(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 December 2004 (09.12.2004)

PCT

(10) International Publication Number
WO 2004/106915 A1

(51) International Patent Classification[7]: G01N 30/86

(21) International Application Number:
PCT/US2004/016797

(22) International Filing Date: 26 May 2004 (26.05.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/474,499 29 May 2003 (29.05.2003) US

(71) Applicant *(for all designated States except US)*: **WATERS INVESTMENTS LIMITED** [US/US]; 109 Lukens Drive, New Castle, DE 19720 (US).

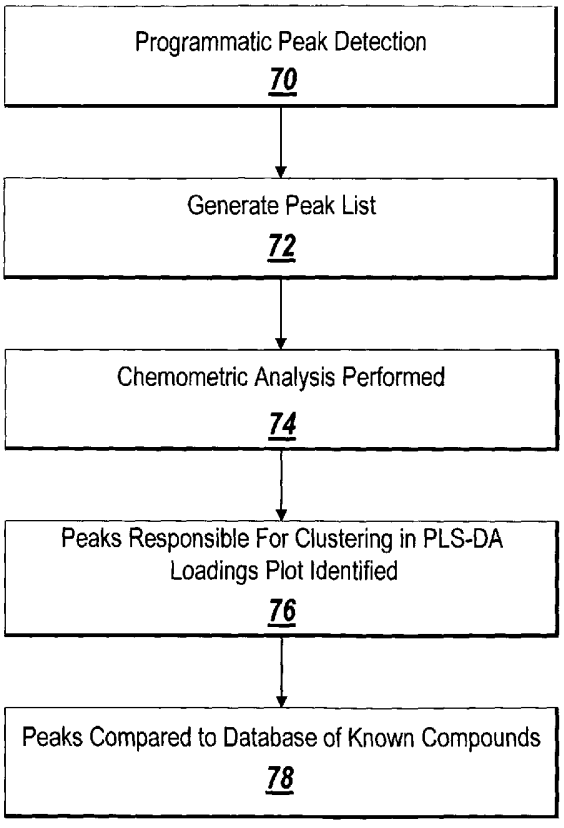
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **CASTRO-PEREZ, Jose** [ES/GB]; 16 Doverston Road, Sale, Manchester M33 6LE (GB). **PLUMB, Robert** [GB/US]; 6 McGill Lane, Milford, MA 01757 (US). **MAJOR, Hilary** [GB/GB]; 2 Adey Road, Lymm, Chesire WA13 9QX (GB). **PREECE, Steve** [GB/GB]; 18 Burnt Acre, Chelford, Macclesfield, Cheshire SK11 9SS (GB). **LUNT, Martin** [GB/GB]; 128 Redwood Firwood Park, Lancashire, Oldham OL9 9 UG (GB). **GOSHAWK, Jeff** [GB/GB]; 18 28 Mellor Brow, Mellor, Blackburn BB2 7EX (GB). **GORENSTEIN, Marc, V.** [US/US]; 313 Brookline Street, Needham, MA 02492 (US).

(74) Agents: **JANIUK, Anthony, J.** et al.; 34 Maple Street, Mailstop: LG, Milford, MA 01757 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD,

[Continued on next page]

(54) Title: METHOD AND DEVICE FOR PROCESSING OF LC-MS OR LC-MS/MS DATA IN METABONOMICS



(57) Abstract: A method of reducing a set of collected LC-MS or LC-MS/MS data such that true chromatographic and MS peaks are identified for use in Metabonomics is disclosed. The identified peaks are used to create a list of LC/MS, GC/MS, DIOS-MS or MALDI-MS signals and responses for a batch of samples which appear in a Master Entity List. The samples in the Master Entity List. The samples in the Master Entity List are then subjected to isotope de-clustering and adduct removal prior to chemometrics being applied to automatically identify biomarkers. An LC-MS/MS or LC/MS, GC/MS, DIOS-MS or MALDI-MS acquisition list is generated for the signals identified as responsible for the PLS-DA or PCA separation. The LC or GC retention time, exact mass and MS/MS spectrum may be compared to databases of known compounds and identified compounds associated with biological parameters may be stored in a new compound database.

WO 2004/106915 A1

MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHOD AND DEVICE FOR PROCESSING OF LC-MS OR LC-MS/MS DATA IN METABONOMICS

Related Application

This application claims the benefit of a United States provisional application entitled "*A System and Method for Metabonomics Directed Processing of LC-MS or LC-MS/MS Data*", Serial No. 60/474, 499, filed on May 29, 2003.

Field of the Invention

The illustrative embodiment of the present invention relates generally to metabolic analysis and more particularly to the programmatic processing of LC-MS and LC-MS/MS data for peak deconvolution and subsequent chemometric analysis.

Background

Metabolism may be defined as the chemical changes that take place in a cell or organisms that are used to produce energy and the basic materials which are needed for important life processes such as mitosis. The byproducts of the chemical reaction may be referred to as metabolites. By analyzing and identifying the metabolites that are present in a sample, it is possible to determine the route of metabolism. For example, an analysis of metabolites in biofluids such as urine may be used to determine what substances were ingested by the individual that produced the urine. The identification and analysis of the metabolites is often performed using liquid chromatography in

combination with mass spectrometry. The profiling of complex metabolic patterns in biofluids is referred to as metabonomics.

Liquid chromatography separates the individual components contained within a sample so that they may be identified. In liquid chromatography two phases are involved, a mobile phase and a stationary phase. A liquid sample mixture (the "mobile phase") is passed through a column packed with particles (the "solid phase") in order to effect a separation of the constituent components. The particles in the column may or may not be coated with a liquid designed to interact with the mobile phase. The constituent components in the mobile phase (i.e.: in the sample) pass through the packed column at different rates based upon a number of factors. The separation of the sample into its constituent components is then analyzed by observing the sample as it exits the far end of the column.

The speed with which the different constituent components pass through the column depends on the interaction of the mobile phase with the solid phase. The components in the sample may physically interact with the particles or a substance coating the particles such that their movement through the column is retarded. Different components in the sample being analyzed will react differently to the particular particle and/or coating by interacting with the particular particles and/or coating with differing degrees of strength depending upon the chemical makeup of the component. Those components which have a greater affinity for the particles and/or coating will pass through the column more slowly than those components which bond weakly or not at all with the particle/coating. In addition to chemical reactions, the size of the components in the sample may dictate the speed with which they pass through the column. For

example, in gel-permeation chromatography, different molecules in the solution being analyzed pass through a matrix containing pores at different speeds thereby effecting a separation of the different molecules in the sample. In size exclusion chromatography the size of the particles and their packing method in the column combine with the size of the components in the sample to determine the rate at which a sample passes through the column(as only certain size components may easily traverse the gaps/interstitial spaces between particles).

The separated sample travels into a detector at the far end of the column where the retention time is calculated for the various components in the sample. The retention time is the time required for the sample to travel from the injection port (where the sample is introduced into the column) through the column and to the detector. The amount of the component exiting the solid phase may be graphed against the retention time to form a chart with peaks which are known as chromatographic peaks. The peaks identify the different components.

The separated components may be fed into a mass spectrometer for further analysis in order to determine their chemical make-up. Systems that have one mass spectrometer stage combined with a liquid chromatography stage are referred to as LC-MS systems. Systems with two mass spectrometer stages are referred to as LC-MS/MS systems. A mass spectrometer takes a sample as input and ionizes the sample to create either positive or negative ions. A number of different ionization methods may be used including the use of an electrospray ionization. The ions are then separated by the mass to charge ratio in a first stage separation commonly referred to as MS1. The mass separation may be accomplished by a number of means including the use of magnets

which divert the ions to differing degrees based upon the weight of the ions. The separated ions then travel into a collision cell where they come in contact with a collision gas or other substance which interacts with the ions. The reacted ions then undergo a second stage of mass separation commonly referred to as MS2.

The separated ions are analyzed at the end of the mass spectrometry stage(or stages). The analysis graphs the intensity of the signal of the ions versus the mass of the ion in a graph referred to as a mass spectrum. The analysis of the mass spectrum gives both the masses of the ions reaching the detector and the relative abundances. The abundances are obtained from the intensity of the signal. The combination of liquid chromatography with mass spectrometry may be used to identify chemical substances such as metabolites. When a molecule collides with the collision gas covalent bonds often break, resulting in an array of charged fragments. The mass spectrometer measures the masses of the fragments which may then be analyzed to determine the structure and/or composition of the original molecule. This feature is significantly enhanced from nominal mass MS when using a mass spectrometer capable of accurate mass measurements e.g. hybrid quadrupole orthoganol TOF instrument or FTICR, allowing analyte elemental composition information to be derived. This information may be used to isolate a particular substance in a sample.

Chemometrics is the mathematical treatment of data such as LC-MS/MS data and includes types of multi-variate analysis such as PCA (Principle Component Analysis) and PLS-DA(Partial Least Squares-Discriminate Analysis) or similar statistical approaches. Chemometrics attempts to reduce large amounts of data to a manageable size and apply a statistically driven model in order to determine latent

variables indicative of hidden relationships between the observed data. Chemometrics may thus be applied to the field of metabonomics. Unfortunately, conventional methods of data acquisition often lose valuable relevant data in the process of reducing the collected data set as the processing/collecting of MS data for chemometric analysis is reliant upon the summing of the whole MS spectrum and thus results in the loss of any retention time data. Additionally, conventional methods do not integrate the raw data, filtered data and statistical analysis into a single data processing application with the result that the mapping of the raw data to filtered data to analyzed data is awkward at best.

Summary of the Invention

The illustrative embodiment of the present invention provides an automated mechanism for rapidly reducing the set of collected LC/MS or LC-MS/MS data such that true chromatographic and MS peaks are identified. The identified peaks are used to create a list of LC/MS signals and responses for a batch of samples which appear in a Master Entity List. The samples in the Master Entity List can then subjected to isotope de-clustering and adduct removal prior to chemometrics being applied to automatically identify biomarkers. An LC-MS/MS acquisition list is generated for the signals identified as responsible for the PLS-DA or PCA group clustering or separation. The LC retention time, accurate mass and MS/MS spectrum may be compared to databases of known compounds and identified compounds associated with biological parameters may be stored in a new compound database.

Brief Description of the Drawings

Figure 1 depicts an environment suitable for practicing the illustrative embodiment of the present invention;

Figure 2 is a flow chart of the sequence of steps used to perform liquid chromatography and mass spectrometry;

Figure 3 depicts a visual display of a Sample List generated by the illustrative embodiment of the present invention;

Figure 4 depicts a visual display of a Master Entity List generated by the illustrative embodiment of the present invention

Figure 5A depicts a visual display of the loadings plot markers graph generated by the illustrative embodiment of the present invention;

Figure 5B depicts a visual display of a trends plot graph generated by the illustrative embodiment of the present invention;

Figure 6 depicts a visual display of the scores plot graph showing group similarities generated by the illustrative embodiment of the present invention

Figure 7 is a flow chart of the overall sequence of steps followed by the illustrative embodiment of the present invention to perform metabonomics-directed processing of LC-MS/MS data; and

Figure 8 is a flow chart of the sequence of steps followed by the illustrative embodiment of the present invention to perform chemometric analysis.

Detailed Description

The illustrative embodiment of the present invention provides a mechanism for using chemometric analysis on programmatically filtered LC-MS or LC-MS/MS data for the purpose of determining metabonomic profiles. Collected LC-MS or LC-MS/MS data is programmatically filtered to determine true chromatographic and MS peaks. A Master Entity List is created from the LC-MS or LC-MS/MS signals and responses for a batch of samples. The samples in the Master Entity List are further filtered and chemometrics are applied to automatically identify metabonomic biomarkers.

Data for the illustrative embodiment of the present invention is performed in a metabolite analyzing system such as an LC-MS/MS system as depicted in **Figure 1**. Other types of metabolic analyzing systems such as LC/MS systems may be used instead of an LC-MS/MS system without departing from the scope of the present invention. Those skilled in the art will recognize that this approach could also be applied to the analysis of LC-UV or other similar hyphenated chromatographic techniques such as GC-MS and DIOS-MS as well as MALDI-MS (Matrix Assisted Laser Desorption/Ionization- Mass Spectroscopy)-MS and DIOS-MS. The metabolite analyzing system 2 includes a chromatography module 4, such as a liquid chromatography module. Also included is an ionization module 10. The ionization module 10 receives as an input sample the output from the chromatography module 4. The ionization module performs ionization of the sample. Those skilled in the art will recognize that there are a number of different ways in which the sample may be ionized, such as by bombarding the sample with a stream of high energy electrons.

The ions produced by the ionization module 10 are passed on to the MS1 first stage mass separation module 12. The mass separation may be performed using any of a number of well-known techniques. For example, the ions may be subjected to magnetic forces which alter the path of the ions based upon the mass of the ion. The separated ions are then be passed into a collision cell module 14 where they are subjected to additional reactions, such as exposure of the ions to a gas designed to react with the separated ions. The sample may be further separated in an MS2 second stage mass separation module 16 prior to arriving at a detector module 18. The detector module 18 is used to generate a mass spectrum based on the detected signal generated by the exiting ions. Those skilled in the art will recognize that a number of different methods of mass separation may be used and different substances may be introduced into the collision cell 14 in order to react with the ions of particular interest. Similarly, the illustrative embodiment of the present invention may also be performed with a number of different metabolite analyzing systems including an
LC-MS system performing only one stage of mass separation.

An electronic device with a processor 6 is interfaced with the detector module 18 and the chromatography module 4. The electronic device 6 may be a server, desktop computer system, laptop, mainframe, network attached device or some other similar device with a processor. The electronic device may also be integrated into one of the modules in the metabolite analyzing system 2 without departing from the scope of the present invention. The electronic device 6 includes storage 8 which is used to record the results of sample runs. Those skilled in the art will recognize that the storage 8 may be located in any location accessible to the metabolite analyzing system 2. Also located on the electronic device 6 is a Toxicological Screening and Biomarker Identification

application 20 that may be used to identify biomarkers for different types of Systems Biology such as Metabonomics, Functional Genomics, Peptidomics, Lipidomics, Glycomics and Proteomics. Those skilled in the art will recognize that this approach could also be used for natural product evaluation, impurity profiling, environmental analysis, food and nutrition and product release. The Toxicological Screening and Biomarker Identification Application 20 is discussed further below. Those skilled in the art will recognize that the Toxicological Screening and Biomarker Application 20 may be located in any location in which it can access the saved raw LC-MS or LC-MS/MS data, including being integrated into the modules of the metabolite analyzing system 2 or on a separate electronic device.

The sequence of steps performed to conduct a single LC-MS or LC-MS/MS run to collect raw data is depicted in the flow chart of **Figure 2**. The sequence begins with a liquid chromatography separation of the components in a sample (step 30). The sample components exiting from the liquid chromatography system are passed into the ionization module 10 where ionization is performed (step 32). The first stage of mass separation is performed (step 34) and the separated ions are passed into the collision cell where they react to the collision cell reactant (step 36). Second stage mass separation is then performed on the reacted ions exiting from the collision cell (step 38). The separated ions are passed into the detector module 18 where a mass spectrum is generated from collected data thereby enabling the identification of metabolites contained within the sample (step 40).

Once the raw LC-MS or LC-MS/MS data has been collected, the illustrative embodiment of the present invention works to identify true chromatographic and MS peaks. The Toxicological Screening and Biomarker Identification Application 20 performs peak deconvolution on the raw LC and MS data. Peak deconvolution identifies the actual analyte signal peaks and filters out noise from the raw LC and MS data. The Toxicological Screening and Biomarker Identification Application 20 next creates a sample list of signals. **Figure 3** depicts a Sample List 50 of signals. The Sample List 50 is used to create a batch of samples from the signals and responses that appear in a Master Entity List 60. **Figure 4** depicts a display of a Master Entity List 60 that is generated by the illustrative embodiment of the present invention. Each sample in the Master Entity List 60 includes an ID 61, a Retention Time 62, a Mass 63, a Significance 64, an Exclusion value 65, and ion intensity/ response value columns 66. As an example, each response value column may be for a separate test animal. The Master Entity List lists all of the similarities of two different groups and may exclude certain masses. Every true peak or analyte detected by the system in each sample is cross-referenced with each of the other samples programmatically. Samples missing a signal are assigned a value.

The Toxicological Screening and Biomarker Identification Application 20 then further filters the sample data. The samples undergo isotope de-clustering and adduct removal to remove unwanted trace elements. Adduct removal refers to the removal of ion such as sodium and potassium or dimmer/trimers etc which if unaccounted for can skew the analysis of the collected data.

Once the samples have undergone isotope de-clustering and adduct removal, the Toxicological Screening and Biomarker Identification Application 20 uses chemometric analysis to identify potential biomarkers in the sample data. The chemometric analysis will identify clusters of interest among the samples. The clusters represent similarities among the samples and are used to identify the metabonomic profiles. A number of different types of chemometric analysis may be used including PCA and PLS-DA.

For example, Principal Component Analysis (PCA) uses mathematical algorithms to determine the differences and similarities in a data set. PCA transforms a number of possibly related variables into a smaller number of unrelated variables which are referred to as principle components. The first principle component accounts for as much of the variability in the data as possible. Each additional component attempts to account for as much of the remaining variability in the data as possible. The collected data may be arranged in a matrix and PCA solves for eigenvalues and eigenvectors of a square symmetric matrix with sums of squares and cross products. The eigenvector associated with the largest eigenvalue has the same direction as the first principle component. The eigennvector associated with the second greatest eigenvalue determines the direction of the second principle component. The sum of the eigenvalues equals the trace of the square matrix and the maximum number of eigenvectors equals the numbers of rows (or columns) of this matrix. Once determined, it is possible to draw screen plots of the calculated eigenvalues. Those skilled in the art will recognize that a number of different algorithms may be used to calculate the eigenvalues and eigenvectors. The data is displayed using two plots: i) the scores plot which shows the group clustering and ii) the loadings plot in which the analytes/ions responsible for the group clustering are identified as those being the greatest distance from the origin.

Chemometric analysis is used to determine latent variables which represent hidden connections between data points. Each data sample has a number of features such as signal intensity, mass and retention time. The chemometric analysis applies a function to the features and graphs the result of the function on an n dimensional plot. Conventional methods of processing the data for plotting involve bucketing data from time intervals of the sample run. This results in the loss of the retention time variable. The illustrative embodiment of the present invention presents a Loadings Plot 70 as shown in **Figure 5A** showing the analytes peak of the various markers. The ions the greatest distance from the origin, using eigen vectors, are those most responsible for the group clustering or separation. The Loading Plot 70 may also be used to create a Trends plot showing the correlation between signal intensity and sample dose. Figure 5B shows a trends plot 73 for a selected ion. A Scores Plot 75 as shown in **Figure 6** indicates the similarities between samples(such as a control sample and a dosed animal sample). The data in Figure 6 shows the PCA of LC/MS data generated from rat urine obtained following the administration of vehicle alone or a candidate pharmaceutical at low and high dose. The display generated for a user visually indicates obvious points of similarity which are ascertainable with the naked eye. These suggestions of similarity may then form the basis for further study.

Figure 7 is a flow chart of the overall sequence of steps followed by the illustrative embodiment of the present invention to perform metabonomics-directed processing of LC-MS or LC-MS/MS data. The sequence begins with the collection of raw LC-MS/MS data and the identification of actual LC and MS peaks as described above (step 70). A list of the identified peaks is then generated (step 72). This list

forms the basis for the samples appearing in the Master Entity List. The samples are further filtered, undergoing isotope de-clustering and adduct removal, and chemometric analysis, such as PCA and PLS-DA analysis, is performed (step 74). Peaks responsible for clustering in PLS-DA loadings are identified (step 76). The peaks are then compared to a database of known endogenous biochemicals in order to identify the compound associated with the peak (step 78). The compounds may be toxic compounds, drugs, chemicals, agricultural chemicals or other compounds. A compound database may be generated from the identified compounds containing retention times, m/z values with accurate mass where appropriate and other biological parameters such as sex, dose levels, day, and toxin. Those skilled in the art will recognize that identified biomarkers may be used in a number of different areas of science such as Metabolomics, Functional Genomics, Peptidomics and Proteomics.

The chemometric analysis performed by the illustrative embodiment of the present invention is further shown in **Figure 8.** The sequence of chemometric analysis steps begins when the LC-MS data is reduced to the samples of the Master Entity List (step 80). Xenobiotics are then removed from the samples to leave only the endogenous metabolites(step 82). PCA and PLS-DA analysis is then carried out on the data batch (step 84). Signals from the PLS-DA plot furthest away from the clusters are removed until there is no separation between groups (step 86).

A user of the Toxicological Screening and Biomarker Identifier Application 20 may thus easily transition between raw data, filtered data and analyzed data all by selecting the appropriate view. Conventional software packages lack this integration between the raw and filtered data and the analyzed data since two or more separate

software packages are required for the task. The requirement of two or more software packages presents a user with difficulty in mapping from analyzed data to the corresponding spot in the raw data.

It will thus be seen that the invention attains the objectives stated in the previous description. Since certain changes may be made without departing from the scope of the present invention, it is intended that all matter contained in the above description or shown in the accompanying drawings be interpreted as illustrative and not in a literal sense. Practitioners of the art will realize that the sequence of steps and architectures depicted in the figures may be altered without departing from the scope of the present invention and that the illustrations contained herein are singular examples of a multitude of possible depictions of the present invention.

We Claim:

1. In a metabolite analysis system, a method, comprising the steps of:

programmatically identifying chromatography peaks and mass spectrometry peaks from a sample run; said mass spectrometry peak being one of an MS peak and MS/MS peak and using nominal or exact mass;

generating a list of sample data having said identified peaks;

performing chemometric analysis on said sample data to identify biomarkers; said chemometric analysis performed without loss of retention time data by the same application performing the programmatic identification of said chromatography and mass spectrometry peaks.

2. The method of claim 1 wherein said chemometric analysis is performed using one of Principle Component Analysis (PCA) and Partial Least Squares Discriminate Analysis (PLS-DA).

3. The method of claim 1, comprising the further steps of:

comparing said identified biomarkers with a database of known compounds.

4. The method of claim 1, comprising the further step of:

removing unwanted material traces from the sample data prior to performing chemometric analysis.

5. The method of claim 1 wherein said unwanted material traces are at least one of xenobiotic traces, dosing vehicle traces, extraneous food traces, and contamination traces.

6. The method of claim 1 wherein said sample data includes mass data, retention time and signal intensity values.

7. The method of claim 1 wherein said biomarkers are used in Systems Biology.

8. The method of claim 1 wherein said chemometric analysis further comprises the steps of:

plotting said sample data on an n-dimensional plot, said n-dimensional plot indicating the analyte peaks of a plurality of said biomarkers.

9. A medium in a metabolite analysis system, said medium holding executable steps for a method, said method comprising the steps of:

programmatically identifying chromatography peaks and mass spectrometry peaks from a sample run; said mass spectrometry peak being one of an MS peak and MS/MS peak and using nominal or exact mass;

generating a list of sample data having said identified peaks;

performing chemometric analysis on said sample data to identify biomarkers; said chemometric analysis performed without loss of retention time data by the same application performing the programmatic identification of said chromatography and mass spectrometry peaks.

10. The medium of claim 9 wherein said chemometric analysis is performed using one of Principle Component Analysis (PCA) and Partial Least Squares Discriminate Analysis (PLS-DA).

11. The medium of claim 9, wherein said method comprises the further steps of:

comparing said identified biomarkers with a database of known compounds.

12. The medium of claim 9, wherein said method comprises the further step of:

removing unwanted material traces from the sample data prior to performing chemometric analysis.

13. The medium of claim 9 wherein said unwanted material traces are at least one of xenobiotic traces, dosing vehicle traces, extraneous food traces, and contamination traces.

14. The medium of claim 9 wherein said sample data includes mass data, retention time and signal intensity values.

15. The medium of claim 9 wherein said biomarkers are used in Systems Biology.

16. The medium of claim 9 wherein said chemometric analysis further comprises the steps of:

plotting said sample data on an n-dimensional plot, said n-dimensional plot indicating the analyte peaks of a plurality of said biomarkers.

17. A metabolite analysis system, comprising:

one of a chromatography-mass spectroscopy type system, MALDI-MS Matrix Assisted Laser Desorption/Ionization-Mass Spectroscopy) system, and DIOS-MS (Desorption Ionization On Silicon) System;

a toxicological screening and biomarker identification facility, said toxicological and biomarker identification facility programmatically identifying analyte peaks and mass spectroscopy peaks from at least one sample run performed on said one of a chromatography- mass spectroscopy type system, MALDI-MS system, and DIOS-MS system, said toxicological and biomarker identification facility further performing chemometric analysis on said sample data to identify biomarkers, said chemometric analysis performed without loss of retention time data.; and

a storage location accessible to said toxicological and biomarker identification facility holding a collection of raw and filtered data from said at least one sample run performed on said one of a chromatography- mass spectroscopy type system, MALDI-MS system, and DIOS-MS system.

18. The system of claim 17 wherein said toxicological and biomarker identification facility is implemented in software on an electronic device interfaced with said one of a chromatography- mass spectroscopy type system, MALDI-MS system, and DIOS-MS system.

19. The system of claim 17 wherein said collection of raw and filtered data includes mass data, retention time and signal intensity values.

20. The system of claim 17 wherein said biomarkers are used in at least one of Metabonomics, Proteomics, Functional Genomics, Lipidomics, Glycomics, Metabolomics and endogenous peptide profiling.

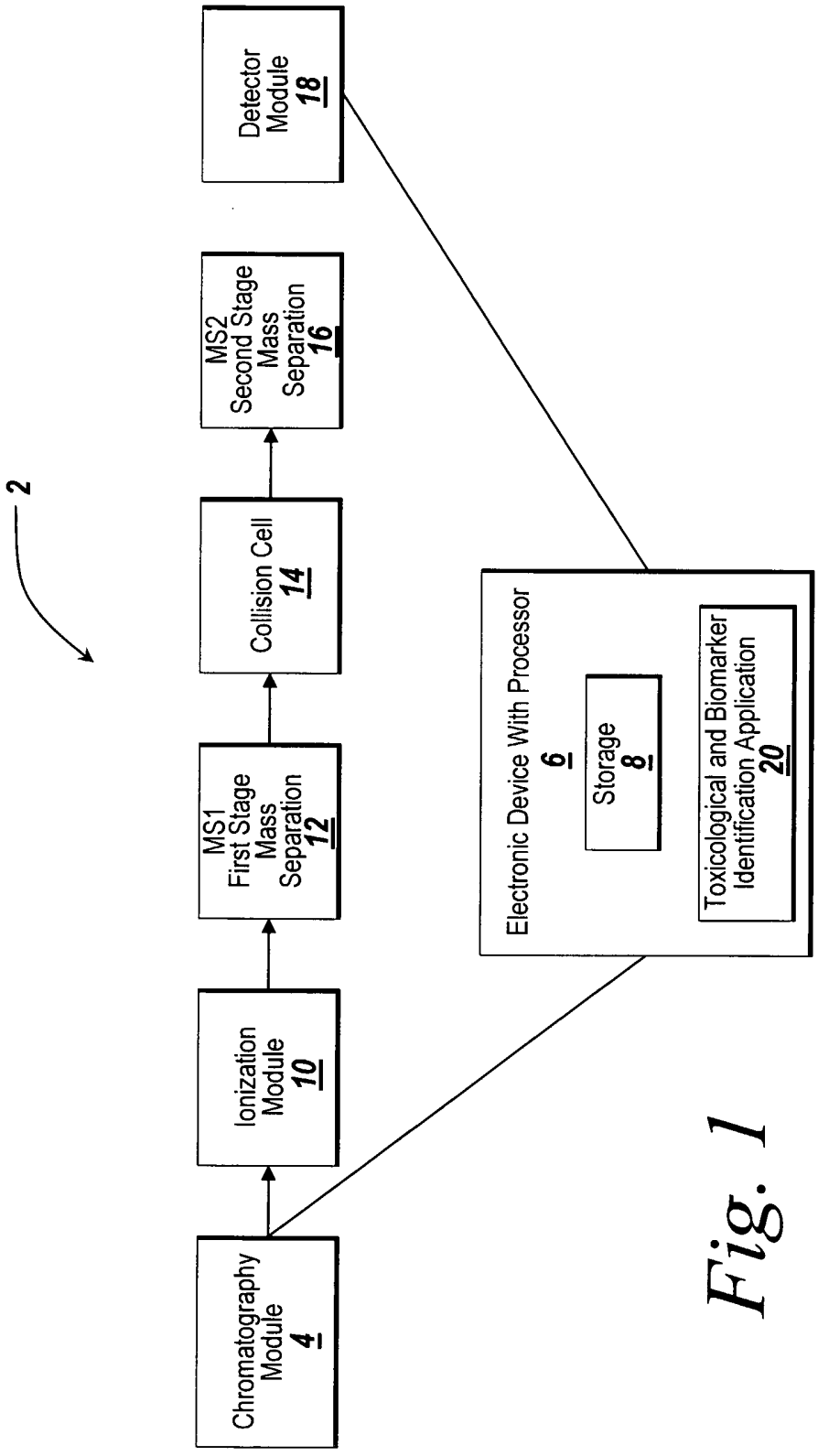

2
Chromatography Module 4
Ionization Module 10
MS1 First Stage Mass Separation 12
Collision Cell 14
MS2 Second Stage Mass Separation 16
Detector Module 18
Electronic Device With Processor 6
Storage 8
Toxicological and Biomarker Identification Application 20

Fig. 1

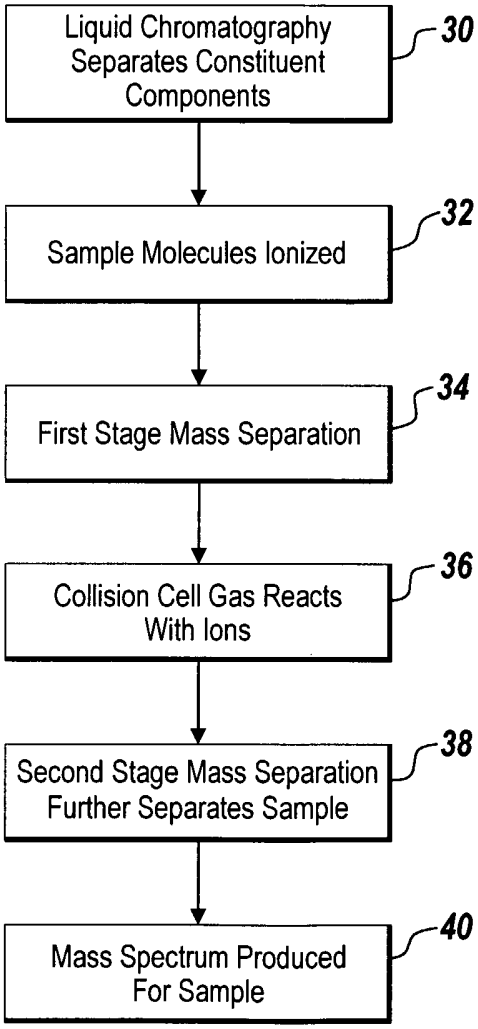

Liquid Chromatography Separates Constituent Components
30
Sample Molecules Ionized
32
First Stage Mass Separation
34
Collision Cell Gas Reacts With Ions
36
Second Stage Mass Separation Further Separates Sample
38
Mass Spectrum Produced For Sample
40

Fig. 2

Sample List

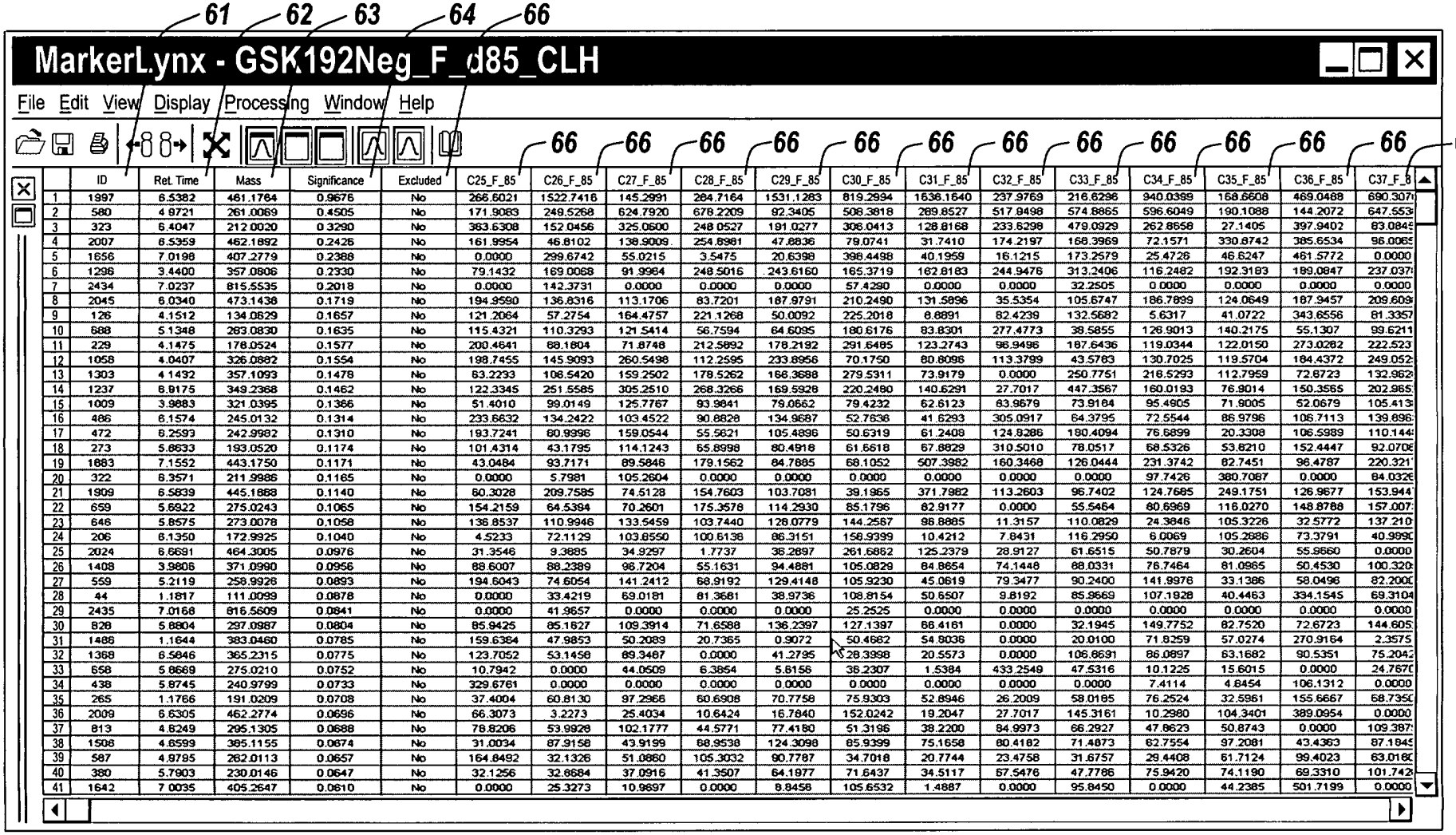
50

MarkerLynx - untitled *

File Edit View Display Processing Window Help

	Excluded	SAMPLEN...	DATAFILE	Internal Std.	ASFILE	INLETFILE	MSFILE	VIALREFE...	USER	SUBMITTER	SAMPLE_ID	TASK
1	No		1_1	0.0000		0_20_95_...	lcms_pos	2:2				
2	No		1_2	0.0000		0_20_95_...	lcms_pos	2:2				
3	No		1_3	0.0000		0_20_95_...	lcms_pos	2:2				
4	No		1_4	0.0000		0_20_95_...	lcms_pos	2:2				
5	No		1_5	0.0000		0_20_95_...	lcms_pos	2:2				
6	No		1_6	0.0000		0_20_95_...	lcms_pos	2:2				
7	No		1_7	0.0000		0_20_95_...	lcms_pos	2:2				
8	No		1_8	0.0000		0_20_95_...	lcms_pos	2:2				
9	No		1_9	0.0000		0_20_95_...	lcms_pos	2:2				
10	No		1_10	0.0000		0_20_95_...	lcms_pos	2:2				
11	No		H030404b_JG_ES+_ENO_041	0.0000		0_20_95_...	lcms_pos	2:5				
12	No		H030404b_JG_ES+_ENO_042	0.0000		0_20_95_...	lcms_pos	2:5				
13	No		H030404b_JG_ES+_ENO_043	0.0000		0_20_95_...	lcms_pos	2:5				
14	No		H030404b_JG_ES+_ENO_044	0.0000		0_20_95_...	lcms_pos	2:5				
15	No		H030404b_JG_ES+_ENO_045	0.0000		0_20_95_...	lcms_pos	2:5				
16	No		H030404b_JG_ES+_ENO_046	0.0000		0_20_95_...	lcms_pos	2:5				
17	No		H030404b_JG_ES+_ENO_047	0.0000		0_20_95_...	lcms_pos	2:5				
18	No		H030404b_JG_ES+_ENO_048	0.0000		0_20_95_...	lcms_pos	2:5				
19	No		H030404b_JG_ES+_ENO_049	0.0000		0_20_95_...	lcms_pos	2:5				
20	No		H030404b_JG_ES+_ENO_050	0.0000		0_20_95_...	lcms_pos	2:5				

Fig. 3



	ID	Ret. Time	Mass	Significance	Excluded	C25_F_85	C26_F_85	C27_F_85	C28_F_85	C29_F_85	C30_F_85	C31_F_85	C32_F_85	C33_F_85	C34_F_85	C35_F_85	C36_F_85	C37_F_8
1	1997	6.5382	461.1764	0.9676	No	266.6021	1522.7416	145.2991	264.7164	1531.1283	819.2994	1636.1640	237.9769	216.6296	940.0389	168.6608	469.0488	690.307
2	580	4.9721	261.0069	0.4505	No	171.9083	249.5268	624.7920	678.2209	92.3405	508.3818	289.8527	517.9498	574.8865	596.6049	190.1088	144.2072	647.553
3	323	6.4047	212.0020	0.3290	No	383.6308	152.0456	325.0600	248.0527	191.0277	306.0413	128.8168	233.6298	479.0929	262.8658	27.1405	397.9402	83.084
4	2007	6.5359	462.1892	0.2426	No	161.9954	46.8102	138.9009	254.8981	47.8836	79.0741	31.7410	174.2197	168.3969	72.1571	330.8742	385.6534	96.008
5	1656	7.0198	407.2779	0.2388	No	0.0000	299.6742	55.0215	3.5475	20.6398	398.4498	40.1959	16.1215	173.2579	25.4726	46.6247	461.5772	0.0000
6	1298	3.4400	357.0806	0.2330	No	79.1432	169.0068	91.9984	248.5016	243.6160	165.3719	162.8183	244.9476	313.2406	116.2482	192.3183	189.0847	237.037
7	2434	7.0237	815.5535	0.2018	No	0.0000	142.3731	0.0000	0.0000	0.0000	57.4280	0.0000	0.0000	32.2505	0.0000	0.0000	0.0000	0.0000
8	2045	6.0340	473.1436	0.1719	No	194.9590	136.8316	113.1706	83.7201	187.9791	210.2490	131.5896	35.5354	105.6747	186.7699	124.0649	187.9457	209.609
9	126	4.1512	134.0629	0.1657	No	121.2064	57.2754	164.4757	221.1268	50.0092	225.2018	8.6891	82.4239	132.5682	5.6317	41.0722	343.6556	81.3357
10	688	5.1348	283.0830	0.1635	No	115.4321	110.3293	121.5414	56.7594	64.6095	180.6176	83.8301	277.4773	38.5855	126.9013	140.2175	55.1307	99.6211
11	229	4.1475	176.0524	0.1577	No	200.4641	68.1804	71.8748	212.5892	176.2192	291.6485	123.2743	96.9496	187.6436	119.0344	122.0150	273.0282	222.523
12	1058	4.0407	326.0882	0.1554	No	198.7455	145.9093	260.5498	112.2595	233.8956	70.1750	80.8096	113.3799	43.5763	130.7025	119.5704	184.4372	249.052
13	1303	4.1432	357.1093	0.1478	No	63.2293	108.5420	159.2502	178.5262	166.3688	279.5311	73.9179	0.0000	250.7751	216.5293	112.7959	72.8723	132.962
14	1237	6.9175	349.2368	0.1462	No	122.3345	251.5585	305.2510	268.3266	169.5928	220.2480	140.6291	27.7017	447.3567	160.0193	76.9014	150.3585	202.965
15	1009	3.9883	321.0395	0.1366	No	51.4010	99.0149	125.7767	93.9841	79.0662	79.4232	62.6123	83.9679	73.9184	95.4905	71.9005	52.0679	105.413
16	486	6.1574	245.0132	0.1314	No	233.6632	134.2422	103.4522	90.8828	134.9687	52.7636	41.6293	305.0917	64.3795	72.5544	86.9796	106.7113	139.696
17	472	6.2593	242.9982	0.1310	No	193.7241	60.9996	159.0544	55.5621	105.4896	50.6319	61.2408	124.8286	180.4094	76.6899	20.3308	106.5989	110.144
18	273	5.8633	193.0520	0.1174	No	101.4314	43.1795	114.1243	65.8998	80.4918	61.6618	67.8829	310.5010	78.0517	68.5326	53.6210	152.4447	92.070
19	1883	7.1552	443.1750	0.1171	No	43.0484	93.7171	89.5846	179.1562	84.7885	68.1052	507.3982	160.3468	126.0444	231.3742	82.7451	96.4787	220.321
20	322	8.3571	211.9986	0.1165	No	0.0000	5.7981	105.2604	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	97.7426	380.7087	0.0000	84.0326
21	1909	6.5839	445.1888	0.1140	No	80.3028	209.7585	74.5128	154.7603	103.7081	39.1965	371.7982	113.2603	96.7402	124.7685	249.1751	126.9677	153.944
22	659	5.6922	275.0243	0.1065	No	154.2159	64.5394	70.2601	175.3578	114.2930	85.1796	62.9177	0.0000	55.5464	60.6969	116.0270	148.8788	157.007
23	646	5.8575	273.0078	0.1058	No	136.8537	110.9946	133.5459	103.7440	128.0779	144.2567	96.8885	11.3157	110.0829	24.3846	105.3226	32.5772	137.210
24	206	6.1350	172.9925	0.1040	No	4.5233	72.1129	103.6550	100.6136	86.3151	158.9399	10.4212	7.8431	116.2950	6.0069	105.2686	73.3791	40.999
25	2024	6.6691	464.3005	0.0976	No	31.3546	9.3685	34.9297	1.7737	36.2697	261.6862	125.2379	28.9127	61.6515	50.7879	30.2604	55.9660	0.0000
26	1408	3.9806	371.0990	0.0956	No	88.6007	88.2389	96.7204	55.1631	94.4881	105.0829	84.8654	74.1448	88.0331	76.7464	81.0965	50.4530	100.329
27	559	5.2119	258.9926	0.0893	No	194.6043	74.6054	141.2412	68.9192	129.4148	105.9230	45.0619	79.3477	90.2400	141.9976	33.1386	58.0496	82.200
28	44	1.1817	111.0099	0.0878	No	0.0000	33.4219	69.0181	81.3681	38.9736	108.8154	50.6507	9.8192	85.9669	107.1928	40.4463	334.1545	69.310
29	2435	7.0168	816.5609	0.0841	No	0.0000	41.9657	0.0000	0.0000	0.0000	25.2525	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
30	826	5.8804	297.0987	0.0804	No	85.9425	85.1627	109.3914	71.6588	136.2397	127.1397	66.4161	0.0000	32.1945	149.7752	82.7520	72.6723	144.605
31	1486	1.1644	383.0460	0.0785	No	159.6384	47.9853	50.2089	20.7365	0.9072	50.4682	54.8036	0.0000	20.0100	71.8259	57.0274	270.9164	2.3575
32	1368	6.5846	365.2315	0.0775	No	123.7052	53.1456	89.3487	0.0000	41.2795	28.3998	20.5573	0.0000	106.8691	86.0897	63.1682	90.5351	75.204
33	658	5.8669	275.0210	0.0752	No	10.7942	0.0000	44.0509	6.3854	5.8156	36.2307	1.5384	433.2549	47.5316	10.1225	15.6015	0.0000	24.767
34	438	5.8745	240.9799	0.0733	No	329.6761	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	7.4114	4.8454	106.1312	0.0000
35	265	1.1766	191.0209	0.0708	No	37.4004	60.8130	97.2966	60.6908	70.7758	75.9303	52.8946	26.2009	58.0195	76.2524	32.5961	155.6667	68.735
36	2009	6.6305	462.2774	0.0696	No	66.3073	3.2273	25.4034	10.6424	16.7840	152.0242	19.2047	27.7017	145.3161	10.2980	104.3401	389.0954	0.0000
37	813	4.8249	295.1305	0.0688	No	78.8206	53.9926	102.1777	44.5771	77.4180	51.3196	38.2200	84.9973	66.2927	47.8623	50.8743	0.0000	109.387
38	1508	4.6599	385.1155	0.0674	No	31.0034	87.9158	43.9199	68.9538	124.3098	85.9399	75.1658	80.4182	71.4873	62.7554	97.2081	43.4363	87.184
39	587	4.9795	262.0113	0.0657	No	164.8492	32.1326	51.0860	105.3032	90.7787	34.7018	20.7744	23.4758	31.6757	29.4408	61.7124	99.4023	63.016
40	380	5.7903	230.0146	0.0647	No	32.1256	32.6684	37.0916	41.3507	64.1977	71.6437	34.5117	67.5476	47.7786	75.9420	74.1190	69.3310	101.742
41	1642	7.0035	405.2647	0.0610	No	0.0000	25.3273	10.9697	0.0000	8.8458	105.6532	1.4887	0.0000	95.8450	0.0000	44.2385	501.7199	0.0000

Fig. 4

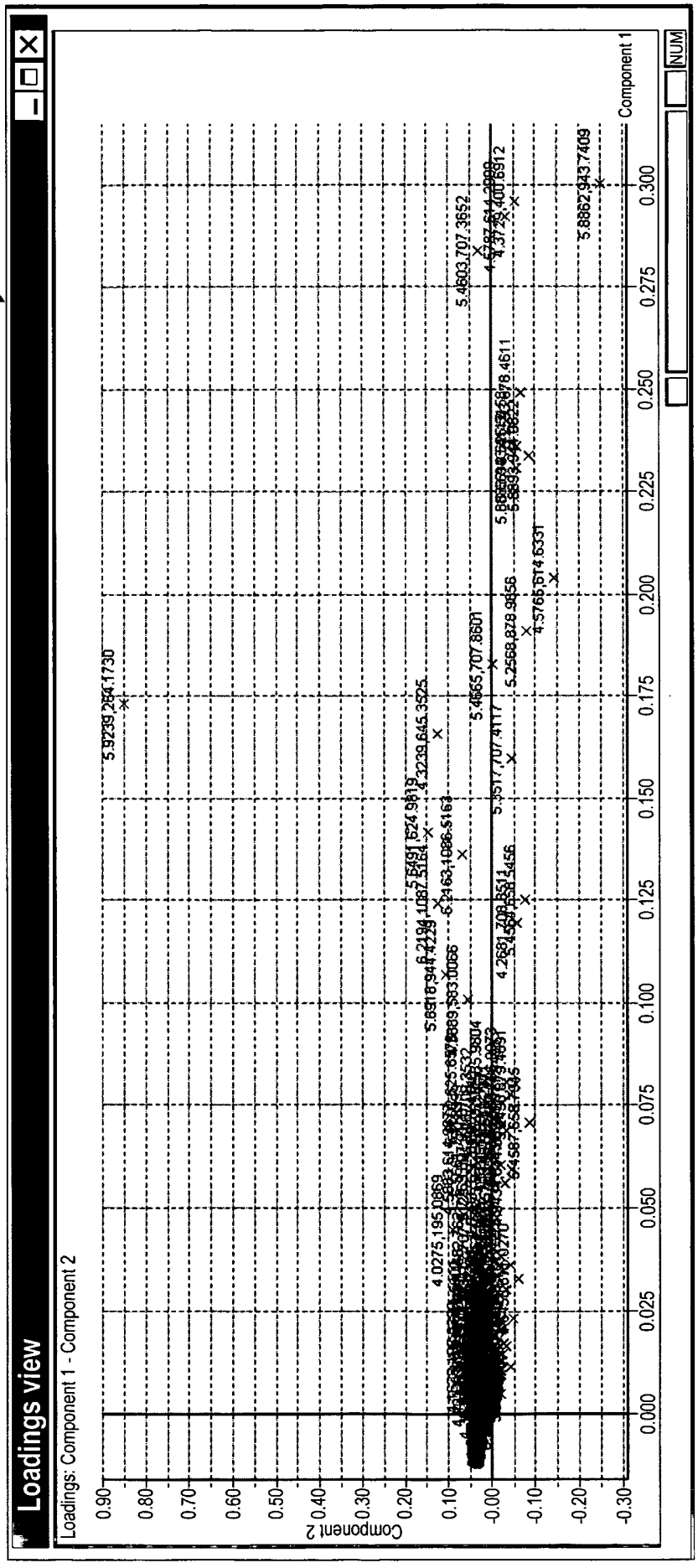

Loadings Plot: Markers
70
Loadings view
Loadings: Component 1 - Component 2
Component 2
Component 1
NUM
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
-0.00
-0.10
-0.20
-0.30
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200
0.225
0.250
0.275
0.300

Fig. 5A

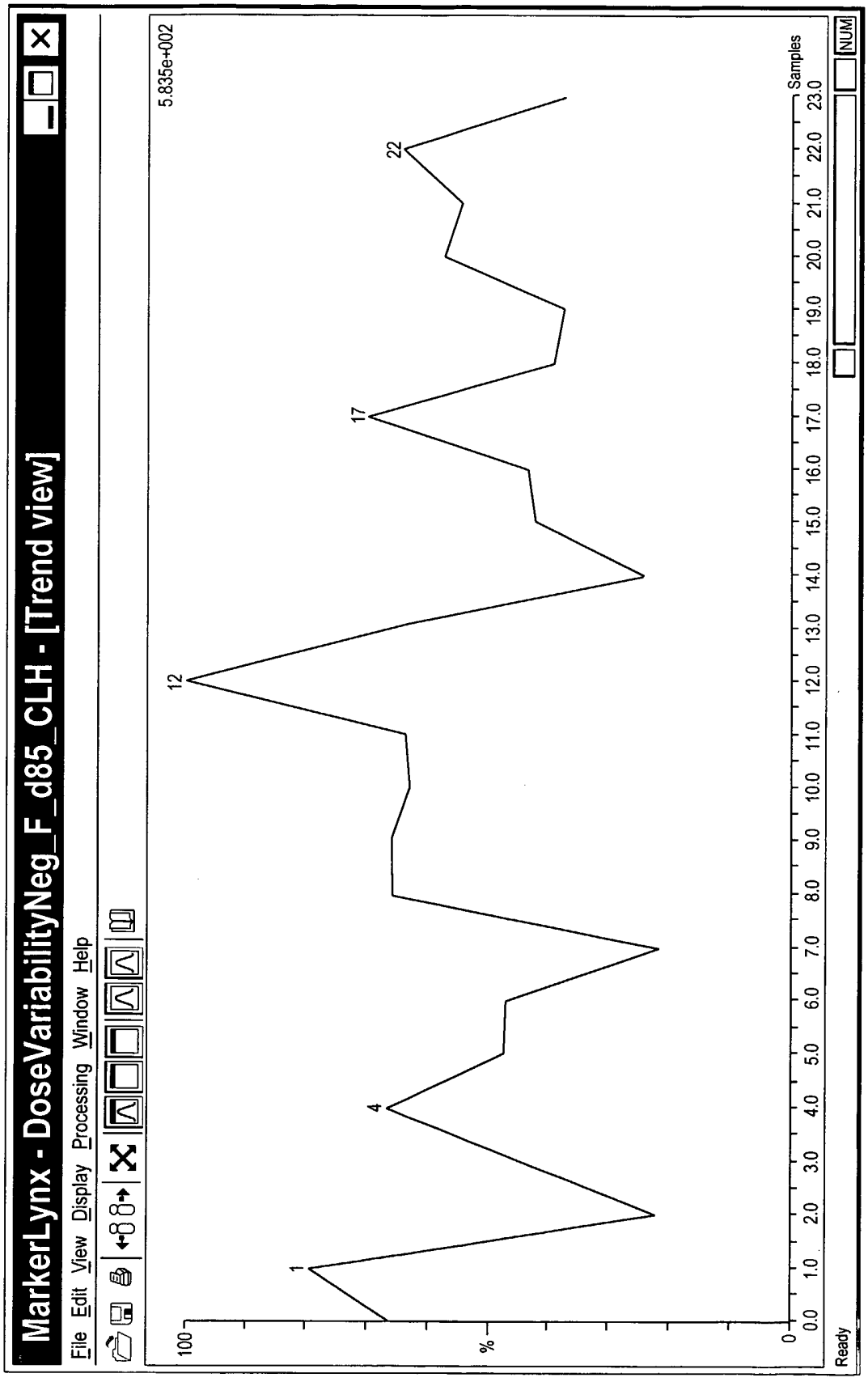

MarkerLynx - DoseVariabilityNeg_F_d85_CLH - [Trend view]
File Edit View Display Processing Window Help
5.835e+002
100
%
0
1
4
12
17
22
0.0 1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0 9.0 10.0 11.0 12.0 13.0 14.0 15.0 16.0 17.0 18.0 19.0 20.0 21.0 22.0 23.0
Samples
Ready
NUM

Fig. 5B

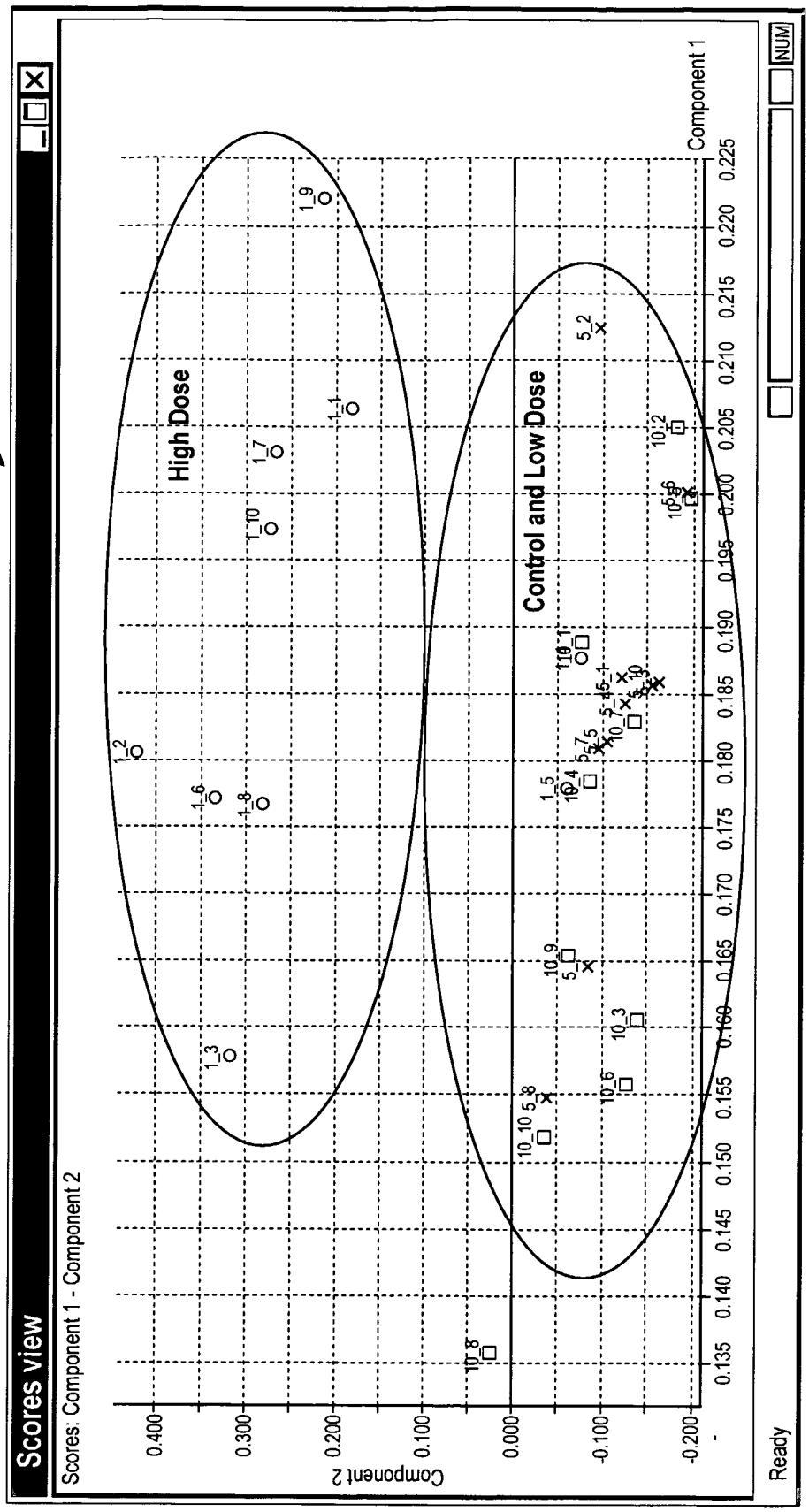

75
Scores view
Scores: Component 1 - Component 2
High Dose
Control and Low Dose
Component 1
Component 2
Ready
NUM

Fig. 6

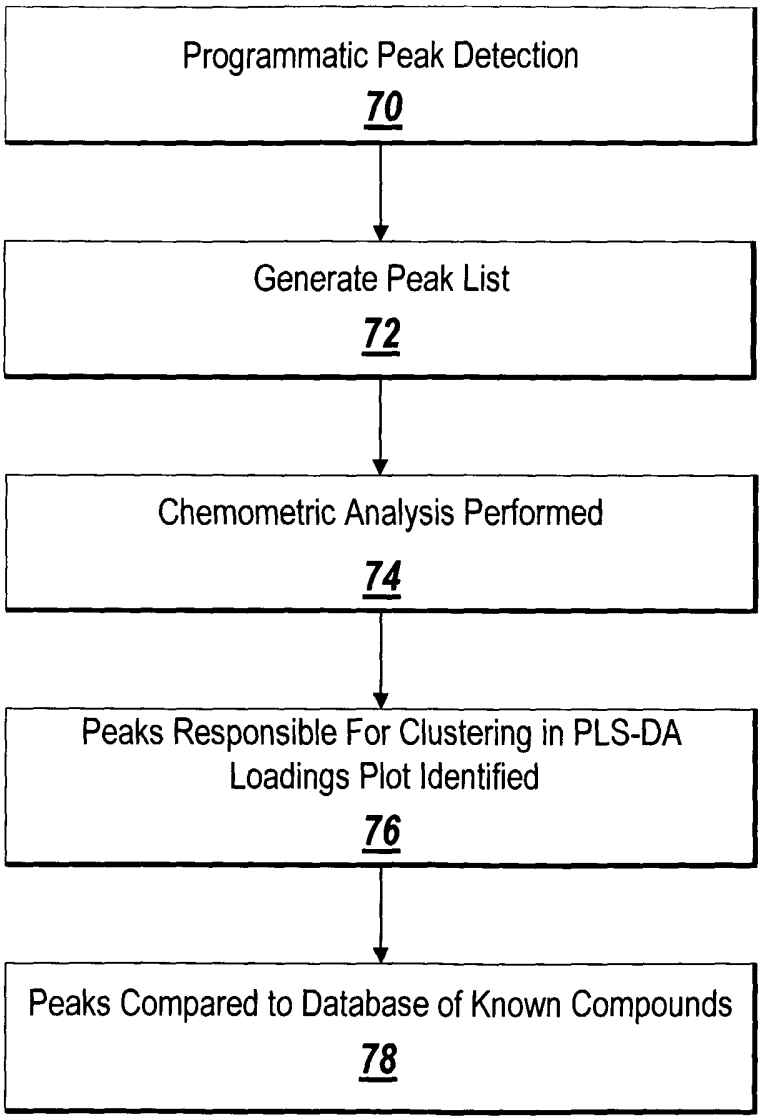

Programmatic Peak Detection
70
Generate Peak List
72
Chemometric Analysis Performed
74
Peaks Responsible For Clustering in PLS-DA Loadings Plot Identified
76
Peaks Compared to Database of Known Compounds
78

Fig. 7

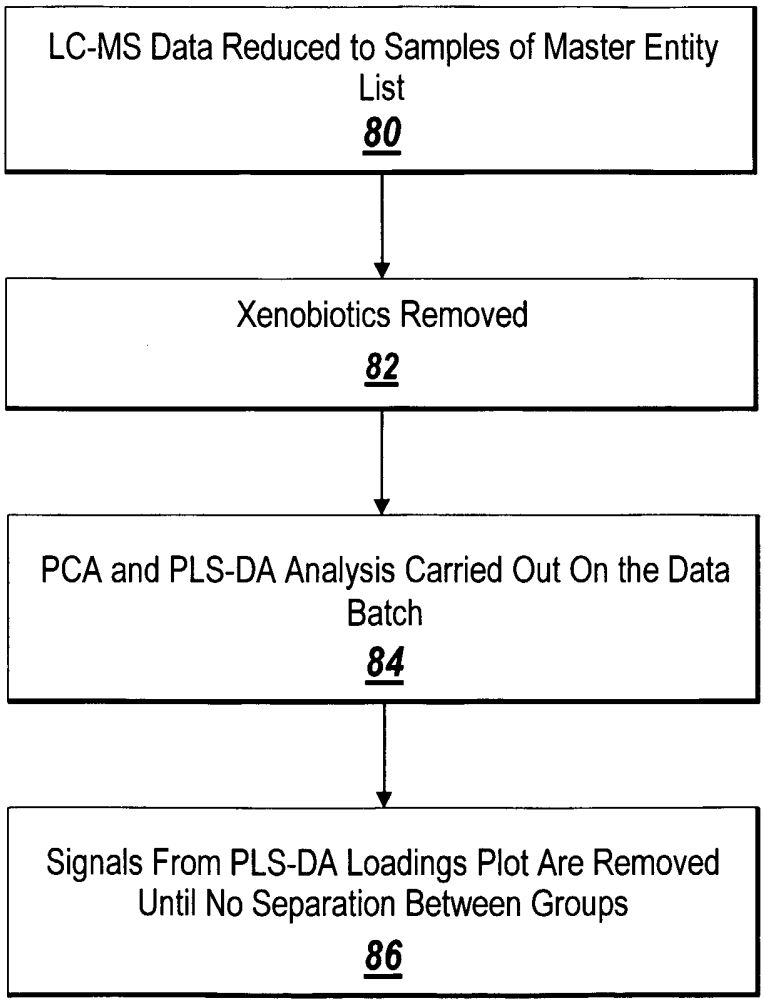

LC-MS Data Reduced to Samples of Master Entity List
80
Xenobiotics Removed
82
PCA and PLS-DA Analysis Carried Out On the Data Batch
84
Signals From PLS-DA Loadings Plot Are Removed Until No Separation Between Groups
86

Fig. 8

INTERNATIONAL SEARCH REPORT

International Application No
/US2004/016797

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G01N30/86

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)
IPC 7 G01N

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 practical, search terms used)
EPO-Internal, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	PLUMB ROBERT S ET AL: "Metabonomics: The use of electrospray mass spectrometry coupled to reversed-phase liquid chromatography shows potential for the screening of rat urine in drug development." RAPID COMMUNICATIONS IN MASS SPECTROMETRY, vol. 16, no. 20, 2002, pages 1991-1996, XP009037995 ISSN: 0951-4198 the whole document	1-20
X	WO 03/017177 A (BEYONG GENOMICS INC) 27 February 2003 (2003-02-27) page 2, line 6 - page 5, line 28	1-20

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in 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 document 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
13 October 2004	27/10/2004

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, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Müller, T

Form PCT/ISA/210 (second sheet) (January 2004)

INTERNATIONAL SEARCH REPORT

International Application No
US2004/016797

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	EP 1 327 883 A (MAX PLANCK GESELLSCHAFT) 16 July 2003 (2003-07-16) column 1, lines 3-6 column 10, lines 9-12 column 12, lines 29-58 -----	1-20

INTERNATIONAL SEARCH REPORT

nformation on patent family members

International Application No
'US2004/016797

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 03017177 A	27-02-2003	CA 2457432 A1	27-02-2003
		EP 1425695 A2	09-06-2004
		WO 03017177 A2	27-02-2003
		US 2003134304 A1	17-07-2003
EP 1327883 A	16-07-2003	EP 1327883 A2	16-07-2003
		WO 03058238 A1	17-07-2003