acquisition des données spectrales, et pour chaque jeu de données spectrales, découper le spectre moléculaires associé à chaque position en plusieurs tronçons de spectre, contenant chacun une partie des informations moléculaires dudit spectre,
- insérer les tronçons obtenus pour chaque position de chaque jeu de données dans une base de données, en sorte qu'à chaque position de chaque jeu de données est associé un ensemble de tronçons de spectre indexés, et optionnellement- sélectionner dans la base de données, suite à une requête relative à une information moléculaire d'intérêt, le ou les tronçons contenant l'information moléculaire d'intérêt, et sélectionner, au sein de chaque tronçon, ladite information moléculaire d'intérêt.

Plus généralement, l'invention propose de traiter des jeux de données, acquis avantageusement de manière automatisée, en découpant en tronçons les spectres moléculaires associés à chacune des positions spatiales des ensembles de positions

spatiales composant lesdits jeux de données. A chaque spectre correspond ainsi une succession de tronçons de spectre, comprenant à eux tous l'ensemble des informations moléculaires du spectre associé. Selon l'invention, l'étape de découpage en tronçon peut être réalisée, avantageusement de manière automatisée, pendant l'acquisition des données spectrales ou pendant l'analyse des données spectrales.

Dans un premier mode de réalisation les tronçons successifs ne se chevauchent pas et sont simplement adjacent, de sorte qu'aucune information moléculaire d'un spectre considéré n'est perdue ni redondante. Dans un autre mode de réalisation, certains des tronçons d'un spectre considéré peuvent se chevaucher partiellement, de sorte que certaines informations moléculaires peuvent être contenues dans deux tronçons de spectre successifs d'un même spectre.

L'invention permet ainsi de réduire de manière substantielle la durée de chargement des données à analyser, dans la mesure où il est possible de récupérer, par sélection de tronçons, l'information utile dans la base de données. Il est également possible de tronçonner les données spectrales dès l'enregistrement des données, pendant leur acquisition, de sorte que les tronçons de données spectrales sont intégrés directement pendant l'analyse de l'échantillon par un procédé d'imagerie moléculaire.

Du fait du découpage en tronçons du signal, sans réduction du spectre, l'invention permet en outre de travailler sans réduction de l'information dans le cadre d'études combinées de pharmacocinétique, de pharmacodynamique et de quantification de molécules endogènes ou exogènes dans une même analyse afin d'observer dans son ensemble des événements physiologiques comme des méthodes de screening (comparaison de distribution de plusieurs candidats médicaments), méthodes de quantification, méthodes d'étude d'occupation de récepteurs ou de criblage tissulaire, d'étude de métabolisme ou de reconstruction 3D. L'invention permet également d'optimiser l'utilisation de la mémoire nécessaire aux traitements statistiques et aux représentations des données d'imagerie moléculaire sur plusieurs jeux de données.

Selon une mise en œuvre, le procédé de traitement ou d'enregistrement de données de l'invention comporte l'étape de visualiser l'information moléculaire d'intérêt sous forme d'un ensemble de cartographies de données correspondant chacune à un jeu de données.

Avantageusement, les spectres moléculaires sont découpés selon un pas déterminé, qui peut être identique ou différent d'un spectre à un autre.

Selon une mise en œuvre, le procédé de traitement ou d'enregistrement de données de l'invention comporte les étapes suivantes:
- découper l'ensemble d'informations moléculaires en tronçons de spectre suivant au moins un pas prédéterminé,
- insérer un axe de référence établissant une correspondance entre des index de points des différents tronçons de spectre et l'information moléculaire correspondante, et
- sélectionner le tronçon de spectre contenant l'information moléculaire d'intérêt en fonction de l'axe de référence et du pas prédéterminé.

Selon une mise en œuvre, avant ou après l'insertion des tronçons dans la base de données, ledit procédé comporte au moins une étape de pré-traitement consistant en un alignement spectral et/ou une soustraction de bruit de fond et/ou une normalisation intra jeu de données.

Selon une mise en œuvre, le procédé de traitement ou d'enregistrement de données de l'invention comporte l'étape de redimensionner les positions spatiales des jeux de données ayant des tailles spatiales différentes de manière à s'aligner sur le jeu de données ayant la taille spatiale la plus fine.

Selon une mise en œuvre, le procédé de traitement ou d'enregistrement de données de l'invention comporte en outre l'étape d'extraire un ensemble de données, telles qu'une intensité maximum, une intensité moyenne, une aire sous un pic, un rapport signal sur bruit de chaque pic d'intérêt de l'ensemble d'informations moléculaires.

Selon une mise en œuvre, l'extraction est effectuée en fonction de critères de sélection de pic définis par un critère de qualité d'un rapport signal sur bruit et/ou une résolution spectrale.

Selon une mise en œuvre, le procédé de traitement ou d'enregistrement de données de l'invention comporte l'étape de normaliser les ensembles d'informations moléculaires de la pluralité des jeux de données entre eux de manière à pouvoir comparer les différents jeux de données entre eux.

Selon une mise en œuvre, pour réaliser l'étape de normalisation, ledit procédé peut comporter les étapes suivantes:

- choisir une molécule ou plusieurs molécules endogènes ou exogènes de référence communes à l'ensemble des jeux de données,
- calculer, pour chaque jeu de données, un facteur de normalisation en fonction de cette ou ces molécules de référence, et
- corriger l'information moléculaire d'intérêt avec ce facteur de normalisation pour obtenir un ensemble de jeux de données normalisés.

Selon une mise en œuvre, le facteur de normalisation est défini à partir d'au moins une molécule de référence présente sur un ensemble d'échantillons correspondant aux jeux de données ou dans une zone particulière de référence des échantillons correspondant aux jeux de données.

Selon une mise ne œuvre, le procédé de traitement ou d'enregistrement de données de l'invention comporte l'étape de visualiser, comparer, et analyser, sous forme d'une pluralité de cartographies de données issues chacune d'un jeu de données normalisé, l'information moléculaire d'intérêt avec une échelle de couleur commune à l'ensemble des cartographies.

Selon une mise en œuvre, l'échelle de couleur commune est basée sur une intensité la plus petite et la plus grande de tous les jeux de données pour la molécule d'intérêt.

Selon une mise en œuvre, lesdits jeux de données sont obtenus par un procédé de spectrométrie de masse.

Selon une mise en œuvre, lesdits jeux de données sont obtenus par un procédé d'Imagerie de type Tomographie par Emission de Positons (TEP) ou d'Imagerie par Résonance Magnétique (IRM).

Avantageusement, tout ou partie des étapes du procédé d'enregistrement et/ou de traitement selon l'invention sont automatisées.

L'invention a également pour objet un serveur de données comportant une mémoire stockant des instructions logicielles pour la mise en œuvre d'au moins une partie des étapes du procédé de traitement d'une pluralité de jeux de données selon la présente invention.

L'invention a également pour objet un support de données lisible par ordinateur comprenant des instructions exécutables par l'ordinateur et adaptées pour permettre à un système informatique d'exécuter au moins une étape du procédé de traitement ou d'enregistrement de données spectrales selon l'invention.
Ainsi, l'invention propose un programme d'ordinateur comprenant des instructions de code de programme pour l'exécution de tout ou partie des étapes exposées ci-dessus, lorsque ledit programme est exécuté sur un ordinateur.
Avantageusement, le programme d'ordinateur comprend des instructions de code de programme pour exécuter au moins l'étape de découpe des spectres moléculaires en tronçons de spectre.

BREVE DESCRIPTION DES FIGURES

L'invention sera mieux comprise à la lecture de la description qui suit et à l'examen des figures qui l'accompagnent. Ces figures ne sont données qu'à titre illustratif mais nullement limitatif de l'invention.

La figure 1 représente de manière schématique les différentes étapes d'un procédé de spectroscopie de masse pour l'obtention de jeux de données exploités par le procédé selon la présente invention;

La figure 2 est un schéma synoptique du système intégré de gestion de base de données permettant la mise en œuvre du procédé de traitement de données selon la présente invention;

La figure 3 est une représentation schématique de l'étape permettant d'obtenir les tronçons des spectres associés à chaque position d'un jeu de données ;

La figure 4 représente les étapes permettant d'établir la correspondance entre une information moléculaire et le tronçon de spectre correspondant ;

La figure 5 montre de manière schématique les étapes de normalisation inter-images permettant de comparer de manière précise plusieurs images entre elles.

Les éléments identiques, similaires, ou analogues conservent les mêmes références d'une figure à l'autre.

DESCRIPTION D'UN EXEMPLE DE REALISATION DE L'INVENTION

L'exemple de mise en œuvre du procédé selon l'invention qui suit est décrit pour une méthode de spectrométrie de masse de type MALDI. Bien entendu, on pourrait de manière sensiblement identique utiliser un autre dispositif d'imagerie que le MALDI, comme par exemple les sources : SIMS, DESI , LAESI, DIOS, ICP, Microscope MALDI, SNOM, SMALDI, LA-ICP, ESI (extraction liquide sur tissu), MILDI, JEDI, ELDI etc. Le procédé est également applicable à toute autre méthode permettant de produire des données exploitables par une méthode d'imagerie moléculaire comme par exemple la TEP ("Tomographie par Emission de Positons") ou l'IRM (Imagerie par Résonance Magnétique).

La figure 1 montre de manière schématique les différentes étapes mises en œuvre par un procédé de spectrométrie de masse MALDI. Une première étape E1 consiste à découper, en général par cryo-section, une tranche de tissu 1 qui est ensuite posée sur une lame 2. On dépose ensuite dans une étape E2 une fine couche uniforme de matrice ionisante 3 sur la tranche de tissu 1. L'étape d'acquisition E3 consiste à générer des impulsions laser automatisées 4 ayant une taille prédéfinie, par exemple de l'ordre de 100 micromètres par zone d'acquisition, afin d'ioniser les molécules de la matrice 3. Les molécules ainsi ionisées sont analysées de façon connue en soi par un spectromètre de masse. On pourra se référer au document FR2973112 pour plus de détails sur la mise en œuvre du procédé.

On obtient alors un jeu de données Jk défini par un ensemble de positions spatiales (Xi, Yj) à chacune desquelles est associé un ensemble d'informations moléculaires. Les positions spatiales sont définies en fonction d'un référentiel de référence d'axes X et Y. L'ensemble d'informations moléculaires est formé en l'occurrence notamment par un spectre S à deux dimensions indiquant une intensité en fonction d'une masse moléculaire. Par exemple, le jeu de données Jk obtenu présente 50000 intensités (soit 50000 points) par spectre sur 20000 positions spatiales Xi, Yj. Le procédé selon l'invention est basé sur le traitement d'une pluralité de jeux de données J1-Jn.

Comme cela est montré sur la figure 3, pour chaque jeu de données Jk susceptible d'être représenté sous forme d'une image, le spectre moléculaire S(Xi, Yj) associé à chaque position (Xi, Yj) est découpé en plusieurs tronçons de spectres T1-Tm. Les tronçons T1-Tm sont découpés suivant un pas prédéterminé P'. Par exemple, pour

chaque spectre de 50000 points par position spatiale, chaque spectre pourra être découpé suivant un pas P' de 10000 points. Autrement dit, le spectre S(Xi, Yj) est découpé en cinq tronçons T1-T5 de 10000 points chacun. En variante, le pas P' utilisé pourra être variable.

On définit par ailleurs un axe de référence Aref, montré sur la figure 4, établissant une correspondance entre des index de points des différents tronçons T1-Tm et la masse moléculaire correspondante.

De préférence, avant insertion dans une base de données BDD, le procédé met en œuvre au moins une étape de pré-traitement des spectres S(Xi, Yj) via un module 100. Cette étape de pré-traitement pourra consister en un alignement spectral. A cet effet, on considère des pics de référence et on décale le spectre de manière à positionner les valeurs de pics correspondants sur ces pics de référence. Il sera également possible d'effectuer une soustraction du bruit de fond en recalant le spectre en fonction d'un écart entre la valeur minimale du spectre et une valeur de référence à laquelle devrait correspondre la valeur minimale du spectre.

En outre, différents types de normalisation du signal (normalisation interne) pourront être réalisés (par un standard interne, par utilisation du total des intensités…). Autrement dit, on réalise ici une normalisation intra jeu de données, c'est-à-dire une normalisation pour chaque jeu de données indépendamment des autres jeux de données, ce qui correspond à des étapes connues en soi de l'homme du métier. Cette normalisation intra jeu de données est effectuée en complément des étapes ultérieures de normalisation entre les jeux de données J1-Jn détaillées ci-après (normalisation inter jeu de données) qui permettront de comparer, de manière précise et sûre, les jeux de données J1-Jn entre eux.

Par ailleurs, le procédé comporte l'étape d'extraire un ensemble de données, telles qu'une intensité maximum Pk1-Pkl, une intensité moyenne, une aire sous un pic, un rapport signal sur bruit de chaque pic d'intérêt des spectres S(Xi, Yj). Il est ainsi possible d'extraire l'intensité maximum de chaque pic Pk1-Pkl du spectre S(Xi, Yj) suivant une méthode dite de "PeakPicking" en anglais. L'extraction est effectuée en fonction de critères de sélection de pics définis par un critère de qualité du rapport signal sur bruit et/ou la résolution spectrale.

Les tronçons T1-Tm obtenus pour chaque position (Xi, Yj) de chaque jeu de données J1-Jn ainsi que l'axe de référence Aref et les pics Pk1-PKl précédemment extraits sont insérés dans la base de données BDD comme cela est montré à la figure 2. Ainsi, à chaque position (Xi, Yj) de chaque jeu de données J1-Jn sont associés un ensemble de tronçons indexés T1-TM et un ensemble de pics Pk1-PKl et de données complémentaires précitées (intensités moyennes, aires de chaque pic).

La base de données BDD est une base de données à haute volumétrie du type NoSQL (pour "Not only Structured Query Langage" en anglais) ou SQL permettant la visualisation, la gestion, l'interprétation et l'analyse statistique des données insérées. Le système de gestion de base de données peut par exemple être choisi parmi les systèmes non exhaustifs suivants: Hypertable, Haddoop, Cassandra, MongoDB, PolyBase, Hadoop on Azure, Hive, Pig. Le système de base de données assure la cohérence, la fiabilité, et la pertinence des données contenues. Des techniques de réplications et de sauvegardes peuvent permettre de garantir la fiabilité de l'information.

On décrit ci-après les étapes permettant la sélection de l'information pertinente dans la base de données BDD. Suite à une requête relative à une information moléculaire d'intérêt, on sélectionne le ou les tronçons des positions des ensembles de jeu de données J1-Jn contenant l'information moléculaire demandée. A cette fin, pour chaque position Xi, Yj de chaque jeu de données J1-Jn, le tronçon Tp contenant l'information moléculaire demandée est sélectionné en fonction de l'axe de référence Aref et de l'information moléculaire demandée.

Ainsi, sur la figure 4, on a illustré une découpe d'un spectre en tronçon suivant un pas P' de cinq points pour faciliter la compréhension de cette étape. Comme cela est souvent le cas pour les spectres issus d'une méthode de spectrométrie de masse, l'échelle des masses moléculaires n'est pas linéaire. Dans le cas où l'utilisateur requiert des informations relatives à une molécule ayant par exemple une masse moléculaire de valeur 3 m/z, l'axe de référence Aref permet d'établir que l'index du point correspondant à cette valeur de masse moléculaire est le 9, en sorte que l'on peut en déduire à l'aide du pas P' que le tronçon Tp contenant l'intensité de la masse moléculaire d'intérêt est le deuxième tronçon T2.

Alternativement, on peut utiliser une table qui stocke les bornes minimum et maximum des masses moléculaires et l'indexation des points de chaque tronçon T1-

Tm correspondant. Dans ce cas, pour une information moléculaire demandée, on sélectionne l'identifiant du tronçon T1-Tm en fonction de ces bornes minimum et maximum.

Au sein de chaque tronçon T1-Tm sélectionné pour chacun des points de l'ensemble des jeux de données J1-Jn, l'intensité correspondant de la masse moléculaire demandée est alors sélectionnée. Il devient alors possible de visualiser, sous forme d'un ensemble de cartographies de données C1-CN correspondant chacune à un jeu de données J1-Jn, la répartition de l'intensité de présence de la molécule d'intérêt sur les différentes cartographies C1-CN. On attribue une couleur en fonction de l'intensité de l'information moléculaire sur une échelle de couleur. Alternativement, ou de façon complémentaire, une extraction des pics Pk1-PKl permet de produire rapidement des images fournissant une première analyse de la répartition de l'intensité d'une molécule d'intérêt.

De préférence, le procédé comporte l'étape de normaliser les informations moléculaires des jeux de données J1-Jn afin de pouvoir comparer plusieurs jeux de données entre eux (cf. module 101). A cet effet, on choisit une molécule de référence Mref commune à tous les jeux de données J1-Jn. Alternativement, on choisit plusieurs molécules endogènes ou exogènes communes à l'ensemble des jeux de données J1-Jn. Il est possible de tenir compte de la présence de la ou des molécules de référence Mref dans les échantillons complets ou plus spécifiquement dans une zone de référence des échantillons correspondants.

Plus précisément, pour chaque jeu de données J1-Jn, un facteur de normalisation Fnor est calculé en fonction de cette ou ces molécules de référence Mref. Les informations moléculaires d'intérêt sont alors corrigées avec ce facteur de normalisation Fnor de manière à obtenir un ensemble de jeux de données J1-Jn normalisés.

Par exemple, pour trois jeux de données correspondant aux cartographies C1-C3 représentées sur la figure 5, on choisit une molécule de référence Mref=500 *m/z* en tant que molécule normalisante (standard normalisant). Cette molécule Mref présente respectivement une valeur de 5, 10 et 15 dans la zone Z des cartographies C1-C3. Pour chaque cartographie C1-C3, on calcule un facteur de normalisation Fnor sur la base de chacune de ces valeurs de Mref :

Fnor= Moyenne des valeurs de Mref pour C1-C3 / valeur du Mref de la cartographie considérée. Dans l'exemple représenté à la figure 5, on obtient un facteur de normalisation Fnor respectivement de 2, 1, et 10/15 pour les cartographies C1, C2 et C3. Ce coefficient est alors appliqué à l'ensemble des valeurs d'intensité des spectres associés aux différentes positions de l'ensemble des jeux de données J1-Jn et en particulier aux valeurs d'intensité d'une molécule d'intérêt Mi=300 *m/z* à la position Xi; Yj. On obtient alors des cartographies normalisées C1'-C3', qui peuvent être directement comparées entre elles.

La normalisation permet ainsi de corriger les intensités liées à des changements de conditions expérimentales dus à la préparation de l'échantillon et à son analyse. En effet, dans l'exemple précédent, on remarque qu'il aurait été possible d'arriver à une conclusion erronée d'une présence plus forte de la molécule d'intérêt Mi=300 *m/z* dans l'échantillon correspondant à la cartographie C3, alors qu'après application du facteur de normalisation Fnor, l'étude fait ressortir une présence homogène de la molécule d'intérêt Mi dans les trois échantillons avec une intensité de 200 à la position Xi; Yj.

Suite à cette normalisation, il est alors possible de visualiser, comparer et analyser, sous forme de cartographies de données C1'-C3' normalisées l'information moléculaire avec une échelle de couleur commune à l'ensemble des cartographies C1'-C3' (cf. module 102). L'échelle obtenue est basée sur l'intensité la plus petite et la plus grande déterminées sur tous les jeux de données pour la molécule d'intérêt Mi.

Bien entendu, il est possible de produire plus de trois jeux de données et cartographies correspondantes pour analyser par regroupements statistiques plusieurs jeux de données J1-Jn de façon transverse sans avoir besoin de tous les charger en mémoire. On peut alors comparer les données d'imagerie pour en faire ressortir des marqueurs potentiels. Il est également possible de réaliser des regroupements de positions ayant un comportement commun afin de permettre une nouvelle interprétation des données.

Autrement dit, l'invention permet ainsi, dans un seul outil, de comparer et interroger les données pour l'obtention d'une information plus pertinente, normalisée et dont la qualité est quantifiable. Plusieurs jeux de données peuvent être chargés avec des

contrôles qualités insérés dans chacune des images permettant ainsi de normaliser les données entre elles.

En outre, le procédé peut le cas échéant permettre de redimensionner les positions de jeux de données J1 et J2 du même échantillon biologique ayant des tailles spatiales différentes de manière à s'aligner sur le jeu de données ayant la plus fine taille spatiale. Il est alors possible de combiner le traitement d'une cartographie issue du spectromètre de masse (imagerie MALDI) avec le traitement d'une cartographie issue d'un système de type PET (Tomographie par Emission de Positons) ou IRM (Imagerie par Résonance Magnétique) ou IRM (Imagerie par Résonance Magnétique).

Il est également possible de représenter plusieurs jeux de données J1-Jn en codistribution dans une image unique. Il est alors possible de réaliser des requêtes du type "sélectionner toutes les valeurs v, w de toutes les positions x et y (de même taille spatiale) pour le champ z et a de la cartographie courante". Les champs z et a représentent les valeurs de deux technologies d'imagerie différente.

Il est également possible de réaliser des traitements statistiques temporaires qui pourront être ajoutés par des requêtes d'insertion dans la base de données BDD. Ainsi, le procédé selon l'invention apporte une grande flexibilité dans la mesure où il est possible d'ajouter de l'information sur chaque position de l'image à tout moment du procédé d'analyse.

Le procédé est basé sur une architecture client/serveur. La base de données BDD est interrogeable par une interface unique. La donnée est stockée et interrogée dans une architecture monoposte ou en cluster de serveurs accessible via en local ou via un réseau (interne ou externe). L'interface logicielle cliente peut être installée sur le poste de travail ou en version interface web. L'invention a également pour objet le serveur de données 10 comportant une mémoire stockant des instructions logicielles pour la mise en œuvre d'au moins une partie des étapes du procédé de traitement d'une pluralité de jeux de données J1-Jn.

De plus, la quantité de données traitée peut croître facilement, en sorte que la volumétrie de la donnée à analyser n'est plus une limitation à l'analyse. L'insertion des données est performante quelle que soit la taille de la base de données BDD.

Bien entendu, la description qui précède a été donnée à titre d'exemple uniquement et ne limite pas le domaine de l'invention dont on ne sortirait pas en remplaçant les détails d'exécution par tous autres équivalents.

REVENDICATIONS

1. Procédé de traitement d'une pluralité de jeux de données spectrales (J1-Jn) destinés à être exploités par un procédé d'imagerie moléculaire, chaque jeu de données spectrales (J1-Jn) étant défini par un ensemble de positions spatiales (Xi, Yj) à chacune desquelles est associée un spectre moléculaire à au moins deux dimensions contenant un ensemble d'informations moléculaires (S(Xi, Yj)), caractérisé en ce qu'il comporte notamment les étapes suivantes :
- pour chaque jeu de données (J1-Jn), découper le spectre moléculaire associé à chaque position (Xi, Yj) en plusieurs tronçons de spectre (T1-Tm) contenant chacun une partie des informations moléculaires dudit spectre,
- insérer les tronçons (T1-Tm) obtenus pour chaque position (Xi, Yj) de chaque jeu de données (J1-Jn) dans une base de données (BDD), en sorte qu'à chaque position (Xi, Yj) de chaque jeu de données (J1-Jn) est associé un ensemble de tronçons de spectre (T1-Tm) indexés,
- sélectionner dans la base de données (BDD), suite à une requête relative à une information moléculaire d'intérêt, le ou les tronçons (T1-Tm) contenant l'information moléculaire d'intérêt, et
- sélectionner, au sein de chaque tronçon (T1-Tm), ladite information moléculaire d'intérêt.

2- Procédé d'enregistrement d'une pluralité de jeux de données spectrales lors d'un procédé d'imagerie moléculaire, chaque jeu de données étant défini par un ensemble de positions spatiales à chacune desquelles est associé un spectre moléculaire à au moins deux dimensions contenant un ensemble d'informations moléculaires, caractérisé en ce qu'il comporte les étapes suivantes :
- lors de l'acquisition des données spectrales, et pour chaque jeu de données spectrales, découper le spectre moléculaires associé à chaque position en plusieurs tronçons de spectre, contenant chacun une partie des informations moléculaires dudit spectre,
- insérer les tronçons obtenus pour chaque position de chaque jeu de données dans une base de données, en sorte qu'à chaque position de chaque jeu de données est associé un ensemble de tronçons de spectre indexés, et optionnellement
- sélectionner dans la base de données, suite à une requête relative à une information moléculaire d'intérêt, le ou les tronçons contenant l'information moléculaire d'intérêt,

et sélectionner, au sein de chaque tronçon, ladite information moléculaire d'intérêt.

3. Procédé selon la revendication 1 ou 2, caractérisé en ce qu'il comporte l'étape de visualiser l'information moléculaire d'intérêt sous forme d'un ensemble de cartographies de données (C1-Cn) correspondant chacune à un jeu de données (J1-Jn).

4. Procédé selon l'une quelconque des revendications 1 à 3, caractérisé en ce qu'il comporte les étapes suivantes:
- découper les spectres moléculaires (S(Xi, Yj)) en tronçons de spectre (T1-Tm) suivant au moins un pas prédéterminé (P'),
- insérer un axe de référence (Aref) établissant une correspondance entre des index de points des différents tronçons (T1-Tm) et l'information moléculaire correspondante, et
- sélectionner le tronçon (T1-Tm) contenant l'information moléculaire d'intérêt en fonction de l'axe de référence (Aref) et du pas prédéterminé (P').

5. Procédé selon l'une quelconque des revendications 1 à 4, caractérisé en ce que, avant ou après l'insertion des tronçons (T1-Tm) dans la base de données (BDD), ledit procédé comporte au moins une étape de pré-traitement consistant en un alignement spectral et/ou une soustraction de bruit de fond et/ou une normalisation intra jeu de données (J1-Jn).

6. Procédé selon l'une quelconque des revendications 1 à 5, caractérisé en ce qu'il comporte l'étape de redimensionner les positions spatiales (Xi, Yj) des jeux de données (J1-Jn) ayant des tailles spatiales différentes de manière à s'aligner sur le jeu de données (J1-Jn) ayant la taille spatiale la plus fine.

7. Procédé selon l'une quelconque des revendications 1 à 6, caractérisé en ce qu'il comporte en outre l'étape d'extraire un ensemble de données, telles qu'une intensité maximum (Pk1-Pkl), une intensité moyenne, une aire sous un pic, un rapport signal sur bruit de chaque pic d'intérêt de l'ensemble des spectres moléculaires.

8. Procédé selon la revendication 7, caractérisé en ce que l'extraction est effectuée en

fonction de critères de sélection de pic définis par un critère de qualité d'un rapport signal sur bruit et/ou une résolution spectrale.

9. Procédé selon l'une quelconque des revendications 1 à 8, caractérisé en ce qu'il comporte l'étape de normaliser les spectres moléculaires de la pluralité des jeux de données (J1-Jn) entre eux de manière à pouvoir comparer les différents jeux de données (J1-Jn) entre eux.

10. Procédé selon la revendication 9, caractérisé en ce que pour réaliser l'étape de normalisation, ledit procédé comporte les étapes suivantes:
- choisir une molécule ou plusieurs molécules endogènes ou exogènes de référence (Mref) communes à l'ensemble des jeux de données (J1-Jn),
- calculer, pour chaque jeu de données (J1-Jn), un facteur de normalisation (Fnor) en fonction de cette ou ces molécules de référence (Mref), et
- corriger l'information moléculaire d'intérêt avec ce facteur de normalisation pour obtenir un ensemble de jeux de données (J1-Jn) normalisés.

11. Procédé selon la revendication 10, caractérisé en ce que le facteur de normalisation (Fnor) est défini à partir d'au moins une molécule de référence (Mref) présente sur un ensemble d'échantillons correspondant aux jeux de données (J1-Jn) ou dans une zone particulière de référence (Z) des échantillons correspondant aux jeux de données (J1-Jn).

12. Procédé selon l'une quelconque des revendications 9 à 11, caractérisé en ce qu'il comporte l'étape de visualiser, comparer, et analyser, sous forme d'une pluralité de cartographies de données (C1'-Cn') issues chacune d'un jeu de données (J1-Jn) normalisé, l'information moléculaire d'intérêt avec une échelle de couleur commune à l'ensemble des cartographies.

13. Procédé selon la revendication 12, caractérisé en ce que l'échelle de couleur commune est basée sur une intensité la plus petite et la plus grande de tous les jeux de données (J1-Jn) pour la molécule d'intérêt.

14. Procédé selon l'une quelconque des revendications 1 à 13, caractérisé en ce que lesdits jeux de données (J1-Jn) sont obtenus par un procédé de spectrométrie de masse.

15. Procédé selon l'une quelconque des revendications 1 à 13, caractérisé en ce que

lesdits jeux de données (J1-Jn) sont obtenus par un procédé d'imagerie de type Tomographie par Emission de Positons (TEP) ou d'Imagerie par Résonance Magnétique (IRM).

16. Serveur de données (10) comportant une mémoire stockant des instructions logicielles pour la mise en œuvre d'au moins une partie des étapes du procédé de traitement d'une pluralité de jeux de données (J1-Jn) défini selon l'une quelconque des revendications précédentes.

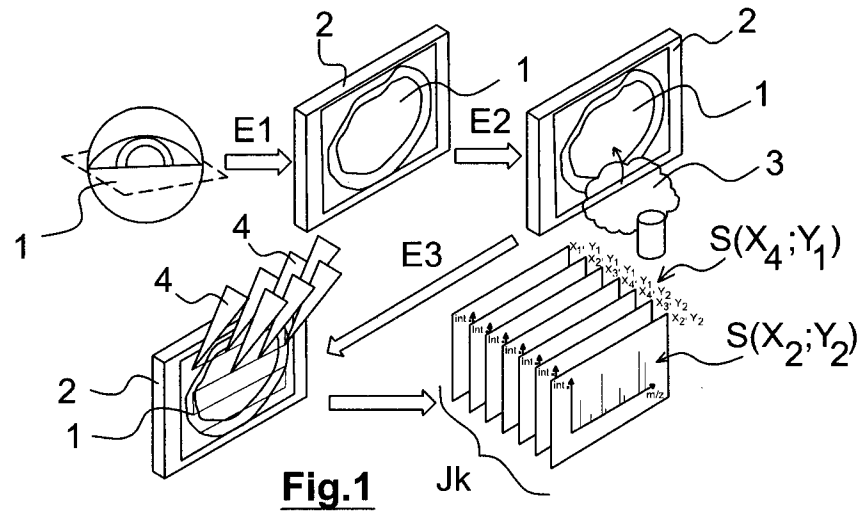
2
1
E1
E2
2
1
3
1
4
E3
S(X4;Y1)
4
S(X2;Y2)
2
1
Jk

Fig.1

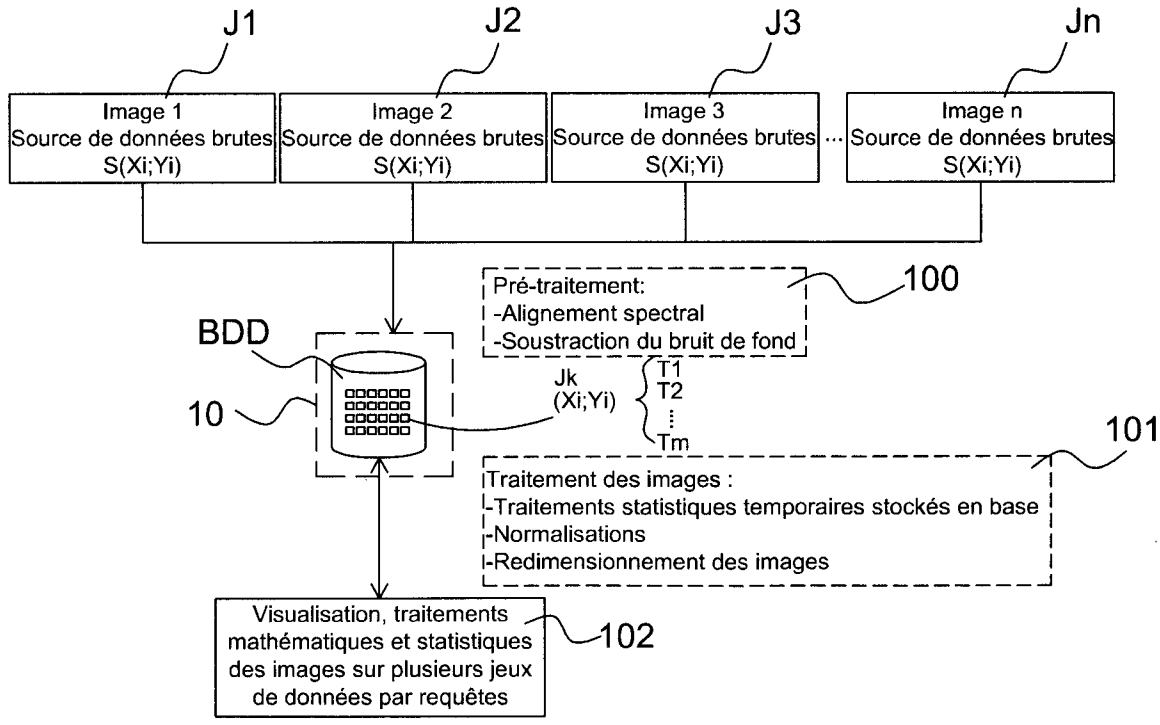
J1
J2
J3
Jn
Image 1
Source de données brutes
S(Xi;Yi)
Image 2
Source de données brutes
S(Xi;Yi)
Image 3
Source de données brutes
S(Xi;Yi)
Image n
Source de données brutes
S(Xi;Yi)
Pré-traitement:
-Alignement spectral
-Soustraction du bruit de fond
100
BDD
10
Jk
(Xi;Yi)
T1
T2
Tm
101
Traitement des images :
-Traitements statistiques temporaires stockés en base
-Normalisations
-Redimensionnement des images
Visualisation, traitements mathématiques et statistiques des images sur plusieurs jeux de données par requêtes
102

Fig.2

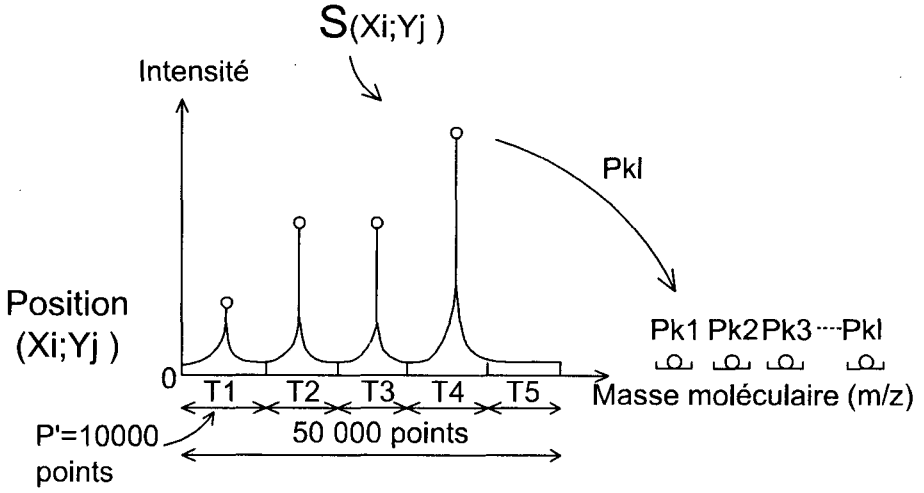
S(Xi;Yj)
Intensité
Pkl
Position
(Xi;Yj)
Pk1 Pk2 Pk3 ···· Pkl
0
T1
T2
T3
T4
T5
Masse moléculaire (m/z)
P'=10000
points
50 000 points

Fig.3

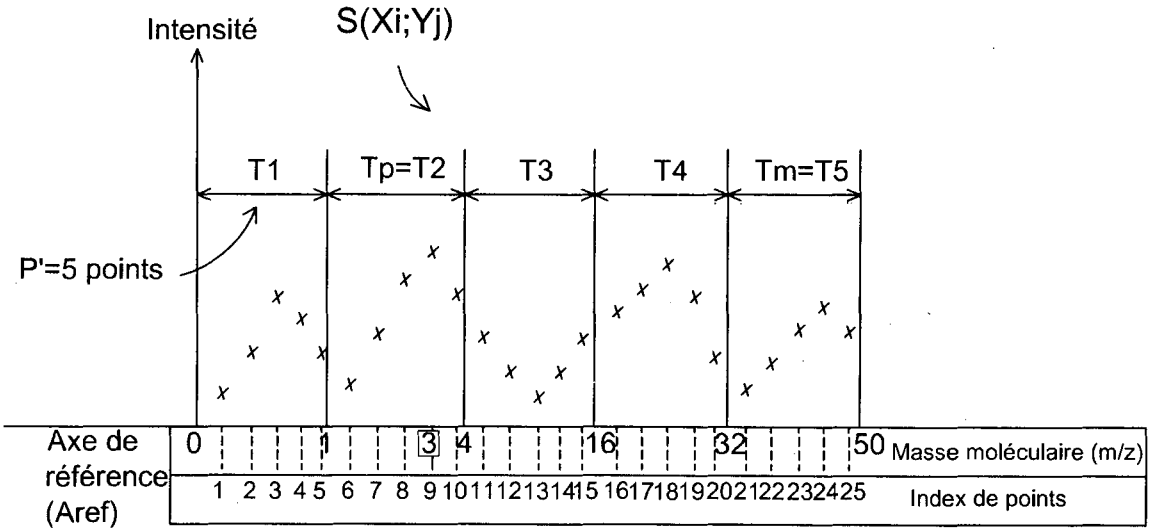
Intensité
S(Xi;Yj)
T1
Tp=T2
T3
T4
Tm=T5
P'=5 points
Axe de
référence
(Aref)
0
1
3
4
16
32
50
Masse moléculaire (m/z)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
Index de points

Fig.4

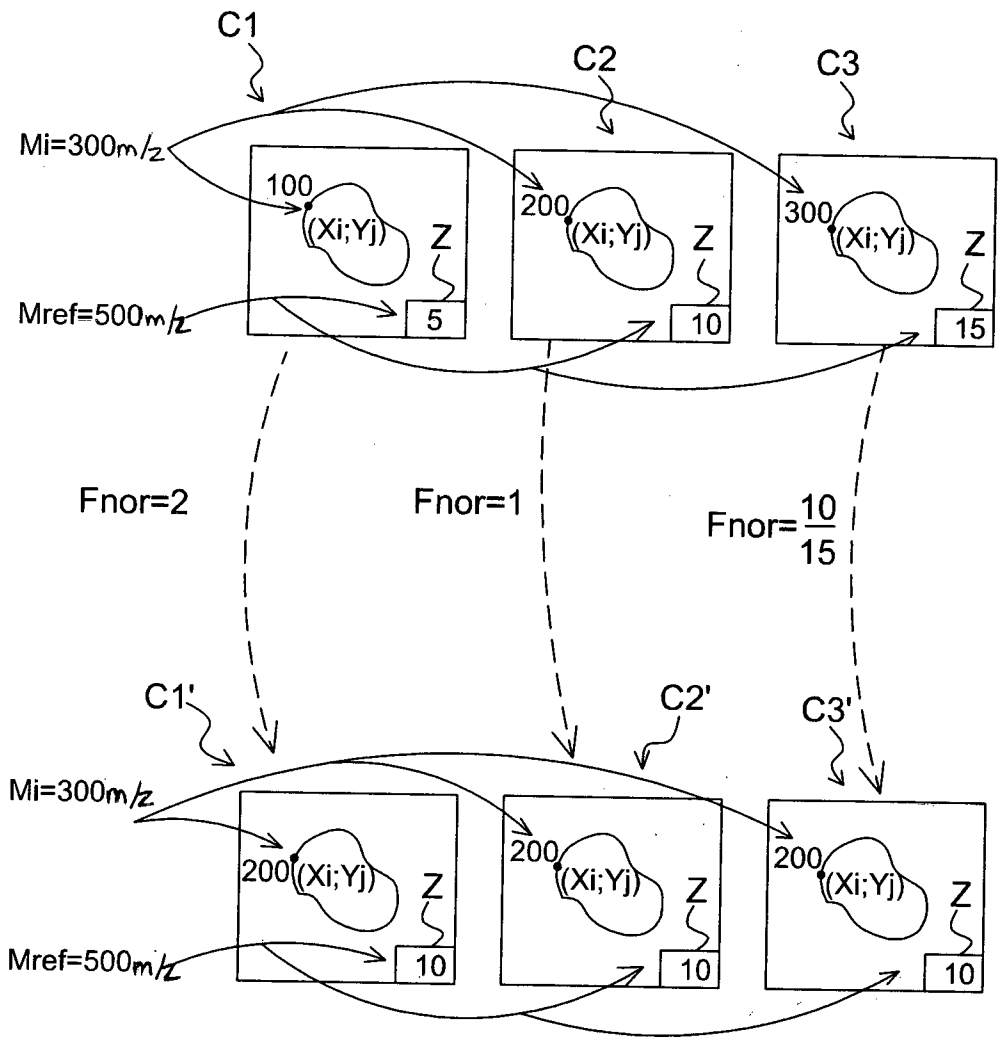
C1
C2
C3
Mi=300m/z
100
(Xi;Yj)
Z
5
200
(Xi;Yj)
Z
10
300
(Xi;Yj)
Z
15
Mref=500m/z
Fnor=2
Fnor=1
Fnor=10/15
C1'
C2'
C3'
Mi=300m/z
200
(Xi;Yj)
Z
10
200
(Xi;Yj)
Z
10
200
(Xi;Yj)
Z
10
Mref=500m/z

Fig.5

INTERNATIONAL SEARCH REPORT

International application No
PCT/FR2015/050051

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F19/26 G06F19/28
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2009/156897 A1 (KONINKL PHILIPS ELECTRONICS NV [NL]; ALSAFADI YASSER H [US]) 30 December 2009 (2009-12-30) abstract page 5, paragraph 2 - page 7, last paragraph claims 1-3,7 figures 1,2 -----	1-16

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 March 2015	31/03/2015

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Hilbig, Matthias

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/FR2015/050051

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2009156897 A1	30-12-2009	NONE	

RAPPORT DE RECHERCHE INTERNATIONALE

Demande internationale n°
PCT/FR2015/050051

A. CLASSEMENT DE L'OBJET DE LA DEMANDE
INV. G06F19/26 G06F19/28
ADD.

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)
G06F

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si cela est réalisable, termes de recherche utilisés)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie*	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	WO 2009/156897 A1 (KONINKL PHILIPS ELECTRONICS NV [NL]; ALSAFADI YASSER H [US]) 30 décembre 2009 (2009-12-30) abrégé page 5, alinéa 2 - page 7, dernier alinéa revendications 1-3,7 figures 1,2 -----	1-16

☐ Voir la suite du cadre C pour la fin de la liste des documents

☒ Les documents de familles de brevets sont indiqués en annexe

* Catégories spéciales de documents cités:

"A" document définissant l'état général de la technique, non considéré comme particulièrement pertinent

"E" document antérieur, mais publié à la date de dépôt international ou après cette date

"L" document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)

"O" document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens

"P" document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

"T" document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention

"X" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément

"Y" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier

"&" document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée	Date d'expédition du présent rapport de recherche internationale
20 mars 2015	31/03/2015

Nom et adresse postale de l'administration chargée de la recherche internationale	Fonctionnaire autorisé
Office Européen des Brevets, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Hilbig, Matthias

1

RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs aux membres de familles de brevets

Demande internationale n°
PCT/FR2015/050051

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
WO 2009156897 A1	30-12-2009	AUCUN	