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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 March 2003 (20.03.2003)

PCT

(10) International Publication Number
WO 03/023409 A2

(51) International Patent Classification[7]: G01N 35/00

(21) International Application Number: PCT/US02/28588

(22) International Filing Date:
6 September 2002 (06.09.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/318,152	7 September 2001 (07.09.2001)	US
60/318,157	7 September 2001 (07.09.2001)	US
60/318,138	7 September 2001 (07.09.2001)	US

(71) Applicant: **TRANSFORM PHARMACEUTICALS, INC.** [US/US]; 29 Hartwell Avenue, Lexington, MA 02421 (US).

(72) Inventors: **JOHNSON, Alasdair, Yate**; 44 Prospect Street, Newburyport, MA 01950 (US). **CIMA, Michael, John**; 184 Mystic Valley Parkway, Winchester, MA 01890 (US). **LEMMO, Anthony, V.**; 3 Washington Street, Sudbury, MA 01776 (US). **ALMARSSON, Orn**; 22 Farmington Drive, Shewsbury, MA 01545 (US). **MORISSETTE, Sherry, Lynn**; 11 Brattle Drive, #10, Arlington, MA 02474 (US). **GONZALEZ-ZUGASTI, Javier**; 15 Angie Road, N. Billerica, MA 01862 (US). **LEVINSON, Douglas**; 111 Maple Street, Sherborn, MA 01770 (US). **MCNULTY, Chris**; 4745 Xerxes Ave., Minneapolis, MN 55410 (US).

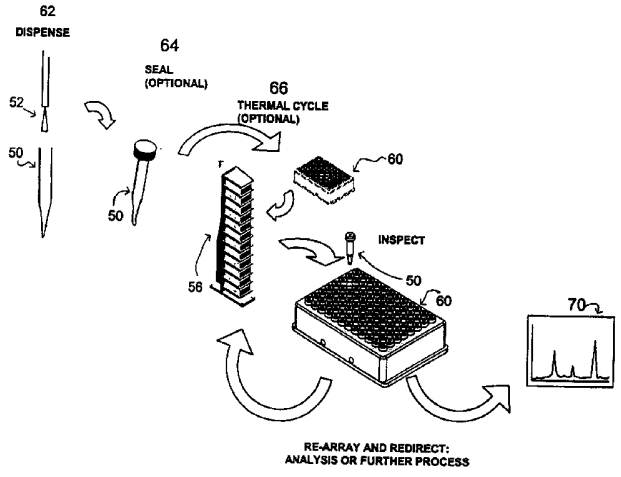
(74) Agents: **GORUZZI, Laura, A.** et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, 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, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),

[Continued on next page]

(54) Title: APPARATUS AND METHOD FOR HIGH-THROUGHPUT PREPARATION AND CHARACTERIZATION OF COMPOSITIONS



(57) Abstract: Systems and methods are described that allow the high-throughput preparation, processing, and study of arrays of samples, each of which comprises at least one compound. Particular embodiments of the invention allow a large number of experiments to be performed in parallel on samples that comprised of one or more compounds on the milligram or microgram quantities of compounds. Other embodiments of the invention encompass methods and devices for the rapid screening of the results of such experiments, as well as methods and devices for rapidly determining whether or not similarities exist among groups of samples in an array. Particular embodiments of the invention encompass methods and devices for the high-throughput preparation of different forms of compounds (e.g., different crystalline forms), for the discovery of new forms of old compounds, and for the discovery of new methods of producing such forms. Embodiments of the invention also allow for the high-throughput determination of how specific compounds or forms of compounds behave when exposed to other chemicals or environmental conditions.

WO 03/023409 A2

Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

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.

APPARATUS AND METHOD FOR HIGH-THROUGHPUT PREPARATION AND CHARACTERIZATION OF COMPOSITIONS

This application claims priority to U.S. provisional application nos. 60/318,152, 60/318,157, and 60/318,138, each of which was filed on September 7, 2001, and each of which is incorporated herein in its entirety.

1. FIELD OF THE INVENTION

This invention generally relates to devices, systems, and methods for conducting and evaluating multiple small-scale experiments. Particular embodiments of the invention encompass methods and devices for the high-throughput preparation and study of a variety of compounds, compositions, and forms of compounds and compositions.

2. BACKGROUND OF THE INVENTION

In recent years, chemical discovery has seen an explosion of new science, such as genomics, proteomic and bioinformatics, as well as high-throughput technologies for identifying and/or creating new compounds or chemical entities, such as combinational chemistry. Such technologies allow the researcher to rapidly synthesize and/or identify large numbers of compounds. At the same time, these technologies have led to the development of more compounds that are larger, greasier and more hydrophobic, and thus more challenging to develop into products.

Conducting large numbers of experiments results in the need to inspect or otherwise analyze hundreds or thousands of samples, *e.g.*, for the presence of the desired result. And, a large number of the pre-selected samples require continuing analysis. The resulting voluminous data must then be processed effectively and efficiently, *e.g.*, within a reasonable amount of time.

The physical form of a compound, particularly that of an active pharmaceutical ingredient (API), plays a role in a number of areas. For example, in order to be developed into a drug, a compound must be able to be delivered to the patient via some suitable device or formulation, and it must also pass criteria in several categories, such as safety, metabolic profile, pharmacokinetics, cost and reliability of synthetic process, stability, and bioavailability.

High-throughput technologies, when possible, enable the discovery of various physical forms of a compound, some of which may be particularly useful as pharmaceuticals, for formulating pharmaceuticals, intermediates for manufacturing drugs, foods, food additives and the like. (*See, e.g.*, International Application Nos. WO00/59627, WO01/09391,

and WO01/51919). Such technologies can result in extraordinary numbers of experiments being conducted very rapidly thereby creating large amounts of data and results that must be reviewed and analyzed by the scientist in order to identify a desired form of the compound. For example, in order to discover various solid forms of a compound, often thousands of experiments, using many different conditions, solvents, additives, pH, thermal cycles, and the like must be conducted. Dozens or even hundreds of the forms must be analyzed before a desired form of the compound can be identified and chosen for further development as a potential product.

Some devices for facilitating large numbers of experiments simultaneously are known. In addition, there are systems consisting of blocks with multiple wells for performing reactions for different applications such as combinatorial chemistry. Examples of such system include the TITAN™ Reactor Clamp and TITAN™ PTFE MicroPlates (both available from Radleys, Shire Hill, Saffron Walden, Essex CBII 3AZ, United Kingdom). A multiple-well tray for crystallization reactions is described in US Pat. No. 6,039,804. There also exist systems of block, tubes, and seals, such as the Radleys TITAN™ Glass Micro Reactor Tube System and the WebSeal System (available from Radleys, Shire Hill, Saffron Walden, Essex CBII 3AZ, United Kingdom). Many tubes or vials of different geometries also exist, including many with crimp, threaded, or snap-on caps.

Spectroscopic techniques such as infrared (IR) and Raman spectroscopy are useful for detecting changes in structure and/or order. In addition, techniques such as Nuclear Magnetic Resonance (NMR), Differential Scanning Calorimetry, ultra-violet (UV) spectroscopy, circular dichroism (CD), linear dichroism (LD), and X-ray diffraction are powerful techniques. However, each of these techniques must be coupled with data analysis and handling techniques to enable data collection and processing of hundred or thousands of samples. All these techniques are not easily adaptable for high-throughput analysis of structural information and order. Indeed, high-throughput analysis still remains a challenge due to the high degree of automation desired in both physical sample handling and in analysis of the collected data. These and many other difficulties are overcome by the system and methods disclosed herein. The invention disclosed herein further extends the reach of high-throughput analysis with a high degree of sensitivity and specificity. Moreover, the disclosed techniques also efficiently use limited test material quantities to enable effective screening at a low cost.

3. SUMMARY OF THE INVENTION

This invention is directed, in part, to methods and systems for determining conditions that when applied to a particular compound or composition provide a particular result (*e.g.*, a compound or composition having particular chemical and/or physical properties). The invention is further directed to methods and systems for the generation, synthesis, and/or identification of various forms of a compound or composition, such as, but not limited to, polymorphs, salts, hydrates, solvates, desolvates, and amorphous forms. The invention is also directed to methods and systems for the generation, synthesis, and/or identification of various forms of solids such as, but not limited to, crystal habit and particle size distribution.

The invention encompasses a complete system for planning and conducting high-throughput experiments, *e.g.*, experiments on one or more arrays of samples. The system encompasses apparatuses and methods that can be used to prepare and process samples, apparatuses and methods that can be used to inspect, process, and screen samples, apparatuses and methods that can be used to collect spectroscopic and other data from one or more of the samples, and apparatuses and methods that can be used to process, interpret, and analyze the data. The system includes robotics, computers, spectral techniques, and various mechanical devices, each designed to conduct high-throughput experiments on large or preferably small amounts of material, including materials on the milligram and microgram scales.

In particular, this invention encompasses methods and devices for the high-throughput preparation, processing, screening, and/or analyzing of samples. Particular methods of the invention utilize arrays of samples, each of which comprises the compound or composition of interest (referred to herein as the "compound-of-interest") in optional contact with one or more solvents or excipients. In specific embodiments of the invention, each sample is held in a container that can be manipulated separately from other samples in the array.

One embodiment of the invention encompasses a high throughput system for evaluating experiments, which comprises: a) a plurality of containers, each of which contains a compound-of-interest and optionally one or more additional compounds; b) a block containing an array of holes for receiving said containers; and c) an imaging device capable of producing images of the samples while in the containers.

Another embodiment of the invention encompasses a method of evaluating experiments which comprises: a) providing a system for evaluating experiments comprising: i) a plurality of containers, each of which contains a compound-of-interest and optionally one or more additional compounds; ii) a block containing an array of holes for receiving the

containers; and iii) an imaging device; b) positioning the block near the imaging device; c) producing images of the contents of each of the containers; d) analyzing the images for the presence of a desired experimental result; and e) identifying containers with the desired experimental result.

Another embodiment of the invention encompasses an automated high throughput method for screening for solid forms of a compound-of-interest, wherein the compound-of-interest is a biologically active small organic molecule, which comprises: a) preparing an array of samples, each of which comprises the compound-of-interest and optionally one or more additional compounds; b) processing the array so as to generate solid forms; c) prescreening the array for solid formation using a digital imaging camera; d) identifying samples with solid formations for further analysis; e) rearranging and reprocessing samples with solids, and optionally repeating steps (b) to (d).

Another embodiment of the invention encompasses a high-throughput system for processing samples and screening for solid forms of a compound-of-interest, which comprises: a) removable containers, each of which contains a compound-of-interest and optionally one or more additional compounds; b) a block made of a thermally conductive material having an array of holes, each hole having a top and a bottom, the top having an opening for receiving the removable container and the bottom having an access hole; c) a thermal processing system for heating and cooling multiple blocks simultaneously; and d) an imaging device (*e.g.*, a camera, preferably a digital video camera) for detecting solid forms.

Another embodiment of the invention encompasses a method for the high-throughput processing and screening of samples, which comprises: a) providing a system for processing a sample comprising: i) removable containers; ii) a block having an array of holes, each hole having a top and a bottom, the top having an opening for receiving a container and the bottom having an access hole; b) placing the containers in the holes; c) dispensing a controlled amount of a compound-of-interest and optionally one or more additional compounds in each container to provide an array of samples; d) processing the array; and e) screening the samples for the presence or absence of solid forms using an imaging device.

Another embodiment of the invention encompasses a method of screening an array of samples which comprises obtaining a Raman spectrum of each sample and determining which, if any, of the spectra share a spectral feature.

Another embodiment of the invention encompasses a method of screening an array of samples for the presence of a particular form of a compound-of-interest, which comprises obtaining a Raman spectrum of each sample.

Another embodiment of the invention encompasses a method of screening an array of samples for the absence of a particular form of a compound-of-interest, which comprises obtaining a Raman spectrum of each sample.

Another embodiment of the invention encompasses a system for detecting similarities among a plurality of samples, which comprises: a) a device for obtaining a spectrum for each sample; and b) a computer configured to analyze each of the spectra and to generate a plurality of bins, wherein each bin corresponds to samples sharing at least one spectral feature.

Another embodiment of the invention encompasses a method of detecting similarities among a plurality of samples, which comprises: a) collecting a spectrum for each of the plurality of samples; b) calculating a similarity metric between the spectrum of one sample and that of at least one other of the plurality; c) clustering, based on the similarity metric, the spectra into bins, each bin containing similar spectra; and d) presenting the clustered spectra with similar spectra located close to each other.

Another embodiment of the invention encompasses a method of analyzing a plurality of samples, which comprises: a) analyzing the samples with a spectrometer to produce spectral data; b) under processor control, identifying similarities between the spectra; and c) grouping the spectra into bins of similarity.

Another embodiment of the invention encompasses a database containing a plurality of spectral samples organized into a plurality of bins, the bins corresponding to a hierarchical organization of the plurality of spectral samples based on pair-wise similarity scores calculated in accordance with a similarity metric.

3.1. BRIEF DESCRIPTION OF THE DRAWINGS

Features and advantages of the invention can be understood with reference to the attached figures. The drawings are not necessarily to scale, emphasis instead being placed upon illustrating the principles of the invention.

FIG. 1 is a schematic diagram of steps associated with a specific embodiment of the invention, wherein tubes are filled with a compound-of-interest and optional other compounds, processed, and inspected.

FIGS. 2A and **2B** are views of a tube in its open and capped configurations, respectively.

FIGS. 3A and **3B** are top and bottom perspective views of a block, respectively, and **FIG. 3C** is a top perspective view of the block filled with capped tubes.

FIG. 4A is a drawing of a temperature-controlled shelf assembly, or "hotel," loaded with twelve blocks.

FIG. 4B is a drawing of a shelf equipped for use with a heating/cooling loop (*e.g.*, using water, ethylene glycol, or another solvent) shown as dotted lines.

FIG. 5 shows a drawing of a thermal cycling system top-level assembly, including an environmental enclosure. In this arrangement, 18 hotels in a semi-circular pattern around a robotic arm are shown.

FIG. 6 shows a flowchart depicting an example of the logic for addressing solid form generation using the vision station approach.

FIG. 7A is a schematic diagram of a specific vision station.

FIGS. 7B and **7C** are drawings of a vision station in operation, showing a side view (**FIG. 7B**) and a perspective view (**FIG. 7C**) of the block wherein a lifting mechanism lifts an entire row of vials at the same time. The light source can be positioned on top of the samples (*e.g.*, normal to the CCD camera) or opposite the samples.

FIG. 8 is a drawing that shows the difference between samples with no birefringence and samples with birefringence.

FIG. 9 is a drawing showing the scattering and diffusion of laser light pointed at a single tube and consecutive tubes.

FIG.10 is a drawing that shows the effect of illuminating of a tube containing a colloidal suspension (a) compared to a tube containing pure water (b). In **FIG. 10**, the laser light is at an angle that is not normal to the camera lens.

FIG. 11 is a flow diagram depicting the use of the vision station to detect birefringence in crystalline solid forms, or to differentiate crystalline versus amorphous solid forms by the observance of birefringence.

FIG. 12 is a flow diagram depicting the use of the vision station for laser light interrogation of samples. Shown are diagrams of nano-suspension compared to true solutions.

FIG. 13A is a perspective diagram of the Raman system.

FIG. 13B is a diagram showing a block of tubes being moved inside an enclosure.

FIG. 13C is a diagram showing the lifting mechanism elevating a tube to be gripped by the tube gripper.

FIG. 13D is a diagram showing the tube gripper and tube having moved in the vertical direction.

FIG. 13E is a diagram showing the tube gripper and tube having rotated.

FIG. 13F is an enlarged diagram showing the tube gripper and tube having moved in the horizontal direction to bring the tube closer to the microscope objective.

FIG. 13G is an enlarged diagram showing the tube gripper and tube having been lowered to a position near the tube holder.

FIG. 13H is an enlarged diagram showing the tube gripper having loaded the tube into the tube holder.

FIG. 13I is an enlarged diagram showing the tube gripper having retracted after loading the tube into the tube holder.

FIG. 13J is an enlarged diagram showing a tube rotator engaging the tube.

FIG. 13K is an enlarged diagram showing the tube being moved under the microscope objective.

FIG. 14A is a perspective view of a tube and microscope objective. The long axis of the tube is preferably at a 90 degree angle to the axis of the objective.

FIG. 14B is diagram of a tube and microscope objective indicating the available axes of motion for the tube.

FIG. 14C is a closer view of a tube with a crystal inside that is in an out-of-focus position with respect to the microscope objective. This figure also indicates the available axes of motion for the tube.

FIG. 14D is a detailed view of a tube with a solid, such as a crystal, inside that is located toward the narrow end of the tube and on the bottom surface with respect to the microscope objective.

FIG. 14E is a detailed view of a tube with a solid, such as a crystal, inside that is being moved in the horizontal direction to bring it closer to the in-focus position beneath the microscope objective.

FIG. 14F is a detailed view of a tube with a solid, such as a crystal inside that is being rotated to bring it closer to the in-focus position beneath the microscope objective.

FIG. 14G is a detailed view of a tube with a solid, such as a crystal, inside that is being moved in the vertical direction to bring it to the in-focus position beneath the microscope objective.

FIG. 15 is a flowchart depicting six stages of a computational binning process of one embodiment of the invention, and one optional stage in such a process.

FIG. 16A is a graph showing Raman intensity plotted as a function of Raman shift (cm^{-1}) for an empty glass vial.

FIG. 16B is a graph showing Raman intensity plotted as a function of Raman shift (cm^{-1}) for a fluorescent sample.

FIG. 16C is comparative graph showing Raman intensity plotted as a function of Raman shift (cm^{-1}) for the pre-filtered sample of **FIG. 17B** compared to the corresponding filtered spectra after the fluorescence has been removed.

FIG. 17A is a screen shot showing the output from the binning software captured during the binning procedure for the flufenamic acid sample set.

FIG. 17B is a screen shot showing the output from the binning software captured during the binning procedure for the theophylline sample set.

FIG. 18 illustrates an implementation of the binning procedure.

FIG. 19A is a comparative graph of X-ray powder diffraction patterns for Form I and Form III of flufenamic acid.

FIG. 19B is a comparative graph of X-ray powder diffraction patterns for anhydrous and monohydrate forms of theophylline.

FIG. 20A is a comparative graph of DSC thermograms for Form I and Form III of flufenamic acid where heat flow (W/g) is plotted as a function of temperature (°C).

FIG. 20B is a comparative graph of DSC thermograms for anhydrous and monohydrate forms of theophylline where heat flow (W/g) is plotted as a function of temperature (°C).

FIG. 21 is a graph showing thermograms obtained for the anhydrous and hydrous forms of theophylline. An inset graph shows an enlargement of the same thermograms between 35 °C and 150 °C.

FIG. 22A is a graph showing Raman intensity (arbitrary units) plotted as a function of Raman shift (cm^{-1}) for Form I and Form III of flufenamic acid.

FIG. 22B is a graph showing the Raman intensity (arbitrary units) plotted as a function of Raman shift (cm^{-1}) for anhydrous and monohydrate forms of theophylline.

FIG. 23A is the output after clustering illustrating sorted cluster diagrams for the flufenamic acid sample set.

FIG. 23B is the output after clustering illustrating sorted cluster diagrams for the theophylline sample set.

FIG. 24A illustrates X-ray crystal diffraction spectra corresponding to the anhydrate and the hydrate forms of Theophylline.

FIG. 24B illustrates the binning of Raman Spectra corresponding to Hydrate distinctly from Anhydrate form of Theophylline.

3.2. DEFINITIONS

As used herein and unless otherwise indicated, the term "array," when used to refer to a plurality of objects (*e.g.*, samples), means a plurality of objects that are organized physically or indexed in some manner (*e.g.*, with a physical map or within the memory of a computer) that allows the ready tracking and identification of specific members of the plurality. Typical arrays of samples comprise at least 6, 12, 24, 94, 96, 380, 384, 1530, or 1536 samples.

As used herein and unless otherwise indicated, the term "compound-of-interest" refers to the substance, compound, molecule, or chemical studied, formulated, or otherwise manipulated using methods or devices of the invention. Examples of compounds-of-interest include, but are not limited to, pharmaceuticals, veterinary compounds, dietary supplements, alternative medicines, nutraceuticals, sensory compounds, agrochemicals, the active components of consumer products, and the active components of industrial formulations. A preferred compound-of-interest is the active component of a pharmaceutical, also referred to as the active pharmaceutical ingredient (API). Specific APIs are suitable for administration to humans. Specific APIs are small organic molecules that are not polypeptides, proteins, oligonucleotides, nucleic acids, or other macromolecules. Small organic molecules include, but are not limited to, molecules with molecular weights of less than about 1000, 750, or 500 grams/mol.

As used herein and unless otherwise indicated, the term "controlled amount" refers to an amount of a compound that is weighed, aliquotted, or otherwise dispensed in a manner that attempts to control the amount of the compound. Preferably, a controlled amount of a compound differs from a predetermined amount by less than about 10, 5, or 1 percent of the predetermined amount. For example, if one were to dispense, handle, or otherwise use 100 μg of a compound-of-interest, a controlled amount of that compound-of-interest would preferably weight from about 90 μg to about 110 μg, from about 95 μg to about 105 μg, or from about 99 μg to about 101 μg.

As used herein and unless otherwise indicated, the term "form" refers to the physical form of a compound or composition. Examples of forms include solid and liquid. Examples of forms of solids, or "solid forms," include, but are not limited to, salts, solvates (*e.g.*, hydrates), desolvates, clathrates, amorphous and crystalline forms, polymorphs, crystal habits (*e.g.*, needles, plates, particles, and rhomboids), crystal color, crystal size, crystal size distribution, co-crystals, and complexes.

As used herein and unless otherwise indicated, the term "pharmaceutical" refers to a substance, compound, or composition that has a therapeutic, disease or condition preventive, disease or condition management, diagnostic, or prophylactic effect when administered to an animal or human, and includes prescription and over-the-counter pharmaceuticals. Examples of pharmaceuticals include, but are not limited to, macromolecules, oligonucleotides, oligonucleotide conjugates, polynucleotides, polynucleotide conjugates, proteins, peptides, peptidomimetics, polysaccharides, hormones, steroids, nucleotides, nucleosides, amino acids, small molecules, vaccines, contrasting agents, and the like.

As used herein and unless otherwise indicated, the term "sample" refers to an isolated amount of a compound or composition. A typical sample comprises a controlled amount of a compound-of-interest, and may also contain one or more excipients, solvents, additives (*e.g.*, stabilizers and antioxidants), or other compounds or materials (*e.g.*, materials that facilitate crystal growth). Specific samples comprise a compound-of-interest in an amount less than about 100 mg, 25 mg, 1 mg, 500 μg, 250 μg, 100 μg, 50 μg, 25 μg, 10 μg, 5 μg, 2.5 μg, 1 μg, or 0.5 μg.

4. DETAILED DESCRIPTION OF THE INVENTION

This invention is based, in part, on the discovery of various methods and devices that can facilitate the rapid and efficient preparation and analysis of compounds, compositions, and various forms of compounds and compositions. For the sake of convenience, methods and devices of the invention may be described with reference to four general aspects of it, which are referred to herein as "Sample Containment and Preparation," "Sample Handling and Processing," "Sample Imaging," and "Spectroscopic Data Collection and Analysis." Each of these aspects encompasses novel embodiments of the invention which can be used alone or together.

For example, a specific method comprises the preparation of an array of samples, each of which is held in a sealed container; exposing the samples to a condition, such as heat or cold, for a particular amount of time; imaging the samples to determine, for example, whether they produced or contain a solid or liquid; and collecting and analyzing spectroscopic data obtained from one or more of the samples. **FIG. 1** provides a general illustration of this method and a system that can be used to implement it. Briefly, one or more tubes **50** are placed in an array of wells in a thermally conductive block **60**. Next, chemical ingredients **52** for each experiment are dispensed into tubes **50**. Optionally, a cap **54**, possibly including a seal **64**, is placed on tube **50** to prevent leakage, evaporation, or

contamination of the contents. The tubes are then optionally processed using a controlled thermal cycling system **56**. Following inspection of the contents of tubes **50** (*e.g.*, using automated imaging equipment), experimental specimens or samples of interest are identified. The samples of interest are then optionally separated from other samples for further analysis or processing.

The other embodiments of the invention can be understood from the more detailed discussions of its Sample Containment and Preparation, Sample Handling and Processing, Sample Imaging, and Spectroscopic Data Collection and Analysis aspects provided below.

4.1. SAMPLE CONTAINMENT AND PREPARATION

4.1.1. Tubes and Blocks

High throughput preparation and analysis of samples is aided by the assembly of arrays of samples, each of which can be the same or different from other samples in the array. In specific embodiments of this invention, arrays of samples are prepared in removable containers (*e.g.*, vials or tubes), which fit in holes, wells, or depressions in a holder, or what is referred to herein as a "block." This system is referred to herein as "tubes and blocks" or "tubes in blocks."

A wide variety of containers known to those skilled in the art can be used to hold the individual samples in an array. Because preferred embodiments of the invention are directed to the high-throughput preparation, processing, and/or testing of samples that contain relatively small amounts of the compound-of-interest, preferred containers are sufficiently small so that many of them can be fit into a block. Preferred containers are also optically transparent or translucent to allow visual inspection of their contents, are chemically inert (*e.g.*, will not chemically react with the compounds they contain), and can withstand physical conditions (*e.g.*, thermal processing) to which it will be exposed. Specific containers are made of glass or polypropylene. Preferred containers can be sealed or closed. For example, a septum that can be pierced by a needle or other device that can add fluids to the container or remove fluids from the container is used in a preferred embodiment. Containers may also be closed with a closure (*e.g.*, a cap or top) that allows light to pass through into the container to illuminate its contents. Moreover, the closure may be used to imprint or otherwise provide an identifier to a single tube or a sub-block. Such an identifier may be in addition to or in lieu of an identifier associated with each block.

FIG. 2 provides a view of a tube container in its open and capped configurations. The specific closure **54** shown in the figure can be crimped, is made of aluminum or some other suitable material, and incorporates a polymer septum **55**.

The blocks that hold the containers preferably allow for the automated removal and insertion of the containers. For example, a particular block has holes with top openings large enough to accommodate containers and smaller bottom openings that allow the containers to be pushed out of their holes with a rod or pin. Such blocks can be used with particular systems of the invention that comprise a lifter mechanism capable of protruding through the access hole in the blocks to elevate one or more containers until at least partially removed from the block. Preferably, the block is thermally conductive and is made from metal (*e.g.*, copper, steel, or aluminum), although other materials, such as plastic, may also be used. As shown in **FIGS. 3A-3C**, a specific block **60** is made of aluminum, and has 96 holes **61** into which containers will fit. Block **60** provides thermal transfer between the tubes and a controlled means for temperature regulation. Block **60** incorporates a set of bottom access holes **63** that provide for physical and optical access to each individual vial. Optionally, block **60** has one or more indents **65** on two opposing sides of block **60** to provide means for preventing slipping or dropping of the block **60** when automated handling, such as by a robotic arm, is used to move block **60**.

The geometry, size, and materials from which a block is made can be readily adapted for use with particular containers, processing conditions, and sample and block handling devices. For example, the holes in a block may be counter-bored, counter-sunk, stepped, tapered, or more complex-shaped to fit different tube and seal shapes, although in **FIG. 3** they are simply shown as illustrative straight through holes.

The tube and block system has several distinct advantages over alternative ways of performing parallel experiments. First, the use of individual containers, instead of using a plate, format allows for the individual handling of each sample, or experiment, in an array. This makes it possible to re-array containers to separate those that show desired properties from the rest, in order to perform further processing or analysis of only some of the experiments. In addition, for experiment samples or products that can exhibit different properties depending on orientation (*e.g.*, samples that contain crystals), the containers can be precisely oriented with respect to an analysis instrument, such as a Raman spectrometer or X-ray diffractometer.

The invention also encompasses the use of various tube materials, including the use of different types of glass such as amber glass vials, which can protect their contents from

degradation due to exposure to light during processing. Optical inspection of the contents of each vial is possible by illuminating the samples with light source having a wavelength able to penetrate the vial walls, and using a detector (camera) for imaging light at that wavelength.

Second, a translucent, transparent, semitransparent, or clear container allows for optical inspection of the experiment from multiple angles, generally perpendicular to the axis of the container, but also from underneath through optional access holes in the block. Also, the use of such containers, including but not limited to, glass or clear polypropylene tubes, allows for optical inspection methods such as machine vision or microscopy. In addition, clear plastic tubes, or tubes fabricated from quartz or any other optically transparent, translucent, or clear material can be used. Chromacol Ltd. (2 Little Mundells, Wellwyn Garden City, Herts AL7 1EW, UNITED KINGDOM) offers many examples of the variety of available vial shapes and materials. The ability to visually inspect each experiment in an array, from all angles, allows analysis of the contents, such as solids or precipitates, in a number of ways, including without limitation, estimating size, color, shape, orientation and location in the container.

Third, because containers can be of any shape, the tubes and blocks system enables the testing of a wide range of experimental volumes. By selecting the shape of the containers, small volume experiments (*e.g.*, about 2µl) are still clearly visible in the narrow, preferably conical, tip of the preferred container, while larger volume experiments (greater than 100µl) may also be tested due to the larger diameter top section of the tubes. Also, the container geometry in the preferred embodiment permits the use of a tightly sealing cap. An airtight seal isolates the contents of the experiments from the environment and prevents evaporation, leakage or contamination of, or changes to, the components in the containers.

Fourth, many containers can be capped. The use of a cap with an integral translucent frit or septum allows for the ability to probe, add, or remove components to/from the experiment, as well as the ability to illuminate the container's contents through the septum. This lighting of the samples in the containers though the septa can be accomplished through the use of light sources such as fiber optic light guides or light-emitting diodes (LEDs).

Fifth, the use of thermally conductive blocks allows for quick heat transfer between a heating or cooling source and the containers, as well as a large thermal mass to maintain the containers at the desired temperature when temporarily not in contact with a heat/cooling source. As noted previously, many metals, plastics, and a variety of other materials can used to build the blocks. Although aluminum does not exhibit the best thermal conductivity or

heat capacity, it is preferred in view of additional considerations such as weight, cost, corrosion-resistance, and ease of manufacturing.

Sixth, the chosen geometry of the block offers certain advantages. For example, the access holes at the bottom of each container hole allows for physical access to the containers, so that they can be partially or fully removed from the block for inspection or rearranging purposes. In addition, the holes also provide a window for optical inspection of the containers from the underside of the block that can be used alone or in conjunction with top lighting of the containers, *e.g.*, through translucent septa, to image the experiments in the containers in a block.

4.1.2. Sample Preparation

The composition of a particular sample in an array will depend on the use to which the particular method or device of the invention is put. For example, if an array is used to provide crystalline forms of a compound-of-interest, each sample might contain one or more solvents or solvent mixtures in addition to the compound-of-interest (which could be evaporated) or to which other solvents (*e.g.*, antisolvents, reagents that affect pH, counterion concentration, or the ionic character of the solvent) or materials (*e.g.*, nucleation promotors) could be added during the processing of the samples. The specific composition of each sample in an array might be the same (to allow redundancy) or different (to allow the simultaneous testing of numerous crystallization conditions). However, the invention also encompasses the use of arrays to attempt the crystalizations of compounds-of-interest from melts, in which case the samples might only contain solid compound-of-interest.

In another example, the array is used to determine various characteristics of a compound-of-interest, or how they change when exposed to particular conditions (*e.g.*, those described below in the Sample Handling and Processing section). Examples of characteristics include, but are not limited to, form, chemical composition, solubility, physical and/or chemical stability, and hygroscopicity.

Whatever the purpose to which an embodiment of this invention is put, each container (apart from any containers used as controls, or blanks), will comprise a controlled amount of the compound-of-interest and, optionally, one or more additional compounds (*e.g.*, solvents, excipients, or nucleation agents). The containers may also contain a stirbar or other device to facilitate stirring, uniform heating, or anything else that is deemed necessary for the partcular use to which the invention is being put. All of these materials are preferably added to containers in an automated fashion. For example, compounds-of-interest and solvents can be

deposited into the vials in a variety of ways, ranging from hand-pipetting to automated liquid and/or solid dispensing. Dispensing of chemicals into the vials is preferably accomplished with an automated reagent dispensing apparatus, such as Cartesian Technologies' PreSys model (available from Cartesian Technologies Inc., 17851 Sky Park Circle, Suite C, Irvine, CA 92614, USA), and multiple-channel liquid dispensers, such as those available from Tecan Group Ltd. (Tecan Group Ltd., Seestrasse 103, 8708 Männendorf, SWITZERLAND). Other models and brands of liquid dispensers can also be used. Solid compounds and compositions can also be dispensed by hand or by automated means known in the art. For example, a solution comprising a compound-of-interest can be dispensed into sample containers, after which the solvent can be removed to provide a controlled amount of the compound-of-interest (*e.g.*, in a milligram or microgram quantity).

After samples have been prepared, the containers that hold them are preferably sealed to prevent leakage, contamination, and evaporation (unless otherwise desired), as well as to prevent outside factors (*e.g.*, humidity changes) from affecting the samples. Preferred containers are vials which can be sealed using crimpable metal caps or compliant gaskets, such as a silicon frits or septa. Other means of sealing containers include, but are not limited to, wax plugs, threaded caps, caps that snap over the vial opening, and compression or adhesive seals. Preferred septa allow for the illumination of the contents of a container from the top, and also allow for the addition or withdrawal of materials or components to/from the tube. The capping or sealing of the containers is preferably accomplished using an automated means, such as a Wheaton Crimpmaster Crimping Station (Wheaton Science Products, 1501 No. 10th Street, Millville, NJ 08332, USA) pneumatically powered crimper. Alternatively, hand powered crimper tools (also Wheaton Science Products) may be used.

The invention encompasses the labeling of either the vial itself or the crimped seal cap that allows the ready identification of individual samples. Both crimp caps and glass vials may be labeled, for instance, through laser and inkjet marking, by using human-readable, alphanumeric codes, as well as using machine-readable codes such as DataMatrix 2-D codes. Such codes may advantageously be scanned and tracked with optical readers. Similarly, other types of barcodes and marking technologies may be used without limitations.

4.2 SAMPLE HANDLING AND PROCESSING

Particular embodiments of the invention encompass exposing the samples in an array to one or more conditions such as, but not limited to, pH, ion concentration, solvent, temperature, and light for a particular amount of time. A typical condition is temperature,

and one embodiment of the invention encompasses a thermal cycling system capable of processing many blocks simultaneously. This system comprises one or more shelves, preferably thermally conductive, onto which blocks can be placed, and heating and/or cooling means such as, but not limited to, chillers, baths (*e.g.*, water), dry baths, hot plates, temperature-controlled rooms, ovens, thermo-electric devices, such as devices employing Peltier-effect cooling and/or joule-heating, and environmental chambers. The temperature of the samples can be controlled by heating or cooling the thermally conductive shelves.

The thermal cycling system can be used to simply incubate an array of samples at a specific temperature for a particular time (isothermal incubation), or can be used to cycle their temperatures, *e.g.*, to vary their temperature as a function of time. When employed, thermal processing comprises varying the temperature of the contents of each vial in a controlled cycle, usually a heating period followed by a cooling period. Heat transfer through the blocks that hold the arrays of containers changes the temperature of the containers. Thus, when thermal processing is used to process the samples, the blocks used should allow heat transfer between a heating/cooling source (*e.g.*, thermally controlled shelves) and the sample containers (*e.g.*, vials).

FIGS. 4A-4B illustrate a specific example of a thermal cycling system, which comprises temperature-controlled shelf assemblies **56**, which are also referred to herein as "hotels." A number of hotels can be arranged to a number of different blocks. **FIG. 5** provides an illustration of one example of a thermal cycling system, which comprises 18 hotels **56** in a semi-circular arrangement around a robotic arm **55**. In a preferred embodiment, a hotel comprises twelve shelves **58** arranged in a vertical fashion as shown in **FIG. 4A**, held in place with supporting members and incorporating locating features for securing the assembly in the desired position. In a preferred embodiment, each individual shelf **58** contains an internal loop through which a liquid, such as water, is circulated to control the shelf temperature. The loops are piped to a bath, *e.g.*, a water bath acting as the cooling/heating source. Finally, the thermal cycling system can optionally include an environmental-control enclosure **57** that regulates the humidity and/or ambient temperature of the air surrounding the blocks, preventing condensation on the containers and other components. One embodiment of an environmental control system **57** is shown in **FIG. 5**. Alternatively, the thermal cycling system can be located in an environmentally-controlled room.

In a specific embodiment of the invention, different water baths (which may also employ various other fluids for conducting heat or cold to the samples) allow for the

processing of multiple blocks at different temperatures. The blocks are located in hotels that are connected to the baths, the temperatures of which are computer controlled. In this embodiment, computers also record the heating/cooling time and temperature for each assembly of shelves, or "hotel." Because each block contains a plurality of sample containers, each of which is identifiable by is location in the block and/or the use of a bar code or other identifier, the conditions to which each sample in a given hotel is exposed is recorded and tracked by computer.

The processing of samples or arrays of samples can involve more than simply subjecting the samples to a particular temperature or range of temperatures. For example, the samples can be exposed to other environmental conditions, such as humidity, using an environment-controled room. As shown in **FIG. 5**, environment control is achieved in one embodiment of the invention using an enclosure **57** that surrounds the shelf assemblies **56**, and is connected to a supply of air that has been treated to provide the desired humidity level inside the enclosure.

Samples in an array can be processed in any number of ways. For example, samples in an array can all be subjected to the same temperature for the same amount of time, or can be processed individually using, for example, robotic techniques. For example, a solvent or antisolvent can be added to just one or a few of the containers held in a block with the aid of automated dispensing devices and robotic arms, such as that shown in **FIG. 5**.

Samples can also be subjected to a combination of different processes. For example, in what is referred to as a "mixed-mode" crystallization process, more than one processing mode is applied to samples in an array either serially or in parallel. For instance, thermal processing (described above), followed by anti-solvent addition to the container(s) and/or partial or complete evaporation of the volatile contents of the container(s) can be used to facilitate crystallization of a compound-of-interest. Here, the term "anti-solvent" refers to a solvent in which the compound to be crystallized has very low solubility. An evaporation process entails allowing the sample solvent systems to evaporate and may involve flowing a dry, inert gas over the samples and/or heating the samples to an extent and for a time sufficient to effect concentration of the compound-of-interest in the sample. In a specific example of mixed-mode processing, a thermal process is followed by an evaporative step in which the sample vessels are opened (uncrimped) and dry nitrogen is blown over the surface of the samples to promote evaporation of the solvent to an extent and for a time sufficient to allow crystallization. In another example of mixed-mode processing, a thermal process is followed by addition of an anti-solvent to the sample vessels in an amount sufficient to allow

crystallization. In still another example of mixed-mode processing, a thermal process is performed on duplicate sets of sample formulations followed by an evaporative step on one set and anti-solvent addition to the other set. A mixed-mode crystallization process may conclude with an incubation step, where the samples are incubated at a temperature and for a time sufficient to allow crystallization. Any combination of individual process steps (*e.g.*, thermal, anti-solvent addition, and evaporation) may be used in serially or sample arrays may be split to allow different process modes to be used in parallel.

Visual inspection of the samples is preferably done at least once during their processing (*i.e.*, their exposure to one or more chemical or environmental conditions). Such inspection can occur at any time before, during, or after the processing of the samples, and is preferably done using automated means. For example, a robotic arm **55** as shown in **FIG. 5** can be used to removed the blocks **60** from the shelves **58** and transfer them to an imaging, or vision station, such as that which is described below in Section 4.3 and elsewhere herein. Depending on the result of the imaging, the block can then be replaced onto a shelf in the thermal cycling system, or its containers can be separated, rearranged into new blocks, or removed entirely for more detailed (*e.g.*, spectroscopic) analysis. As mentioned above, the location and processing history of each sample is preferably tracked and recorded, so that it can be located, analyzed, and reproduced at any time. Because each container can be imaged separately from others in the array to which it belongs, this invention allows the rapid identification of samples that can be further processed or removed for detailed analysis even when such samples are just a few of hundreds or even thousands of samples being processed.

In one embodiment of the invention, the processing of one or more samples in an array is stopped at a specific time using what is referred to herein as a "quenching station." It is at such a station that the condition(s) to which a sample is exposed are removed. For example, if the condition to which a compound-of-interest has been exposed involves contact with a particular solvent, the samples can be quenched by extracting any fluid component that remains in each container. This can be accomplished by puncturing the seal of the container, or tube, with a needle that can extract the liquid from the tube and provide a relief path through which air can flow into the tube, so as not to create a vacuum. In addition, samples can be air-dried after removal of the liquids in a vial by using a similar needle assembly to punch through the septum and inject dry air into the vial for a specific amount of time. The dry air (or other gases) removes remaining liquids from the sample through evaporation, and vents them outside the vial. As with sample preparation, quenching can be automated, and can be triggered by a human operator or by computer.

4.3. SAMPLE IMAGING

A result of conducting a large number of small scale experiments using various processing methods creates the need to interrogate or inspect each of the samples in the containers for the presence (or absence) of solid forms or other products of interest. Although visual inspection can be done manually, preferred embodiments of the invention utilize what is referred to herein as a "vision station," which is an automated system that allows for the rapid and efficient imaging and screening of samples. Preferred vision stations are designed for the analysis of samples contained in tubes and blocks-type arrangements, as discussed above in Section 4.1 and elsewhere herein.

In one embodiment of the invention, the vision station comprises a device for capturing an image of small particles, such as a microscope/camera system with a highly magnifying lens to capture images of small (down to sub-micron) particles onto a CCD such as the Canty Particle Size Vertical Imaging Microscope (JM Canty Inc., Buffalo, NY USA). Another example is the published report from December 2000 on image analysis of protein crystals: *An optical system for studying the effects of microgravity on protein crystallization*, Alexander McPherson *et al.*, application note from American Biotechnology Laboratory, December 2000 issue, which is incorporated herein in its entirety by reference.

Depending on the use to which the invention is put, sample imaging can be used to determine the presence of a solid form in a sample or container. Alternatively, the absence of solids can be also be detected. Consequently, vision stations of the invention can be used to determine the stability of liquid formulations (*e.g.*, drug formulations for intravenous administration to patients) and the stability of a formulation in a simulated body (*e.g.*, gastric) fluid.

Samples can be imaged at any time after their preparation. Consequently, imaging information can be used to determine whether or not a sample should be processed, how it should be processed, and whether or not it should be subjected to more detailed, (*e.g.*, spectroscopic) analysis.

A typical vision station of the invention comprises a light source and a camera. A suitable camera can be any unit capable of yielding photographic images of the contents of containers, *e.g.*, the presence or absence of solids or solid forms, but is preferably capable of digital capture. In a preferred embodiment of this invention, a charge coupled device (CCD) camera provides adequate sensitivity, but other digital capture devices may also be used. The light source is selected based on the types of containers being used and the design of the experiment. Examples of light sources include, but are not limited to, visible light, laser light

of varying wavelengths, monochromatic laser, plane-polarized, or circularly polarized light. In an example embodiment, the light source is white light from one or more tungsten lamps. Depending on the mode of application of the vision station, light can be brought in from the top of the array, the bottom, or from the side. Blocks containing removable containers allow improved access by light to the sample due to the ability to elevate the containers from the block, either by hand, or using an automated means.

In one embodiment, the vision station system is adapted for use with the tubes and blocks system. In this embodiment, the vision station system comprises a camera, a light source, and, optionally, a mechanism to elevate containers (*e.g.*, tubes) from a block, thus presenting the containers to the camera. The mechanism to elevate containers from a block can lift containers out of a block individually, or in groups, including without limitation, lifting all the containers in one or more rows or columns of a block at the same time, and preferably, lifting all the containers in one row or column at the same time. Additionally, the system can employ software to capture, store, and analyze images and digitally flag or select tubes containing contents of interest, *e.g.*, solid forms, in a series of images. Furthermore, the vision station system may optionally comprise a database for warehousing of the results and collation of information on the identity, composition and history of samples in order to allow further detailed analysis of the combined data.

Ultimately, the vision station system enables the automatic selection of specific samples (or containers containing samples) from an array based on their appearance. Advantages of the vision station system include, but are not limited to, speed of acquisition coupled with the details of the solid form, such as gross crystal habit, color, form, and location of solids (*e.g.*, crystals) in a container. Such information about where solid formation occurs (such as where a crystal nucleates, *e.g.*, at the air-liquid interface or in the bulk solution), and shape of the crystals or precipitate is useful in studying and controlling crystallization. The vision station also provides many automation opportunities (both in hardware as well as in software analysis of images) and the ability to capture a variety of data regarding the detailed physical form of the compound-of-interest (*e.g.*, its crystallinity, amorphous character, physical stability, and size range information). In terms of speed, embodiments of the vision station system can observe 96 sample tubes in less than one minute, and the image capture is rapid (on the order of 30 milliseconds with current digital camera technology).

In a specific embodiment of the invention, the vision station system can accept different arrays or blocks of containers for analysis in rapid succession. Using the vision

station system, the information obtained can include: (1) detection of solids based on illumination (*e.g.*, white light) and image capture; (2) observation of birefringence (backlit crystalline samples seen with the help of cross-polarized light); (3) observation of nano-particle presence (using laser beams at various angles to the camera lens); and (4) temporal information (nucleation kinetics and kinetic stability of colloidal suspensions toward growth and phase separation are two examples). In addition, automated exemplary and example machine vision algorithms further enhance the utility of the system by obviating the need for a user to manually select tubes that are of interest.

In another specific embodiment, the vision station system is adapted to process blocks that contain about 96 containers in an arrangement of 8 rows of 12 columns. **FIG. 6** illustrates the logic used in one embodiment to address solid form generation using the vision station approach. The embodiment in the flowchart **80** involves tubes in blocks, but the process can be adapted for use with other container systems. This embodiment comprises a vision station analysis **82**, which includes optical inspection of vials holding samples. If no solids are detected **84** in the vial, a determination is made whether or not that particular reaction is of further interest **88**. If it is not of further interest, the experiment may be stopped **92** for that particular vial. If a sample in a vial is still of interest, it may be returned to the block 94 and sent back for further processing, such as in the thermal cycling system. The process may then be repeated as to that vial. Alternatively, it may be removed without further processing. If the experiment is designed to detect solids, vials that contain solids are sent to a re-arraying process **86** whereby multiple vials with solids present are grouped together in the same output block. At every step, the address of each vial is tracked and updated if necessary. Such tracking can be done using various methods known to the skilled artisan, including without limitation using bar codes. Optionally, the entire output block can then be sent for detailed (*e.g.*, spectral) analysis **90**. The output block is preferably entirely filled, but it need not be.

FIG. 7A shows a schematic diagram **100** of a vision station. A block **60** containing an array of vials **50** is placed before an imaging device **104**. While it is recognized that a camera can be oriented underneath the block so as to be capable of viewing the contents of the vials through the access holes in the bottom of the block, the preferred embodiment contemplates raising the vials at least partially out of the blocks into view of the imaging device. Thus, **FIGS. 7B** and **7C** are drawings of a vision station in operation, showing a side view **111** (**FIG. 7B**) and a perspective view **113** (**FIG. 7C**) of the block **60**, wherein a lifting

mechanism **115** lifts an entire row of vials **50** at the same time. Alternatively, vials **50** can be lifted one by one.

As shown in **FIGS. 7A-7C**, vial **50** is illuminated by a light **102** that can be placed in a variety of locations to light up different portions of the vial **50**, depending on where in the vial **50** illumination is desired. The level of illumination is determined by inspecting the resulting images for the desired contrast and is controlled by the operator adjusting the level or voltage until the desired contrast is obtained. Alternatively, the level of illumination can be automatically adjusted using appropriate sensors and/or algorithms. The resulting illumination provides sufficient contrast for an image or picture to be captured. Various software **108** can be used to capture the image, such as Component Works IMAQ Vision (National Instruments). In a preferred embodiment, the image capture software is integrated into a custom VB software. The camera **104** then takes a picture of the vial **50**. The picture is then stored on a computer. A hardware card, such as National Instruments Image Capture Card, model number PCI-1422, is used to capture the image in conjunction with image capture software **108**. A custom software application then displays the picture of the vial and the vial can then be designated as containing various results, such as, but not limited to, solids, lack of solids, sediment, phase separation. The pictures and the vial **50** designations can then be stored in a database **110**. The process is repeated for all of the vials **50** in a block **60**, and for all the blocks in a given experiment run or design. The vials can be all processed at one time, or it can be done intermittently.

A preferred embodiment of the vision station system comprises a camera **104**, preferably a CCD camera, for example, a CCD camera manufactured by Roper Scientific (model MegaPlus ES:1.0) (now Redlake MASD , Inc., 11633 Sorrento Valley Road, San Diego, CA 92121 USA) with an 9x9 mm image array with a total of 1008x1008 pixels. Another suitable source of imaging cameras is Spectral Instruments, Inc., Tucson AZ that provides a CCD camera that can be cooled to –50°C. Alternatively, image plate technology based on CMOS can be used for obtaining images, but CCD is the preferred capture mechanism.

In one implementation, an area of the width of roughly 72 mm is observed when 8 tubes (a row at a time) are pushed out of a block for vision analysis, although, for instance, tubes may be viewed in groups of fewer than 8 such as single tubes or two tubes per captured image. This observed area leads to a pixel resolution of about 70 microns/pixel. A resolution range from about 5 to 1000 microns is useful in the many embodiments of this invention, since most organic crystalline materials in a powdery state range in particle size from a few

microns to hundreds of microns. Single crystals are often a few hundred microns on the shortest edges, while on the other hand extreme colloidal particles, such as titania (TiO_2) and silica (SiO_2) can be stably prepared in the nanometer size range.

The vision station can be used to identify amorphous, as well as crystalline, solids. The amorphous form can be of significant interest with regard to certain compounds-of-interest, such as, but not limited to, increased solubility relative to crystalline forms. Generally, amorphous forms of a given compound are thermodynamically unstable compared with crystalline forms, but can be rendered kinetically stable toward physical form change, *e.g.*, as a glass. Amorphous particles are typically irregular in size, and the material lacks the property of optical birefringence. This is defined as the ability of most crystalline materials to interact with polarized light by changing the direction of the polarization as it passes through the crystals. Plane-polarized light is generally rotated upon traveling through a crystalline material. If the light is subsequently sent through an analyzing filter (this is another plane polarized filter where the polarization direction is 90 degrees perpendicular to the first filter) at a right angle to the plane-polarizing filter on the light source, the rotated light escapes the analyzer. Therefore true crystals appear as bright spots on a dark background. Conversely, amorphous disordered materials generally do not rotate plane-polarized light such that minimal light (equal to background) escapes the filter resulting in a dark image. It may be advantageous to look for the presence or absence of crystallinity in this way, and by comparison of birefringence image with the plain image rather than simply looking for the presence of solids.

The lighting used to capture white light images of elevated tubes is flexible, in that it can be brought in (a) from the top of the tubes (if the top is either open or any seal is transparent), and (b) from the side of the tubes, behind the camera. The latter is referred to as backlighting and this approach is required when one wants to capture birefringence information. In principle, the light can be brought in at a number of angles, but the preferred orientations are either vertical or horizontal. The lighting can be provided by fiber optics (for example, NT39-366 from Edmund Industrial Optics,101 East Gloucester Pike, Barrington, NJ 08007), although white light strips (for example, Stocker Yale, Imagelite brand) can also be used. Various polarizing filters can be obtained from a number of commercial sources, *e.g.*, polarizing filters, such as NT45-669, are available from Edmund Industrial Optics.

FIG. 8 illustrates an embodiment of the vision station adapted for the detection of birefringence. On the left side, a pair of tubes with water **112** and another pair of tubes with varying amounts of glycine crystals **114** are shown with backlighting without a polarizing

filter on the camera lens. On the right are the same samples **116** and **118** with a polarizer in place. With use of color images, one can capture polychromism (*i.e.*, multi-color crystals) information from the experiment with a suitable camera, or simply run the analysis with black and white images and look for bright pixels. In addition, a quarter-wave retarder filter can be used to confirm the presence of crystals by causing a color shift when the filter is applied.

FIG. 9 shows the use of a combination of white light and laser scattering. A laser beam **124** (which can be of any color, such as red, green, or blue) can be brought into proximity with a tube or vial **50**. Single tube analyses are typically preferred, due to some scattering and diffusion of the laser beam in cases where one attempts to send the beam **124** through several tubes consecutively (**FIG. 9B**). The laser beam, which can be generated by any number of laser devices such as with a He-Ne Class II laser pointer at <1 mW power, will interact with sub-micron particles inside the tubes and the radiation is scattered, resulting in a contiguous trail of laser light through the tube. If no significant colloidal component is present (the sample is a true solution) no such trail of laser light will be observed in the image. Using this application, the vision station system with the laser beam can be used to obtain kinetic information regarding colloidal stability (*i.e.*, how long it takes a suspension to settle or ripen to microcrystals), solution physical stability (how stable is a solution toward nucleation), or phase segregation.

Another embodiment of the invention utilizes laser light at an angle different from 90 degrees (*e.g.*, at a 45 degree angle) relative to the camera lens. This is shown in the example of **FIG. 10**, where the image **126** in panel (a) clearly shows a contiguous path of laser light due to the presence of the colloids. In contrast, the image **128** in panel (b) shows a single point of scatter on the right side of the tube (where the laser beam hits the tube). This effect is due to partial scattering of the laser light by the glass, and becomes more pronounced in the image as the angle between the camera lens and the laser beam is decreased.

FIGS. 11 and **12** show flow charts for the logic used in specific vision station systems for the detection of birefringence **130** and laser light interrogation **132**, respectively. The funnel widths roughly represent the number of samples at a given stage of the experimental workflow. The charts illustrate how a vision station system can facilitate analysis of crystallinity or lack thereof in a set of solid forms (**FIG. 11**), and also allows analysis of nano-particulate and true solutions along with the stability of each (**FIG. 12**).

In a preferred embodiment, the analysis of images obtained by the vision station is automated. For example, software (*e.g.*, National Instruments IMAQ VISION software) is

employed in image acquisition and analysis. When image analysis is performed manually, an operator flags the samples that satisfy the criteria used in the particular experiment (*e.g.*, which ones contain a solid) using a software interface. Such software can perform a variety of function, such as, but not limited to, automated capture and storage of images, creating and storing logic for each sample (*e.g.*, which ones contain solid, was a sample in solution at the start of the experiment), and ultimately containing algorithms for time-based measurements as well as automated isolation of containers that satisfy given criteria. Such software can also inform the user which samples are of interest, and facilitates the re-array of hit tubes from the source block into a destination block for further off-line processing or characterization. Preferred software provides an actual image of vials that allows a user to observe and manually select vials of interest for further processing.

In another specific embodiment of the invention, the vision station system further comprises a means of determining the optimal laser light configuration relative to the tubes for interrogation of colloidal suspensions (*e.g.*, as to the size of the particles they contain). In another embodiment, the vision station system comprises a means of optimizing the capture of birefringence information, including the investigation using a quarter wave plate and other filters in concert with plane or other polarizers to ensure that light scattering is not interfering with image analysis and interpretation.

In a specific embodiment of the invention, once a number of blocks have been processed through the vision station system, there will be one or more output blocks holding vials containing solids. Optionally, in a preferred embodiment, these blocks are then processed further (*e.g.*, moved to a quenching station) as described above in Section 4.2 and elsewhere herein.

4.4. SPECTROSCOPIC DATA COLLECTION AND ANALYSIS

In a typical embodiment of the invention, one or more samples in an array are analyzed using spectroscopic techniques. In preferred embodiments of the invention, the sample(s) that are analyzed have been screened or selected (*e.g.*, using the methods or devices described above in Section 4.3) from an original array of samples. For example, the vision station can be used to identify samples that contain solids, and the contents of those samples are then analyzed further using spectroscopic techniques.

The specific analysis done will depend on the purpose to which a particular embodiment of the invention is put. For example, if the invention is used to prepare solid forms of compound-of-interest, the solids that have been identified in samples can be

analyzed to determine their chemical and physical form, *e.g.*, whether they are salts or solvates (*e.g.*, hydrates) of the compound-of-interest, whether or not they are crystalline, and, if they are crystalline, the nature of their crystal form (*e.g.*, their crystal structures). Spectroscopic analysis can also be used to determine if any of the compounds in a sample (*e.g.*, the compound-of-interest) decomposed or reacted with other compounds in that sample.

Spectroscopic techniques can also be used to identify samples that share one or more characteristics. For example, if a solid compound-of-interest can exist in more than one solid form, and each of a plurality of samples comprises solid compound-of-interest, it may be desirable to identify which samples contain compound-of-interest of which form. The grouping of samples as a function of a particular characteristic (*e.g.*, a spectral characteristic unique to a particular solid form) is referred to herein as "binning." Such binning provides a means of avoiding unnecessary duplication of further experiments. For example, if a group of samples are binned based on a particular spectral characteristic which corresponds to a previously unknown solid form of the compound-of-interest, further analysis of that solid form need not require a detailed analysis of each sample in the group.

Examples of spectroscopic techniques that can be used to bin or analyze samples are numerous, and will be readily apparent to those skilled in the art. Some specific examples include, but are not limited to, optical absorption (*e.g.* UV, visible, or IR absorption), optical emission (*e.g.*, fluorescence or phosphorescence), Raman spectroscopy (including resonance Raman spectroscopy), nuclear magnetic resonance spectroscopy (*e.g.*, single and multi-dimensional ^{1}H and ^{13}C), X-ray diffraction (*e.g.*, powder X-ray diffraction), neutron diffraction, and mass spectroscopy. For the sake of convenience, other methods of analysis are encompassed by the term "spectroscopic technique," as it is used herein, include, but are not limited to, microscopy (*e.g.*, light and electron microscopy), second harmonic generation, circular dichroism, linear dichroism, differential scanning calorimetry (DSC), thermal gravimetric analysis (TGS), and melting point. Preferred embodiments of the invention utilize Raman spectroscopy.

4.4.1. Raman Spectroscopy

The use of Raman spectroscopy for the high-throughput screening and/or analysis of multiple samples is believed to be novel, particularly in view of the relatively low intensity of Raman scattering as compared to other spectroscopic techniques. When coupled with the devices and techniques disclosed herein, however, Raman spectroscopy has been found to be particularly useful in the high-throughput screening and analysis of samples.

The Raman spectrum of a compound can provide information both about its chemical nature as well as its physical state. For example, Raman spectra can provide information about intra- and inter- molecular interactions, inclusions, salts forms, crystalline forms, and hydration states (or solvation states) of samples to identify suitable or desirable samples, or to classify a large number of samples. With regard to the hydration states of molecules, methods and devices of this invention, particularly the binning methods discussed in more detail below, allow their determination *in situ*.

Raman spectroscopy can also be used in this invention to examine kinetics of changes in the hydration-state of a sample or compound-of-interest. Moreover, the ability of Raman spectroscopy to distinguish, in certain situations, forms with different hydration states is comparable to X-ray diffraction, thus promising specificity and sensitivity. The lack of a strong Raman signal from water, a common solvent or component in preparations allows collection of Raman data *in-situ* in a manner relevant to many applications.

This invention also encompasses the use of Raman spectroscopy to determine the amount of a compound-of-interest that is dissolved in a particular sample. Advantageously, it has been discovered that for many compounds-of-interest and solvents, a correlation between the amount of compound-of-interest dissolved in a liquid sample and certain characteristics of its Raman spectrum can be obtained using one solvent, yet can be applied to the high-throughput analysis of samples prepared using a variety of other solvents.

These and other aspects of the invention are made possible by the utilization of several devices and methods described herein, which overcome problems inherent to Raman spectroscopy that would otherwise limit its usefulness as a high-throughput analytical technique. Examples of such problems include, but are not limited to, weak signals, background (*e.g.*, solvent) emissions, and signals due to other solids or liquids in a sample, as well as the sample container itself.

Improvements in reproducibly obtaining Raman spectra for samples of interest include rapid and sensitive spectra acquisition and rejection of background noise. The strength of Raman emissions is improved by the use of lasers to excite the target substance. Use of a carefully selected wavelength also results in resonance Raman spectra. Sample preparation techniques resulting in adsorbing of a target to a surface further increase Raman signals, although such preparation is not always possible or desirable in the case of *in-situ* data collection. Since the strength of the Raman signal can vary depending on many factors, it is important to use on-line data analysis in order to determine when a sufficient quality and quantity of data have been collected to meet the goals of the measurement (*e.g.* a prescribed

signal-to-noise threshold). Of course, optical amplifiers further improve sensitivity and specificity. Each of these techniques or process steps may be used alone or in combination.

Filtering techniques encompassed by the invention that can be used to reject noise include but are not limited to temporal, spatial, and frequency domain filtering. Spatial filtering requires collecting emissions from a small area to reject noise from surrounding sources. Such confocal techniques, for instance with the target in the focus of an objective and/or using a pinhole arrangement, allow scanning of a target to reduce unwanted noise due to emissions from the material surrounding the target area.

The invention also encompasses temporal filtering, which rejects or accepts signals received in a particular time window. In the case of Raman spectra, temporal filtering relies on the different times taken for emission of Raman spectra and the background fluorescence spectra. Notably, Raman emissions, although weak, can be detected much earlier than fluorescence following excitation. Furthermore, fluorescent radiation continues over a significantly longer period, thus making possible selection of time windows for collecting Raman signal with a higher S/N ratio than otherwise. An example of such filtering is provided by Matsousek *et al.* in "*Fluorescence suppression in resonance Raman spectroscopy using a high-performance Picosecond Kerr Gate*," in *J. Raman Spectroscopy*, vol. 32, pages 983-988 (2001). The Kerr gate realized by Matousek *et al.* exhibits a response time of about 4 picoseconds, thus allowing collection of Raman emissions during a window of 4 picoseconds following an exciting laser pulse. This example should be regarded as illustrative and not limiting as to temporal considerations in collecting and filtering spectra in possible embodiments since other gates, including virtual gating techniques are also intended to be within the scope of the claimed invention. Such filtering techniques, which can be used separately or together, can be augmented with mathematical filtering, *e.g.*, convolution with the characteristic shape of a Raman line to further reduce the noise and reject unwanted frequencies and emissions.

In another aspect, the invention encompasses the use of polarized excitation and detection. Raman scattering emissions are sensitive to the orientation of the polarization of the exciting light relative to the molecules being examined. If the exciting light (typically from a laser) is polarized and the molecules in a crystal have fixed orientations, the Raman signal varies as a function of the orientation of the crystal. This property, while useful for detecting and evaluating crystalline samples, presents challenges in collecting representative Raman spectra due to the change in the amplitude of individual lines. The use of spectral binning, which is discussed elsewhere herein in more detail, can be used to overcome such

challenges. Following the collection of a plurality of spectra that are, optionally, preprocessed to remove contaminating signals, as described more fully below, it is possible to identify peaks in each of the spectra. Optionally, from these identified peaks of the spectra it is possible to generate, for instance, a peak height or binary spectra reflecting the peak positions. The use of binary spectra reduces the computational overhead in binning and otherwise interpreting the data while taking into account variations due to orientation and the like. Filtered raw spectra, peak height spectra generated from identified peaks of filtered raw spectra, or binary spectra may be used to calculate similarity scores using any suitable metric, and the similarity scores allow binning of the spectra in accordance with various clustering techniques.

4.4.2. Data Collection

Spectroscopic data can be obtained for one or more samples by manually removing the containers that contain them from the block holding them, and presenting the containers to the particular analytical device being used (*e.g.*, Raman spectrometer). Preferably, a mechanical system (such as an automated robotic arm) is used to select, or "cherry-pick," particular containers (*e.g.*, those identified as satisfying certain criteria by the vision station) from the block(s) that contain them.

In a specific embodiment of the invention used to detect and/or characterize solid forms of compounds-of-interest, a container is presented to a Raman spectrometer, and is imaged down the centerline at predetermined x, y positions. At each x, y position, two predetermined z positions are selected in order to focus imaging on the upper and lower inside face of the container (*e.g.*, the upper and lower inside glass faces of a glass tube). Preferably, at least one position is used to focus imaging. This image acquisition step is repeated for different angles of rotation of the container until the entire inside surface of the container (*e.g.*, glass tube) is imaged. After each image capture, an analysis is performed to determine where the "areas of interest" in a container are, where "areas of interest" can include solids or solid forms (*e.g.*, crystals), and in some instances, any remaining droplets of solution or solvent.

A vision algorithm designed to automatically detect areas of interest (*e.g.*, solid forms) in a container carries out the following: 1) locates or recognizes the presence or absence of a container; 2) locates the meniscus, if any, of the sample in a container; and 3) searches the area between the meniscus and the bottom of the container for particles, solids, solid forms, or other areas of interest.

After identifying areas of interest in a container, the Raman stage is moved to the center of the excitation source (*e.g.*, laser) on to each of areas of interest in a container, and the Raman detection apparatus is focused using manual or automated means.

In one embodiment, auto-focusing of the Raman spectrometer can be performed. One way in which auto-focusing can be performed is by taking a series of Raman spectra at various z positions (to change the focus), for each x, y position representing an area of interest in a container. The one with the "best" Raman signal is marked, wherein the "best" Raman signal is defined by predetermined criteria, including, for example, by filtering each spectrum for a location and taking the maximum peak other than the normal peak associated with the effects of the container (*e.g.*, glass tube). The resulting series of "best" Raman spectra for various areas of interest in a container can then be sorted based on similarities, and clustered into bins with spectra from other containers in an experiment. Automated focusing of a Raman spectrometer can result in a series of "best" Raman spectra for various areas of interest. These spectra can be sorted to distinguish droplets of solution or solvent from solids and clustered with data (spectra) from the other containers in the experiment.

When multiple spectra are obtained, one or more of the following can be also done: (1) find the one "best" spectra of a set of spectra for an area of interest or a solid form, with best being defined in a predefined way, including without limitation, highest peak signal, highest average signal, best S/N ratio, most peaks, and the like; (2) construct an average spectrum of all the spectra for an area of interest or a solid form, and use this spectrum in further processing; (3) construct an "agglomerated spectrum" that contains the highest peak of the set for every peak window, wherein a peak window is defined as a region in which peaks are considered to be the same; and/or (4) keep all of the spectra and perform downstream analysis on all of the spectra.

In processing (*e.g.*, sorting and clustering) spectral data, the knowledge that several spectra come from each sample can used to score the clustering results, or the labeled spectra can be used to influence the clustering run. For example, a k-means clustering run can be altered in the following manner: for each step of the k-means run, cluster assignments are made in the traditional sense, such that each point is assigned to the cluster with the nearest centroid, resulting in precluster assignments that are not the final assignments for the step; the precluster assignments for all points coming from the area of interest or solid form are then compared, and the most popular cluster assignment is assigned to all of the points in the group as the final assignment; and new centroids are determined from these final cluster assignments.

4.4.3. Data Analysis

In particular embodiments of the invention, spectroscopic data is processed using what is referred to herein as a "spectra binning system," which allows the rapid analysis and identification of samples in an array by creating, for example, a family or similarity map. Preferred embodiments of the spectra binning system comprise a hardware-based instrumentation platform and a software-based suite of algorithms. The computer software is used to analyze, identify and categorize groups of samples having similar physical forms, thus identifying a group from which the operator, or scientist, can then select a few samples for further analysis. This selection can be performed independently by the scientist or using an automated means, such as software designed to automatically select samples of interest. Although, many applications made possible by the spectral binning system will be apparent to those skilled in the art, preferred systems of this invention is used to identify and characterize samples or compounds-of-interest. Particular binning and analytical methods useful in the invention are disclosed in U.S. patent application no. 10/142,812, filed May 10, 2002, the entirety of which is incorporated herein by reference.

The spectral binning system is generally used in this invention to detect similarities in the properties of a plurality of samples by observing their binning behavior. Thus, the number of forms of a substance can be estimated by binning spectra. The plurality of samples are examined with a device for generating a corresponding spectrum of acceptable quality, *i.e.*, sufficient S/N ratio. Spectral peaks or other features are next identified to obtain a binary fingerprint. Advantageously, the spectra are compared pairwise in accordance with a metric to generate a similarity score. Other comparisons that use more than two spectra concurrently are also acceptable, although possibly complex.

One or more clustering techniques can be used to generate bins that are preferably well defined, although this is not an absolute requirement since it is acceptable to generate a reduced list of candidate forms for a given substance as an estimate of the heterogeneity of the substance's structure. Advantageously, the generation of bins facilitates the ready evaluation of structure heterogeneity among samples. For instance, frequency, frequency shift, amplitude, and other similar measurements based on Raman spectra are often limited by the lack of suitable standards. However, the number of bins generated from evaluation of Raman spectra obtained by sampling a substance of interest is a measure that does not directly depend on having a good standard.

The invention also encompasses the use of hierarchical clustering to represent the data in the form of a similarity matrix having similar spectra/samples listed close together. Such a

similarity matrix may be sorted to generate similarity regions along a diagonal. The resulting sorted similarity matrix may be used as a basis for setting the number of clusters for k-means clustering or other clustering techniques based on a specified number of clusters such as Gaussian Mixture Modelling.

Advantageously, although the clusters are actually in higher dimensional space, they can be projected into 2 or 3 dimensional space and visualized. Therefore, the binning procedure allows for both steady state and kinetic evaluation of states (*e.g.*, hydration states, crystalline states, and other states, or forms, that can vary over time). This method is well suited for such measurements since individual Raman spectra can be collected rapidly (*e.g.*, in a few seconds). Preferably, the turn-around time for generating a spectrum and assigning the spectrum to a bin is less than about two minutes, one minute, ten seconds, or one second. Moreover, limited real time processing is often possible if an acquired spectrum is to be assigned to existing bins, or, in a preferred embodiment of the invention, a library of binned spectra is updated with newly acquired spectra. In a preferred embodiment, newly acquired spectra from a single sample may all be binned into a single bin based on a majority of them being more related to the single bin in accordance with a metric, such as those discussed below and elsewhere herein.

Once the spectra from all of the samples to be analyzed have been collected, they are processed by a series of algorithms. These algorithms facilitate the binning of sample spectra according to one or more spectral features. Examples of such features include, but are not limited to, the locations of peaks, peak shoulders, peak heights, and peak areas. In a preferred embodiment, the spectral binning process bins spectra based on the locations of their scattering peaks and peak shoulders, expressed as wavelength or Raman shift (cm^{-1}).

In the spectra binning system, the collected spectra can be binned using the raw or filtered spectra, peak height spectra generated using peaks selected from the raw or filtered spectra, and binary spectra generated using the raw or filtered spectra.

FIG. 15 represents the computational process applied by a specific embodiment of the spectra binning system. As shown in the flow chart **270**, the process can be divided into preprocessing **271**, peak finding **275**, similarity matrix calculation **281**, spectral clustering **283**, and visualization **285** stages An optional binary spectra generation stage **279** can also be used. Each of these stages, which are applicable to the analysis of data obtained using a variety of spectroscopic techniques. For the sake of convenience, however, each is discussed in more detail below in reference to Raman spectroscopy.

4.4.3.1. Preprocessing

In specific embodiments of the invention, preprocessing is used to eliminate artifacts of the Raman spectra that are not caused by Raman scattering. Preprocessing can also be used to make the Raman scattering peaks as sharp as possible. Raman spectra often contain large fluorescence peaks spread over a broad spectral range and much smaller, narrower peaks caused by scattering from containers (*e.g.*, glass) and instrument noise. Several different filtering techniques can be used to eliminate such noise including, but not limited to, Fourier filtering, wavelet filtering, matched filtering, and averaging. A preferred method uses a matched filter approach, wherein the filter kernel is a zero-mean, symmetric product of sinusoids matched approximately to an average Raman peak width. The specific form of the matched filter is given by the following equation:

$$k[n] = \sin\left(\frac{3\pi n}{N-1}\right) \cdot \sin\left(\frac{\pi n}{N-1}\right)$$

where N is the length of the kernel. Preferably, the matched filter equation includes a normalization term:

$$k[n] = -\sqrt{\frac{4}{N-1}} \sin\left(\frac{3\pi n}{N-1}\right) \cdot \sin\left(\frac{\pi n}{N-1}\right)$$

The normalization factor ensures that the magnitude of the "passed" peaks in a filtered signal are about the same as the magnitude of the original peaks, and that all peaks point in the right direction. In one embodiment, filtered points having a value less than zero are automatically set to equal zero.

In a specific embodiment of the invention, the bandwidth of the main kernel peak is set to be equal to or slightly smaller than the bandwidth of an average Raman peak. When matched filters of this type are viewed in the Fourier domain, they perform as bandpass filters, almost completely attenuating low and high frequency spectral components. Furthermore, with the bandwidth of the filter kernel chosen to be equal to or slightly smaller than the average Raman peak bandwidth, this filter detects peaks that are very close to each other. A raw, unfiltered spectrum will often display two close peaks as a main peak with a "shoulder" on one of its sides. After a matched filtering step, though, the shoulder will often be distinguished as a separate peak. This separation is useful for the selection, or finding, of peaks used for binning.

An example of the effect of such filtering means is provided in **FIG. 16.** Specifically, **FIG. 16A** shows Raman intensity plotted as a function of Raman shift (cm^{-1}) for an empty

glass vial. The resulting waveform shows the pattern of absorbance present. **FIG. 16B** shows a Raman intensity of a fluorescent sample as a function of Raman shift. **FIG. 16C** shows the same pre-filtered plot as that of **FIG. 16B**, but also shows the corresponding filtered spectra after the fluorescence has been removed.

4.4.3.2. Peak Finding

The process of finding peaks in a spectrum is an essential aspect of many spectral processing techniques, so there are many commercially available programs for performing this task. The many variations of peak finding algorithms can be found in the literature. An example of a simple algorithm is to find the zero-crossings of the first derivative of a smoothed or unsmoothed spectrum, and then to select the concave down zero-crossings that meets certain height and separation criteria. In a specific embodiment, the peak finding function available in the software provided with the Almega dispersive Raman spectrometer (Thermo Nicolet, OMNIC software) is used. This function allows the threshold and sensitivity values to be set by the user. The threshold sets the lowest peak height that will be counted as a peak, and the sensitivity controls how far apart each peak must be to count as a separate peak.

In an optional step, once the peaks have been found for all of the spectra, binary spectral representations are created for all of the initial spectra. These binary spectra are essentially vectors of ones and zeros. Each zero represents the absence of a peak feature and each one represents the presence of a peak feature. A peak feature is simply a peak that occurs within a certain spectral range, usually a few wave numbers. The vectors for all of the spectra are the same length and corresponding elements of these vectors correspond to peak features occurring at nearly the same locations in the spectra.

In order to create these binary spectra, peaks are clustered into ranges of peak features. The process used to perform this peak clustering is a modified form of a 1-dimensional iterative k-means clustering algorithm. The process begins with the peaks picked from a single spectrum. These peak positions are used to define the centers of the spectral bins, peak feature bins, for the creation of the binary spectra. A spectral bins cover a range of wave numbers that may be specified by the operator (in one embodiment, the default is five wave numbers). The rest of the spectra are then iteratively added to the peak feature representation. At each step, any peak that fits into a pre-existing peak feature bin is added to that bin. For any peak that does not fit into a bin, a new bin is created. The centers of the bins are not permitted to move resulting in overlapping peak feature ranges. Then, the

centers of all of the ranges are re-calculated, optionally with a modified range of wave numbers, and the peak feature bins are re-defined relative to the new centers. This process can leave some peaks outside of an existing peak feature range. In this case, a new range is created for these peaks. This process creates a matrix with each row of the matrix corresponding to a binary spectrum specified in terms of the bins in which its peaks fall. An example of such a matrix for five spectra is given below in **TABLE 1**.

TABLE 1

Peak Position	270	350	390	430	510
Spectrum 1	1	1	0	1	1
Spectrum 2	1	0	0	1	1
Spectrum 3	1	1	0	0	0
Spectrum 4	0	1	1	1	0
Spectrum 5	1	1	0	1	1

In this matrix, Spectrum 1, for example, has a peak in each of the bins corresponding to wave numbers 270, 350, 430 and 510, but does not have a peak in the bin associated with wave number 390. This optional step of binary spectra generation has several benefits over "spectrum to spectrum" or peak height spectra comparisons, which include, but are not limited to, yielding data on differences between spectra that is useful in refining spectra collection and peak-finding algorithms, and reducing or removing orientation-dependent peak amplitudes from the spectra.

From either the spectra themselves, the peak height spectra, or from binary spectra such as those generated using the process described above, similarity between all of the spectra in the matrix can be calculated. This similarity measure is utilized to identify and create cluster boundaries. Illustrative, but not limiting, similarity measures include Hamming or Euclidean distance, or non-metric similarity indices such a the Tversky similarity index (or its derivatives such as the Tanimoto or Dice coefficients) or functions thereof.

In order to describe a spectra-to-spectra similarity matrix, the following notation can used: N_{mn} = number of peak values in a first spectrum falling in the same peak feature bin in a second spectrum; N_m = number of peak values in the first spectrum; and N_n = number of peak values in the second spectrum. Similarity can then be calculated using various methods, *e.g.*,

$$\text{Tanimoto (m,n)} = \frac{N_{mn}}{N_m + N_n - N_{mn}}$$

The matrices shown in **FIGS. 17A** and **17B** were generated using this foregoing method.

For similarity matrices based on binary spectra, the following notation can be used: a = number of 1's in a first spectrum that are zeros in a second spectrum; b = number of 1s in a second spectrum that are zeros in the first spectrum; and c = number of 1s in the first spectrum that are ones in the second spectrum. These values can be calculated in the following manners:

Hamming distance: $d = a + b$

Euclidean distance: $d = \sqrt{a+b}$

Tversky index: $t = \frac{c}{\alpha a + \beta b + c}$

Some of these metrics are related. For instance, the Tanimoto coefficient is equal to the Tversky index with α and β equal to 1. The Dice coefficient is equal to the Tversky index with α and β equal to 0.5. In a preferred embodiment, 1 - Tanimoto coefficient is used as the (dis)similarity measure. It should be noted that additional metrics, including metrics based on other metrics, may be used in alternative embodiments of the invention.

Once a particular way of measuring the similarity of sample spectra has been selected, the similarity of the spectra is determined. This determination typically results in a symmetric similarity matrix with each element (i,j) of the matrix representing the similarity between spectra i and j.

Using the similarity matrix or the binary spectra matrix, several different clustering methods can be employed to assign spectra into bins. Hierarchical clustering, k-means clustering, Gaussian mixture model clustering, and self-organizing map (SOM) based clustering are just some of the methods that can be used. These and other methods are well described in the literature. *See* Kohonen, T., "Self-organizing Maps", Springer Series in Information Sciences, Vol. 30, Springer, Berlin, Heidelberg, New York, 3rd Extended Edition (2001); Duda, R., Hart, P., and Stork, D., "Pattern Classification", John Wiley & Sons, 2nd Edition (November 2000); and Kaufman, L., Rowseeaww, "Finding Groups in Data", John Wiley & Sons, (1990). In a preferred embodiment, hierarchical clustering is used as a first-pass method of data analysis.

Using the information from the hierarchical clustering run, k-means clustering is then performed with user-defined cluster numbers and initial centroid positions. In another

embodiment, the number of clusters can be automatically selected in order to minimize some metric, such as the sum-of-squared error or the trace or determinant of the within cluster scatter matrix. *See* Duda, R., Hart, P., and Stork, D., "Pattern Classification", John Wiley & Sons, 2nd Edition (November 2000).

Hierarchical clustering produces a dendrogram-sorted list of spectra, so that similar spectra are very close to each other. This dendrogram-sorted list can be used to present the similarity matrix in a coded manner, wherein similarity indicia are used for each similarity region, including without limitation different symbols (such as cross-hatching), shades of color, or different colors. In a specific embodiment, the coded similarity matrix is presented in a color-coded manner, with regions of high similarity in hot colors and regions of low similarity in cool colors. Using such a visualization, many clusters become apparent as hot-colored square regions of similarity along the matrix diagonal. These square regions represent the high degree of similarity between all of the spectral (i,j) pairs in those regions. However, it should be noted that the failure of the coded similarity matrix to present a diagonal form is to be expected with some types of samples, although the matrix is still useful in representing more complex similarity relationships. Furthermore, in some cases there can be similarity regions along more than one possible diagonal that correspond to different rearrangements. Such rearrangements result in off-diagonal similarity square regions becoming part of the diagonal similarity square regions.

Along with the matrix representation of the cluster data, it is also useful to show where all of the spectra and the cluster boundaries lie in a dimensionally reduced space (usually 2-dimensions). There are several ways to perform this dimensionality reduction. In a preferred embodiment, a linear projection is made of a binary spectra matrix onto its first two principal components. Alternatively, the chosen similarity matrix could be used in order to create a map of the data using multidimensional scaling.

FIG. 18 illustrates an implementation of the binning procedure. At step **1800** the filter spectra are obtained. In one branch of the possible procedure, the peaks are located and corresponding binary spectra constructed in step **1805**. The binary spectra are used to create the similarity matrix during step **1810**. Next, hierarchical clustering results from sorting the similarity matrix to place similar spectra close to each other during step **1815**. This matrix is suitable for visualization during step **1830**. In one of the many alternative ways of processing the raw spectra of step **1800**, the peaks are located and instead of binary vectors of step **2405**, peak height vectors are generated during step **1820**. Control can flow to step **1810** for the construction of a similarity matrix or directly point based clustering may be performed during

step **1825** followed by visualization of the results in step **1830**. Other alternative embodiments include control flowing from step **1805** following generation of binary vectors to point based clustering in step **1825** and then onto visualization in step **1830**.

An example Raman binning application is written in Visual Basic (VB). This VB program allows a user to select a group of spectra and set processing parameters. Preprocessing is performed within the VB application and then the filtered spectra are sent to OMNIC for peak finding through the Macros/Pro DDE communication layer provided by OMNIC. Once peaks are found, binary spectrum and distance matrix generation is performed in the main VB application. Then, the distance matrix is sent to MATLAB through a socket communication layer. Using a program such as MATLAB, clusters are generated and visualizations are created. These visualizations are made available to the main VB application through a web server. The resulting visualization allows for the easy identification of groups of samples that all have similar physical structure.

5. EXAMPLES

Some specific, non-limiting examples of particular features of the invention are provided below.

5.1. EXAMPLE 1: RAMAN DATA ACQUISITION SYSTEM

An automated robotic mechanism has been constructed and integrated with a microscope to facilitate selecting the sample containers (*e.g.*, tubes) from the blocks and positioning the containers under the microscope objective for spectral acquisition. The spectral data collection system comprises a dispersive Raman microscope (Almega dispersive Raman by Thermo Nicolet, 5225 Verona Road, Madison, Wisconsin 53711, USA), which is a research grade dispersive Raman instrument, combining a confocal Raman microscope and a versatile macro sampling Raman spectrometer. The highly automatable and versatile system offers multiple laser options under software automation, for optimized sensitivity, spatial resolution, and confocal operation. The Almega dispersive Raman spectrometer is capable of housing up to two lasers. Selection of the lasers and control of laser power is accomplished through software. In addition, the appropriate Rayleigh rejection filters, apertures and gratings are automatically selected when the laser excitation wavelength is changed. The high-resolution setting provides better than 2 cm^{-1} resolution for all laser wavelengths. The spectral range of operation for CCD based detection is 400-1050nm, allowing collection of Raman spectra over the full range for laser wavelengths. The example system is equipped

with a 785nm laser, a 256k x 1024k CCD detector, and a NTSC video camera to monitor samples in the microscope with a spectral range, when using the 785nm laser, of 100 –3200 cm^{-1}.

FIG. 13A shows a schematic of an entire automated spectra collection in Raman system **150**. Material handling automation has been designed around the microscope to allow automated sample handling in and out of Raman system **150**. Operationally, block **60** containing samples, in tubes **50**, is placed into block nest **25**. Block nest **25** is attached to XY stage **27** (Parker) to allow individual addressing of tubes within the block. XY stage **27** is attached to linear actuator **16** (Parker) that moves in the X direction. When block **60** is placed in block nest **25**, a sensor (commercially available from Keyence) is activated, causing linear actuator **16** to move block **60** into light-tight enclosure **18** surrounding the Raman microscope, shown in **FIG. 13B**. Once inside enclosure **18**, bar code reader **22** (commercially available from Keyence) reads the bar code on the sample block in order to track the contents of block **60**. Referring to **FIG. 13C**, lift mechanism **24** presents individual tube **50** to tube gripper **26** by pushing tube **50** up through the access hole in the bottom of block **60**. In **FIG. 13D**, tube gripper **26** and tube **50** are raised vertically from block **60**. Tube gripper **26** is attached to linear actuator **32** and rotary actuator **30**. **FIG. 13E** shows rotary actuator **30** rotating 90 degrees counter-clockwise to position tube **50** in a horizontal direction. Tube gripper **26** is then ready to travel horizontally along linear actuator **32** to move tube **50** near tube holder **36** in input stage **40** of microscope **20** as shown in **FIG. 13F**. **FIG. 13G** shows tube gripper **26** lowering tube **50** to tube holder **36**. Tube **50** is then placed in tube holder **36**, as shown in **FIG. 13H**, and tube gripper **26** is retracted vertically, as shown in **FIG. 13I**. Tube rotator **46** then engages tube **50** in tube holder **36**, shown in **FIG. 13J**. Finally, in **FIG. 13K**, microscope input stage **40**, under computer control, then actuates tube holder **36** under objective **38** for Raman analysis.

FIGS. 14A-G shows a procedure for focusing the Raman spectrometer on a solid form inside tube **50**. The solid form is typically found attached to an inside surface of the tube. Therefore, the Raman spectrometer is preset to first look at that position and depth. This focusing also reduces the noise due to out of focus fluorescent emissions. Although confocal techniques are not used in this example implementation, in alternative embodiments of the invention they provide greater reduction in the noise since only the radiation through a pinhole is used at any time with integration over time to reconstruct the entire image. Naturally, data collection is over longer periods of time.

Returning to the described embodiment, it becomes necessary to properly position the tube beneath objective **38** of the microscope so that the solid form is at the right depth. As shown in **FIG. 14B**, this is accomplished by moving the entire microscope stage (not shown) supporting the tube holder (tube holder **36**, as shown in **FIG. 14K**) in the X, Y and Z directions, as indicated by arrows **152**, as well as rotating tube **50**, as shown by arrow **154**, to present the solid form at the depth expected by the spectrometer. In a preferred embodiment, tube **50** has the geometry as shown in **FIG. 2**, where the top half of tube **50** is cylindrical, but the bottom half of tube **50** is tapered. If the solid form is located in the tapered portion of the tube, then moving the tube in the XY plane to position it under the objective will also result in changing the Z-height of the sample with respect to the spectrometer, thus the need for controlling the Z-height of the stage as well.

FIG. 14C shows a detailed perspective view of solid **160** inside tube **50** in an out-of-focus position and indicates the available axes of motion **152** and **154**. **FIG. 14D** shows a solid **160** that is out of focus because it is attached to the inside bottom (with respect to objective **38** located above) wall of tube **50** and is also located closer to the end of tube **50** than where the objective is focused. **FIG. 14E** shows how the microscope (not shown) stage is moved in the horizontal direction to bring solid **160** closer to the focal position. **FIG. 14F** then shows how tube **50** is rotated **154** to bring solid **160** closer to the focal position. However, after this rotational movement, solid **160** is now closer to the focal point and needs to be lowered. **FIG. 14G** shows how tube **50** is moved in vertical direction **152** to complete the process of bringing solid **160** into focus just below the inner surface of tube **50**.

Using spectral signal intensity feedback from the Raman CCD, focal distance is "auto-focused" by computer controlling the XY position and the Z-height between the tube and the microscope objective. This auto-focus capability allows for the automated collection of Raman spectra once the tube is in place under the objective. Additionally, the NTSC video camera on the Raman allows for video capture and frame grabbing of the sample as it is being analyzed. This feature further allows for a spatial "history" to be created whereby the exact location of laser on the tube can be associated with a specific Raman spectrum. In order to implement the previously mentioned auto-focus capability, the tube holder has a computer controlled, motorized rotation axis. This controllable rotation allows the system, again under feedback control, to rotate the tube under the microscope objective in order to scan the entire inside surface of the tube.

When this feature is used, it is often not quite as important to pre-align the samples in the tube so that the sample is in the field of view as discussed above. Moreover, this feature

allows for rotation during collection of a Raman spectrum. This is important to minimize so-called orientation effects that are sometimes observed in Raman spectra from anisotropic crystalline samples. Orientation effects exist when a sample has two or more unequivocal crystallographic "faces" that can be targeted by the laser source. Depending on the analyzed face, different spectra are generated, although the sample is physically unchanged. These different spectra might cause one to draw the conclusion that two or more different samples were present.

Once the sample in the tube is analyzed, the tube gripper removes the tube from the tube holder and returns it to the original location in the tube block followed by the XY stage indexing to the next tube to be analyzed.

5.2. EXAMPLE 2: DATA COLLECTION AND BINNING

The effectiveness of binning was demonstrated using two test sets that included the Raman spectra of a polymorphic material and a material with two hydration states. First, the authenticity of the samples was validated. Next, Raman spectra for each sample under varying acquisition conditions were collected. The spectra were then filtered and binned using the previously described algorithms and method. Finally, the results were cross-checked by comparison of the known sample identification to the bin/cluster assignment. Each of these steps is outlined below.

Authentic polymorphic forms (polymorphs) and anhydrate/hydrate forms for a given material each exhibit a unique x-ray powder diffraction pattern and melting transition. Such criteria were deemed sufficient evidence to verify authenticity of each sample. Representatives of each of the forms of sample sets 1 and 2 were therefore characterized using x-ray diffraction (XRD) and differential scanning calorimetry (DSC), generating x-ray powder diffraction patterns and thermal transition data to determine sample uniqueness. Aliquots of samples from set 2 were further characterized using thermo-gravimetric analysis (TGA) to confirm the hydration state (*i.e.*, water content) of the samples.

5.2.1. Materials and Experimental Methods

Two test sets were used to demonstrate the binning procedure for Raman spectra. Set 1 had two polymorphic forms of Flufenamic acid (2-[[3-(Trifluoromethyl)phenyl]-amino]benzoic acid), and set 2 had the anhydrate and monohydrate of theophylline (3,7-Dihydro-1, 3-dimethyl-1- H- purine-2, 6-dione). Anhydrous theophylline was obtained from Fluka Biochemica (Lot & Filling Code 403967/1 13700). The monohydrate was prepared by

suspending 4.0g of anhydrous theophylline in 20ml of methyl alcohol. While stirring, 20ml of de-ionized water was added to the suspension and the as-diluted suspension was warmed to approximately 40°C to promote conversion to the hydrated form. The resulting suspension was continuously stirred and allowed to cool to 25°C under ambient conditions. An aliquot of the suspension was collected by filtration after 6 hours and allowed to air dry. The solid obtained was characterized as described below to verify its hydration state.

All x-ray powder diffraction patterns were obtained using the D/Max Rapid X-ray Diffractometer (Rigaku/MSC, The Woodlands, TX, U.S.A.), which uses as its control software RINT Rapid Control Software, Rigaku Rapid/XRD, version 1.0.0 (©1999 Rigaku Co.), equipped with a copper source (Cu/K 1.5406), manual x-y stage and 0.3mm collimator. Samples were loaded in to 0.3mm quartz capillary tubes supplied by Charles Supper Company by tapping the open end of the capillary into a bed of the powdered sample. The loaded capillary was mounted in a holder that was placed into the x-y stage. Diffractograms were acquired under ambient conditions at a power setting of 46kV at 40mA in transmission mode, while oscillating about the omega-axis from 0-5 degrees at 1 degree/s and spinning about the phi-axis at 2 degrees/s. Exposure times were 30 minutes unless otherwise specified. The diffractograms obtained were integrated over 2-theta from 2–60 degrees and chi (1 segment) from 0-40 degrees at a step size of 0.02 degrees using the cyllnt utility in the RINT Rapid display software version 1.18 provided by Rigaku with the instrument. No normalization or omega, chi or phi offsets were used for the integration.

5.2.2. Results and Discussion

The resultant X-ray powder patterns, plotted as intensity (arbitrary units) as a function of 2-theta (degrees), are shown in **FIGS. 19A** and **19B** for the flufenamic acid **176** and theophylline **178** samples, respectively. Comparison of the x-ray powder patterns within each set clearly shows unique reflections (*e.g.*, shifted peaks) in each pattern, indicating structural differences between the samples within each set and hence, validating the authenticity of the samples. Note, comparable x-ray powder patterns for the anhydrous and monohydrate forms of theophylline have been reported by Zhu *et al.* (*International Journal of Pharmaceutics*, 135:151-160 (1996))).

Further confirmation of the authenticity of the test sets was provided by DSC thermal analysis. An aliquot of each sample was weighed into an aluminum sample pan obtained from TA Instruments (pan number 90078.609, lid number 900779.901). Pans containing flufenamic acid samples were crimped closed, whereas pans containing theophylline samples

were fit pressed to avoid pressure build up due to potential water vaporization. Sample pans were loaded into the apparatus and thermograms were obtained by individually heating the samples at a rate of 10°C /min from 20°C to 350°C using an empty crimped aluminum pan as a reference.

The DSC thermograms for the flufenamic acid **178** and theophylline **180** sample sets are shown in **FIG. 20A** and **20B**, respectively, where heat flow (W/g) is plotted as a function of temperature (°C). The melt transition (peak temperature) of flufenamic acid samples was observed at 134.4°C for Form I showing it to be pure Form I, whereas Form III exhibited a melt at 126.1°C, followed by re-crystallization and another melt at 133.9°C indicating the conversion of Form III (melting point = 126.1°C) to Form I upon heating. The DSC thermogram for anhydrous theophylline shows a single sharp endotherm at 273.1 °C corresponding to the melting transition of the sample. The DSC curve for the hydrated sample exhibits two endotherms, the first occurring at a peak temperature of approximately 77.4°C where dehydration of the sample is expected. This is followed by an endotherm at 273.1°C, where the anhydrous form melts.

Thermo-gravimetric analysis (TGA) was performed on samples from set 2 to verify water content. An aliquot of each sample was transferred into a platinum sample holder obtained from TA Instruments (#952019.9061) and loaded in to the apparatus. Thermograms were obtained by individually heating the samples at 10°C /min from 25°C to 300°C under flowing dry nitrogen (balance purge 40ml/min; sample purge 60 ml/min).

The thermograms obtained for the anhydrous and hydrous forms of theophylline are shown in **FIG. 21**, where the weight change (%) is plotted **182** as a function of temperature (°C). As illustrated in **FIG. 21**, the hydrated sample undergoes a two-step weight loss. The first weight loss 184 of 9.2% begins at approximately 25°C and continues until approximately 70°C. This weight change is associated with loss of loosely bound water from the hydrate structure and corresponds to a water mole fraction of 0.50, indicating the sample is a monohydrate of theophylline. For comparison, the theoretical water content for the monohydrate of theophylline is 9.09%. The small deviation in the measured sample is attributed to surface absorbed water, typically ranging from 0.0-0.3%. At approximately 172°C, the second weight loss **186** indicative of decomposition of the compound is observed. Note, the anhydrous theophylline sample exhibits only one weight loss **187** corresponding to decomposition beginning at approximately 172°C.

5.2.3. Reference Raman Spectra

For reference, Raman spectra were collected for each of the samples in sets 1 and 2. An aliquot of the sample was transferred to a glass slide that was positioned in the sample chamber. The measurement was made using the Almega™ Dispersive Raman system fitted with a 785 nm laser source. The sample was manually brought into focus using the microscope portion of the apparatus with a 10x power objective, thus directing the laser onto the surface of the powdered sample atop a glass slide. The spectra were acquired using the parameters outlined in the following table:

TABLE 2. Raman spectral acquisition parameters

Parameter	Setting Used
Exposure time (s)	2.0
Number of exposures	10
Laser source wavelength (nm)	785
Laser power (%)	100
Aperture shape	pin hole
Aperture size (μm)	100
Spectral range	105 – 4252
Grating position	single
Temperature at acquisition (°C)	24.0

The unfiltered Raman spectra generated for each sample are shown in **FIGS. 22A** and **22B**, where the Raman intensity (arbitrary units) is plotted as a function of Raman shift (cm^{-1}). Note that the appearance or disappearance of peaks and/or shifts in peak position between the samples within a set was observed. For example, the spectra shown in **FIG. 22A** for flufenamic acid polymorphs of sample set 1 show a doublet **190** centered around 450 cm^{-1} for form I and a singlet **192** for form III at that position, as well as significant shifting of the three peaks **193** in the 1150-1250 cm^{-1} range. Such peak appearance/disappearance and/or shifts in peak position indicate a unique crystal packing configuration, thus differentiating the forms and showing that the Raman spectra can be used as a unique signature for a given form.

5.2.4. Filtering of Raman Spectra

Evaluation of the filtering and binning algorithms was carried out by acquiring at least 20 spectra for each of the samples from sets 1 and 2, filtering the spectra to remove background signals, and binning the spectra. To collect the Raman spectra, an aliquot of each sample was transferred onto a glass slide or into a glass vial. Measurements were made by directing the laser onto the surface of the powdered sample atop the glass slide (theophylline) or through the glass vial (flufenamic acid). The acquisition parameters used are provided in **TABLE 2** with the exception that half of the spectra collected for each polymorph of set 1 (flufenamic acid) were collected using the 50x microscope objective rather than the 10x objective. The sampling location was varied either by moving the glass slide or rotating the glass vial.

All spectra were filtered to remove background signals, including glass contributions and sample fluorescence. This is particularly important as large background signal or fluorescence limit the ability to accurately pick and assign peak positions in the subsequent steps of the binning process. Such background contributions to the Raman spectra are shown in **FIGS. 16A** and **16B** for representative glass and fluorescent samples, respectively. Spectra from all samples of test sets 1 and 2 were filtered using a matched filter of feature size 25. An example of the original and filtered spectra for a fluorescent sample is shown in **FIG. 16C**.

5.2.5. Binning of Raman Spectra

Filtered spectra were binned using the algorithm described above under the peak picking and binning parameters given in **TABLE 3** and screen shots showing the output from the binning software captured during the binning procedure are provided in **FIGS. 17A** and **17B** for the flufenamic acid and theophylline sample sets, respectively.

TABLE 3. Peak picking and binning parameters used.

Parameter	Setting Used for Binning
QC Parameters	
Peak Height Threshold	1000
Region for noise test (cm^{-1})	0-10000
RMS noise threshold	10000
Automatically eliminate failed spectra	yes

Region of Interest	
Include (cm^{-1})	120 – 1800
Peak Pick Parameters	
Peak Pick Sensitivity	99
Peak Pick Threshold	100
Peak Comparison Parameters	
Peak Window (cm^{-1})	3
Analysis Parameters	
Number of clusters	2

The sorted cluster diagrams **194** and **196** showing the output for each sample set are illustrated in **FIGS. 23A** and **23B** and the corresponding cluster assignments for each spectral file are provided in **TABLES 4** and **5**, respectively.

TABLE 4. Cluster assignments for each spectral file for flufenamic acid sample set.

File Name	Cluster Number	Original Number	Sorted Number
Filtered flufenamic I 10x	1	1	1
Filtered flufenamic I 10x 10.SPA	1	2	5
Filtered flufenamic I 10x 2.SPA	1	3	6
Filtered flufenamic I 10x 3.SPA	1	4	9
Filtered flufenamic I 10x 4.SPA	1	5	4
Filtered flufenamic I 10x 5.SPA	1	6	7
Filtered flufenamic I 10x 6.SPA	1	7	8
Filtered flufenamic I 10x 7.SPA	1	8	15
Filtered flufenamic I 10x 8.SPA	1	9	2
Filtered flufenamic I 10x 9.SPA	1	10	3
Filtered flufenamic I 50x 1.SPA	1	11	16
Filtered flufenamic I 50x 10.SPA	1	12	11
Filtered flufenamic I 50x 2.SPA	1	13	17
Filtered flufenamic I 50x 3.SPA	1	14	18
Filtered flufenamic I 50x 4.SPA	1	15	20

Filtered flufenamic I 50x 5.SPA	1	16	12
Filtered flufenamic I 50x 6.SPA	1	17	19
Filtered flufenamic I 50x 7.SPA	1	18	13
Filtered flufenamic I 50x 8.SPA	1	19	14
Filtered flufenamic I 50x 9.SPA	1	20	10
Filtered flufenamic III 10x 1.SPA	2	21	21
Filtered flufenamic III 10x 10.SPA	2	22	28
Filtered flufenamic III 10x 11.SPA	2	23	29
Filtered flufenamic III 10x 2.SPA	2	24	26
Filtered flufenamic III 10x 3.SPA	2	25	22
Filtered flufenamic III 10x 4.SPA	2	26	23
Filtered flufenamic III 10x 5.SPA	2	27	31
Filtered flufenamic III 10x 6.SPA	2	28	30
Filtered flufenamic III 10x 7.SPA	2	29	27
Filtered flufenamic III 10x 8.SPA	2	30	24
Filtered flufenamic III 10x 9.SPA	2	31	25
Filtered flufenamic III 50x 1.SPA	2	32	33
Filtered flufenamic III 50x 10.SPA	2	33	34
Filtered flufenamic III 50x 2.SPA	2	34	36
Filtered flufenamic III 50x 3.SPA	2	35	35
Filtered flufenamic III 50x 4.SPA	2	36	32
Filtered flufenamic III 50x 5.SPA	2	37	37
Filtered flufenamic III 50x 7.SPA	2	38	39
Filtered flufenamic III 50x 8'.SPA	2	39	38
Filtered flufenamic III 50x 9.SPA	2	40	40

TABLE 5. Cluster assignments for each spectral file for theophylline sample set.

File Name	Cluster Number	Original Number	Sorted Number
Filtered Theophylline Hydrate1.SPA	1	1	1
Filtered Theophylline Hydrate10.SPA	1	2	14
Filtered Theophylline Hydrate11.SPA	1	3	7
Filtered Theophylline Hydrate12.SPA	1	4	8
Filtered Theophylline Hydrate13.SPA	1	5	9
Filtered Theophylline Hydrate14.SPA	1	6	15
Filtered Theophylline Hydrate15.SPA	1	7	10
Filtered Theophylline Hydrate16.SPA	1	8	11
Filtered Theophylline Hydrate17.SPA	1	9	16
Filtered Theophylline Hydrate18.SPA	1	10	12
Filtered Theophylline Hydrate19.SPA	1	11	17
Filtered Theophylline Hydrate2.SPA	1	12	19
Filtered Theophylline Hydrate20.SPA	1	13	2
Filtered Theophylline Hydrate3.SPA	1	14	3
Filtered Theophylline Hydrate4.SPA	1	15	4
Filtered Theophylline Hydrate5.SPA	1	16	18
Filtered Theophylline Hydrate6.SPA	1	17	13
Filtered Theophylline Hydrate7.SPA	1	18	5
Filtered Theophylline Hydrate8.SPA	1	19	6
Filtered Theophylline Hydrate9.SPA	1	20	20
Filtered Theophylline1.SPA	2	21	21
Filtered Theophylline10.SPA	2	22	27
Filtered Theophylline11.SPA	2	23	33
Filtered Theophylline12.SPA	2	24	28
Filtered Theophylline13.SPA	2	25	34
Filtered Theophylline14.SPA	2	26	29
Filtered Theophylline15.SPA	2	27	30
Filtered Theophylline16.SPA	2	28	22
Filtered Theophylline17.SPA	2	29	31
Filtered Theophylline18.SPA	2	30	23

Filtered Theophylline19.SPA	2	31	40
Filtered Theophylline2.SPA	2	32	36
Filtered Theophylline20.SPA	2	33	24
Filtered Theophylline3.SPA	2	34	38
Filtered Theophylline4.SPA	2	35	25
Filtered Theophylline5.SPA	2	36	26
Filtered Theophylline6.SPA	2	37	39
Filtered Theophylline7.SPA	2	38	37
Filtered Theophylline8.SPA	2	39	32
Filtered Theophylline9.SPA	2	40	35

In each sample set, two distinct clusters are observed represented by sorted spectra numbers 1-20 and 21-40 that correspond to the file names and sample identifications provided in **TABLES 4** and **5**. In comparing the cluster assignments to the sample identification (by file number), 100% binning accuracy is observed for each test set. For example, all form I samples are binned in cluster 1 and all form III samples are binned together in cluster 2 for flufenamic acid test set 1.

While the invention has been described in connection with what is presently considered to be the practical and preferred embodiments, the invention is not limited to the disclosed embodiments. In particular, it will be clear to those skilled in the art that this invention may be embodied in other specific forms, structures, and arrangements, and with other elements, and components, without departing from the spirit or essential characteristics thereof. One skilled in the art will appreciate that the invention may be used with many modifications of structure, arrangement, and components and otherwise, used in the practice of the invention, which are particularly adapted to specific environments and operative requirements without departing from the principles of this invention. The presently disclosed embodiments are therefore to be considered in all respects as illustrative and not restrictive, the scope of the invention being indicated by the appended claims.

CLAIMS

What is claimed is:

1. A high throughput system for evaluating experiments, which comprises:
   a) a plurality of containers, each of which contains a compound-of-interest and optionally one or more additional compounds;
   b) a block having a top surface, a bottom surface, and a plurality of holes for receiving the containers, wherein each hole has a top opening on the top surface and a bottom opening on the bottom surface, wherein the top opening is of a dimension sufficient to accommodate a container and the dimension of the bottom opening is of a dimension that will not accommodate the container; and
   c) a lifting mechanism which comprises a pin that when inserted in the bottom opening of the hole is of a length sufficient to at least partially push the container from the hole.

2. The system of claim 1 wherein the block is thermally conductive.

3. The system of claim 1 wherein the block is metal.

4. The system of claim 3 wherein the metal is aluminum.

5. The system of claim 1 wherein the containers are optically clear, transparent, semi-transparent, or translucent.

6. The system of claim 5 wherein the containers are made of glass or plastic.

7. The system of claim 1 wherein the containers are tubes.

8. The system of claim 1 further comprising closures for each container.

9. The system of claim 8 wherein the closures are adapted to allow the addition or extraction of fluid components.

10. The system of claim 8 wherein the closures are adapted to be pierced by a device adapted to remove or add fluid from the container.

11. The system of claim 8 wherein the closures allow light to enter the container for optical inspection of the contents of the container.

12. The system of claim 1 which further comprises an automated means of sealing the containers.

13. The system of claim 1 wherein the plurality of containers comprises at least 24 containers.

14. The system of claim 1, wherein the experiment is a salt forming experiment.

15. The system of claim 1 further comprising a robotic system for moving the blocks through the system.

16. The system of claim 1 further comprising an environmental control enclosure.

17. The system of claim 1 wherein the compound-of-interest is an active pharmaceutical ingredient and the optional additional compound is a solvent or excipient.

18. The system of claim 1 wherein the experiment is a crystallization or precipitation experiment.

19. A method of obtaining a solid form of a compound-of-interest which comprises:

a) providing a system which comprises:

i) a block having a top surface, a bottom surface, and a plurality of holes for receiving the containers, wherein each hole has a top opening on the top surface and a bottom opening on the bottom surface, wherein the top opening is of a dimension sufficient to accommodate a container and the dimension of the bottom opening is of a dimension that will not accommodate the container;

ii) a plurality of containers held in the block; and

iii) a thermal processing system for heating and cooling multiple blocks simultaneously;

b) dispensing a controlled amount of a compound-of-interest and optionally one or more additional compounds in each of the containers;

c) optionally sealing the containers; and

d) placing the block in the thermal processing system for an amount of time.

20. The method of claim 19 wherein an anti-solvent is added to at least one of the containers.

21. The method of claim 19 wherein at least one of the containers is not sealed so that any volatile compounds or solvents it contains are allowed to evaporate.

22. The method of claim 19 wherein at least one of the containers contains a compound-of-interest and a solvent, and the compound-of-interest is exposed to heat, is contacted with an anti-solvent, and/or the solvent is allowed to evaporate.

23. The method of claim 19 wherein the solid form is a crystal.

24. A high-throughput system for processing samples and screening for solid forms of a compound-of-interest, which comprises:

a) removable containers, each of which contains a compound-of-interest and optionally one or more additional compounds;

b) a block made of a thermally conductive material having an array of holes, each hole having a top and a bottom, the top having an opening for receiving the removable container and the bottom having an access hole;

c) a thermal processing system for heating and cooling multiple blocks simultaneously; and

d) an imaging device for detecting solid forms.

25. The system of claim 24 wherein the imaging device comprises a CCD camera.

26. The system of claim 24 which further comprises a computer configured to recognize images of solids.

27. The system of claim 24 wherein the compound-of-interest is an active pharmaceutical ingredient and the optional additional compound is a solvent or excipient.

28. A method for the high-throughput processing and screening of samples, which comprises:

a) providing a system for processing a sample comprising:

i) removable containers; and

ii) a block having an array of holes, each hole having a top and a bottom,

the top having an opening for receiving a container and the bottom having an access hole;

b) placing the containers in the holes;

c) dispensing a controlled amount of a compound-of-interest and optionally one or more additional compounds in each container to provide an array of samples;

d) processing the array; and

e) screening the samples for the presence or absence of solid forms using an imaging device.

29. The method of claim 28 wherein the compound-of-interest is an active pharmaceutical ingredient and the optional additional compound is a solvent or excipient.

30. The method of claim 28 wherein the screening comprises:

a) imaging the contents of the containers; and

b) analyzing the images for a desired experimental result.

31. The method of claim 28 which further comprises prescreening containers prior to the processing of the array.

32. The method of claim 31 wherein the prescreening comprises:

a) imaging the contents of the containers; and

b) analyzing the images for a desired experimental result.

33. The method of claim 28 which further comprising re-arraying the containers which contain the desired experimental results into an output block.

34. The method of claim 28 wherein processing comprises raising or lowering the temperature of containers in the array.

35. An automated high throughput method for screening for solid forms of a compound-of-interest, wherein the compound-of-interest is a biologically active small organic molecule, which comprises:

a) preparing an array of samples, each of which comprises the compound-of-interest and optionally one or more additional compounds;

b) processing the array so as to generate solid forms;

c) prescreening the array for solid formation using a digital imaging camera;

d) identifying samples with solid formations for further analysis;

e) rearranging and reprocessing samples with solids, and optionally repeating steps (b) to (d).

36. A method of evaluating experiments which comprises:

a) providing a system for evaluating experiments comprising:

i) a plurality of containers, each of which contains a compound-of-interest and optionally one or more additional compounds;

ii) a block containing an array of holes for receiving the containers; and

iii) an imaging device;

b) positioning the block near the imaging device;

c) producing images of the contents of each of the containers;

d) analyzing the images for the presence of a desired experimental result; and

e) identifying containers with the desired experimental result.

37. The method of claim 36 wherein the contents of the containers are analyzed for birefringence.

38. The method of claim 36 wherein the contents of the containers are analyzed for the scattering of laser light.

39. The method of claim 36 wherein the contents of the containers are subjected to processing.

40. The method of claim 39, wherein the processing is heating, cooling, mixing, changing pH, changing pressure, changing the concentration of the compound-of-interest, or placing the compound-of-interest in contact with one or more additional compounds.

41. The method of claim 36 wherein the containers without the desired result are subjected to further processing.

42. The method of claim 36 wherein the identified containers having the desired experimental result are re-arrayed to an output block for further analysis.

43. The method of claim 36 wherein the compound-of-interest is an active pharmaceutical ingredient and the optional additional compound is a solvent, excipient, or additive.

44. The method of claim 36 wherein the experiment is a crystallization or precipitation experiment.

45. A method of screening an array of samples which comprises obtaining a Raman spectrum of each sample and determining which, if any, of the spectra share a spectral feature.

46. The method of claim 45 wherein the spectral feature is unique to a particular form of a compound-of-interest.

47. A method of screening an array of samples for the presence of a particular form of a compound-of-interest, which comprises obtaining a Raman spectrum of each sample.

48. A method of screening an array of samples for the absence of a particular form of a compound-of-interest, which comprises obtaining a Raman spectrum of each sample.

49. The method of claim 45, 47, or 48 wherein the particular form is a solid form.

50. The method of claim 49 wherein the solid form is a crystalline or amorphous form.

51. The method of claim 45, 47, or 48 wherein the particular form is a hydrated form.

52. A system for detecting similarities among a plurality of samples, which comprises:

a) a device for obtaining a spectrum for each sample; and

b) a computer configured to analyze each of the spectra and to generate a plurality of bins, wherein each bin corresponds to samples sharing at least one spectral feature.

53. The system of claim 52 wherein the device is an infrared spectrometer, near infrared spectrometer, NMR spectrometer, X-ray diffractometer, neutron diffractometer, light microscope, electron microscope, second harmonic generator, circular dichroism spectrometer, linear dichroism spectrometer, differential scanning calorimeter, thermal gravimetric analyzer, or melting point analyzer.

54. The system of claim 52 wherein the device is a Raman spectrometer.

55. The system of claim 52 wherein the computer is further configured to generate a binary spectral representation for a spectrum that reflects the presence or absence of a spectral feature.

56. The system of claim 52 wherein the computer is configured to mutually compare a plurality of spectra and generate a hierarchical clustering dendrogram.

57. The system of claim 52 wherein the computer is configured to cluster the plurality of spectra.

58. The system of claim 57 wherein the computer is configured to cluster the plurality of spectra in accordance with iterative k-means clustering.

59. The system of claim 52 wherein the computer is configured to cluster the plurality of spectra such that if a majority of spectra obtained from a single sample are assigned to a particular bin, then all spectra from that sample are assigned to that bin.

60. The system of claim 52 wherein the computer is configured to assign newly obtained spectra to at least one of the plurality of bins.

61. The system of claim 52 wherein the computer is configured to modify, in response to an analysis of newly obtained spectra, at least one of the plurality of bins.

62. The system of claim 52 wherein the computer is configured to add, in response to an analysis of newly obtained spectra, at least one bin to the plurality of bins.

63. The system of claim 52 wherein the computer is configured to generate a similarity matrix representing the similarity between at least two of the plurality of samples.

64. The system of claim 63 wherein the computer is further configured to sort the samples such that they are arranged to reflect their similarity.

65. The system of claim 63 wherein the computer is further configured to sort the similarity matrix such that a diagonal in the matrix represents samples exhibiting the greatest similarity.

66. A method of detecting similarities among a plurality of samples, which comprises:

a) collecting a spectrum for each of the plurality of samples;

b) calculating a similarity metric between the spectrum of one sample and that of at least one other of the plurality;

c) clustering, based on the similarity metric, the spectra into bins, each bin containing similar spectra; and

d) presenting the clustered spectra with similar spectra located close to each other.

67. The method of claim 66 wherein the spectra are preprocessed after they are collected.

68. The method of claim 66 wherein the positions of one or more spectral peaks in the preprocessed spectra are used to generate real value vectors.

69. The method of claim 68 wherein binary spectra are generated from the vectors.

70. The method of claim 66 wherein the spectra are Raman spectra.

71. A method of analyzing a plurality of samples, which comprises:

a) analyzing the samples with a spectrometer to produce spectral data;

b) under processor control, identifying similarities between the spectra; and

c) grouping the spectra into bins of similarity.

72. The method of claim 71 wherein the spectrometer is a Raman spectrometer.

73. A database containing a plurality of spectral samples organized into a plurality of bins, the bins corresponding to a hierarchical organization of the plurality of spectral samples based on pair-wise similarity scores calculated in accordance with a similarity metric.

74. The database of claim 73 wherein the similarity metric is a Tanimoto coefficient, Tversky index, Euclidean distance, or Hamming distance.

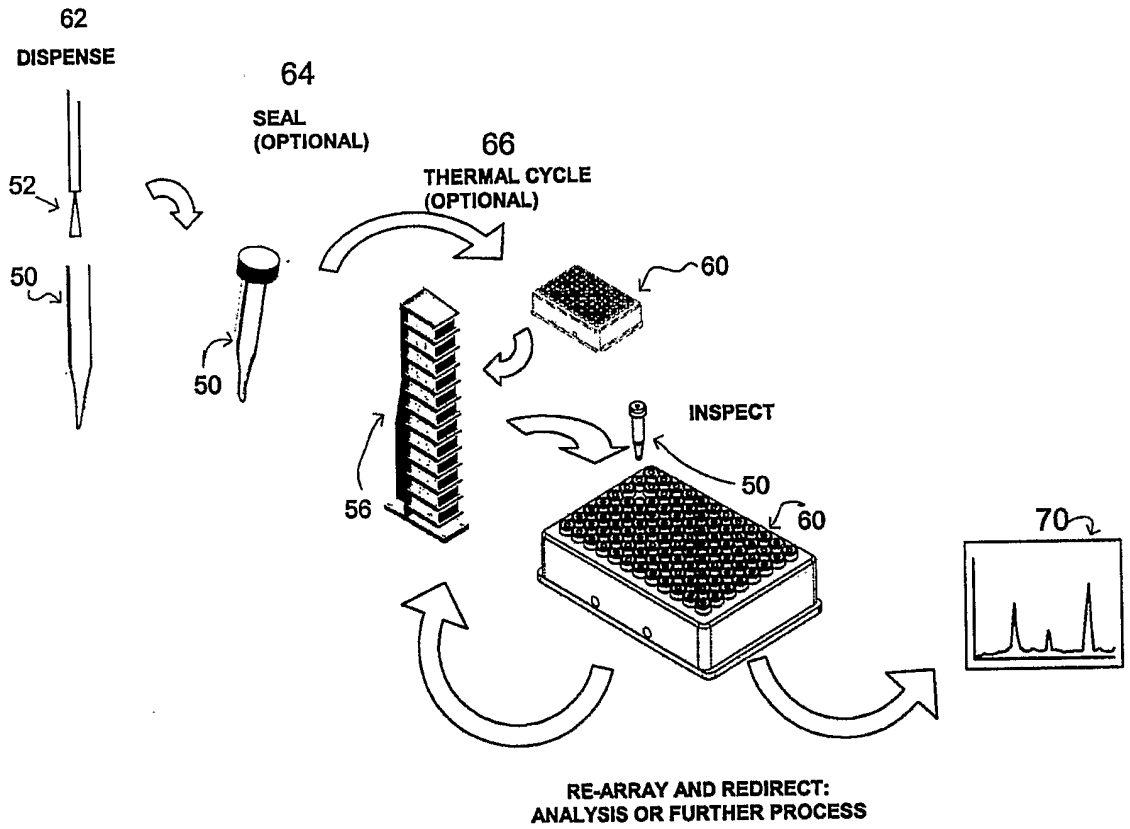

62
DISPENSE
64
SEAL
(OPTIONAL)
66
THERMAL CYCLE
(OPTIONAL)
52
50
50
60
INSPECT
50
56
60
70
RE-ARRAY AND REDIRECT:
ANALYSIS OR FURTHER PROCESS

Fig. 1

Fig. 2A

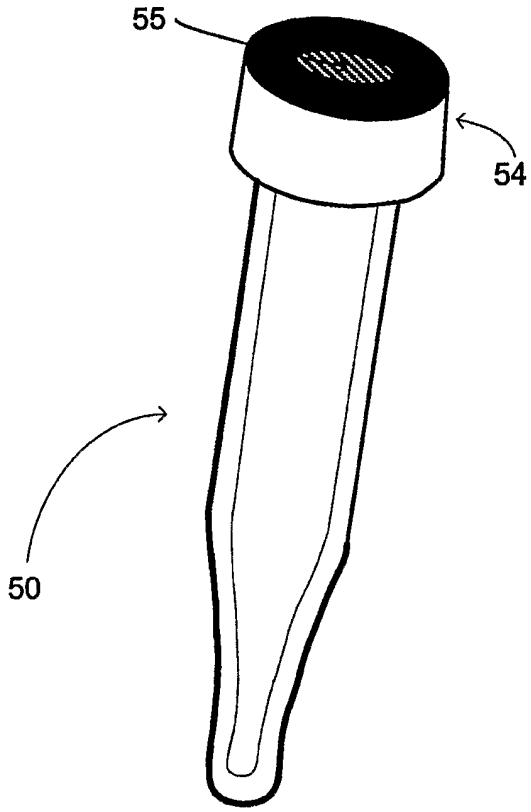


Fig. 2B

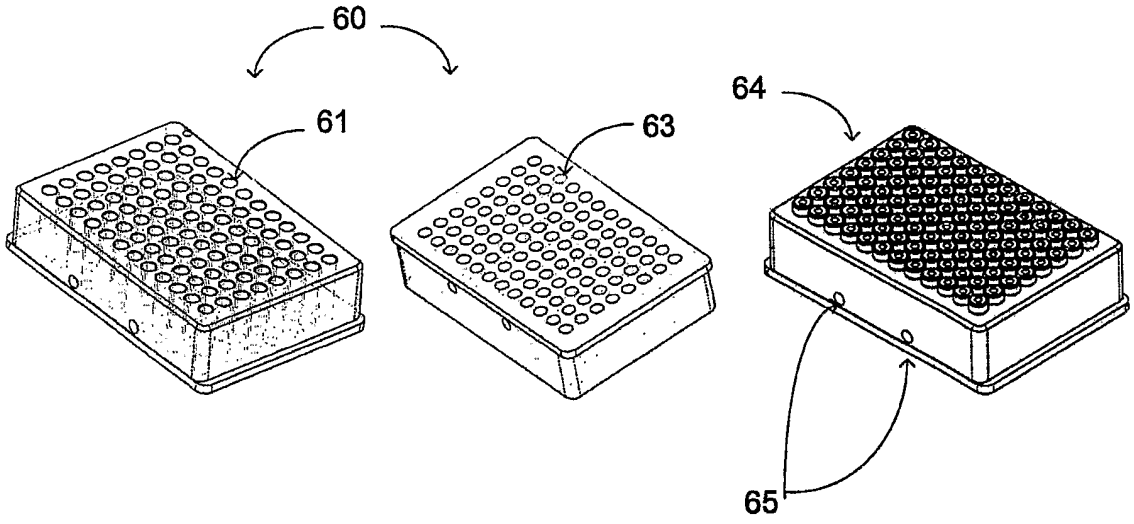


Fig. 3A Fig. 3B Fig. 3C

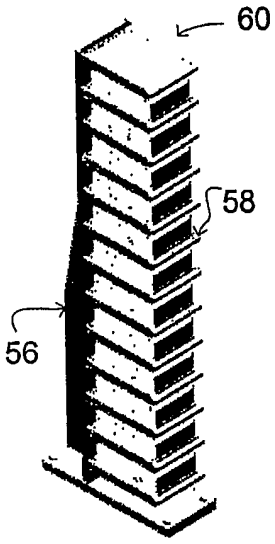


Fig. 4A

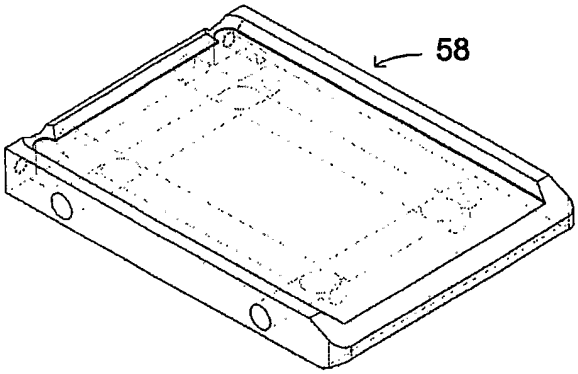


Fig. 4B

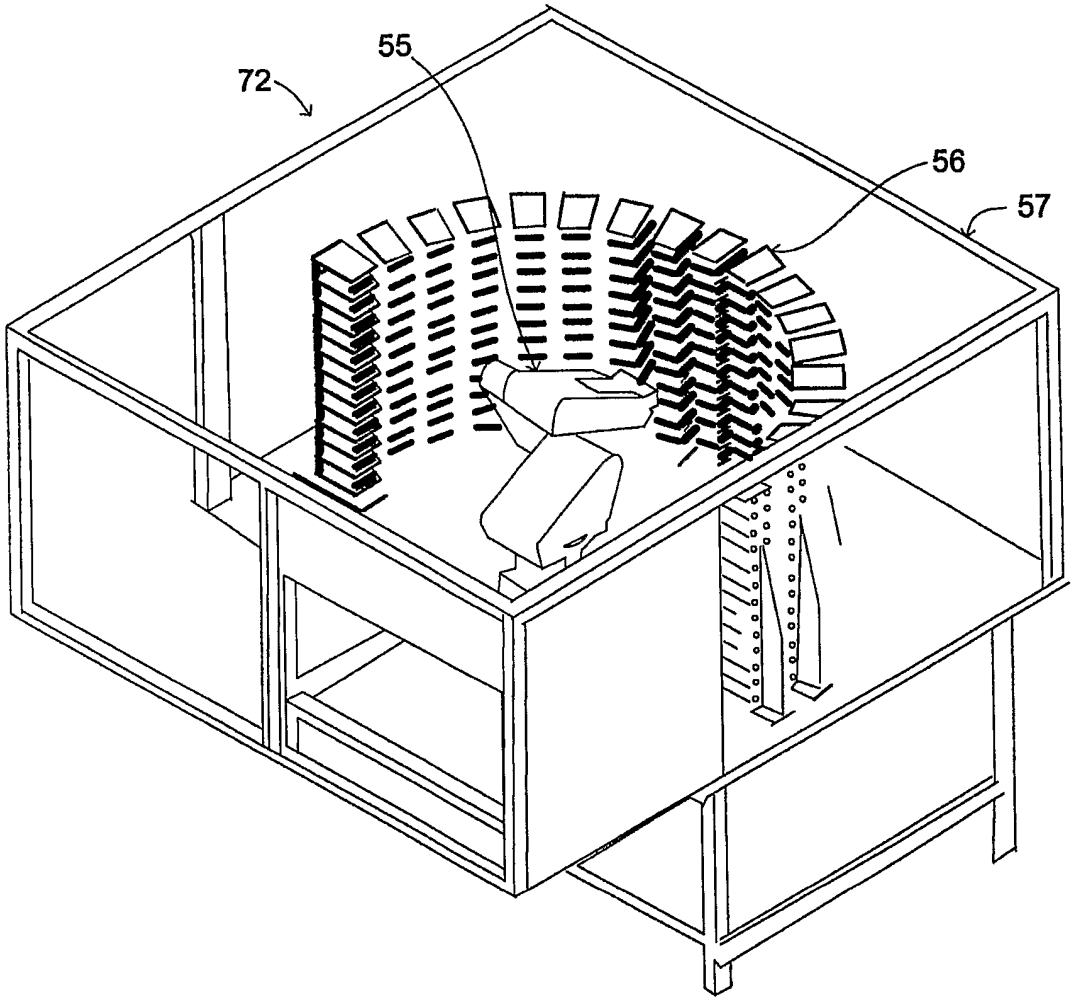


Fig. 5

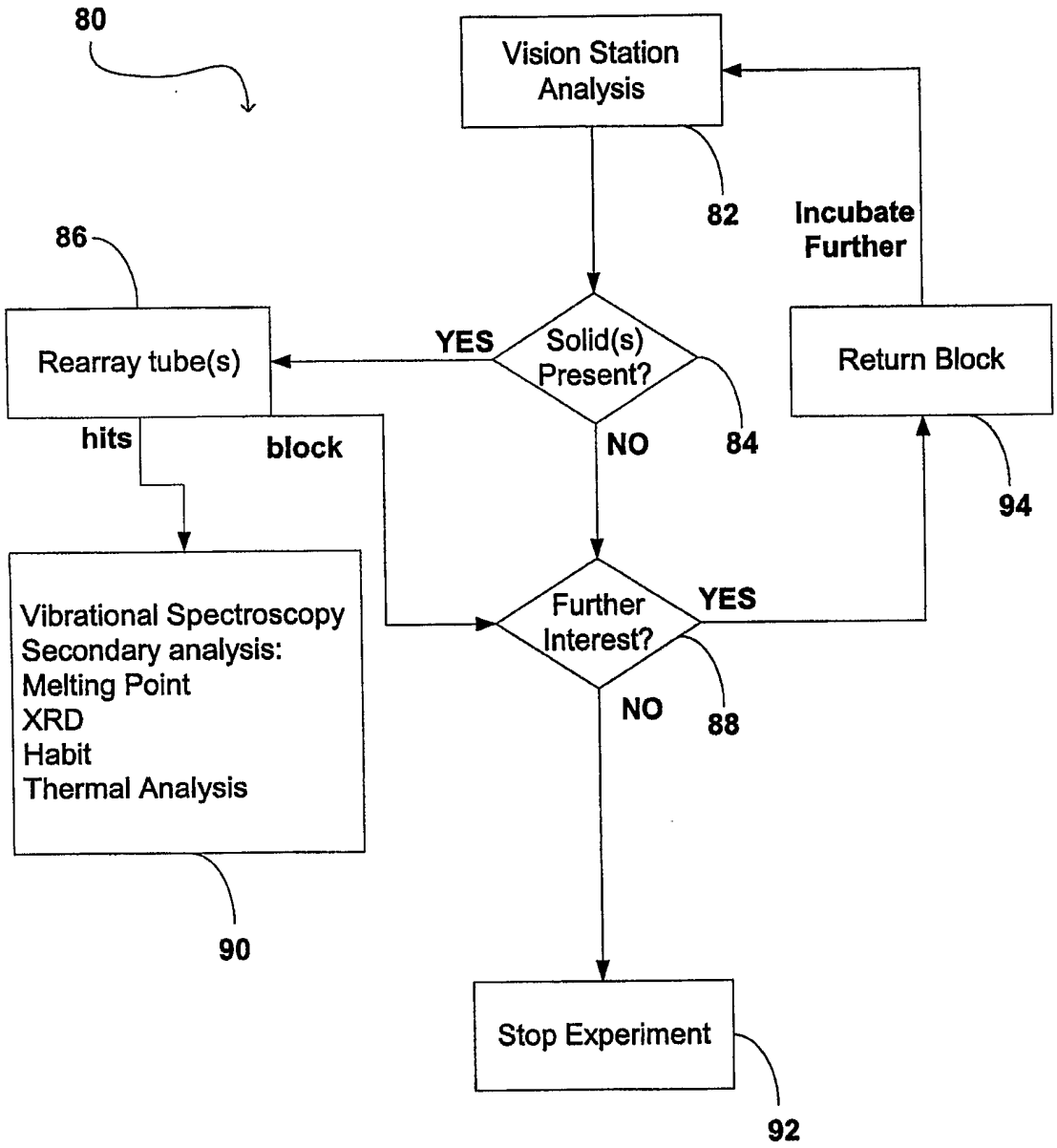

80
Vision Station Analysis
82
Incubate Further
86
Rearray tube(s)
YES
Solid(s) Present?
Return Block
hits
block
NO
84
94
Vibrational Spectroscopy
Secondary analysis:
Melting Point
XRD
Habit
Thermal Analysis
Further Interest?
YES
NO
88
90
Stop Experiment
92

Fig. 6

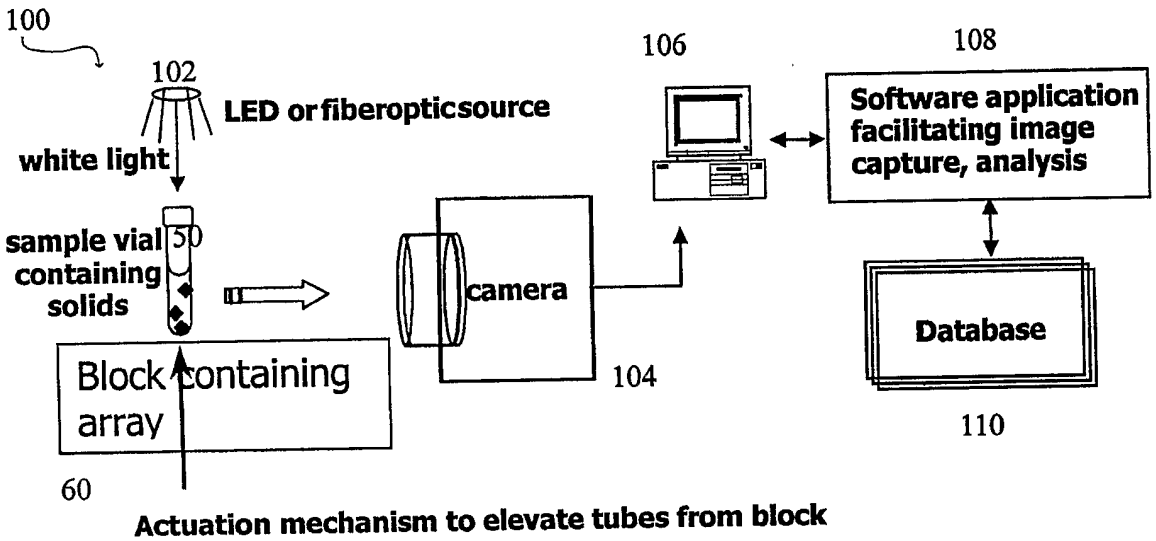

100
102
LED or fiberopticsource
white light
sample vial containing solids
50
Block containing array
60
Actuation mechanism to elevate tubes from block
camera
104
106
108
Software application facilitating image capture, analysis
Database
110

Fig. 7A

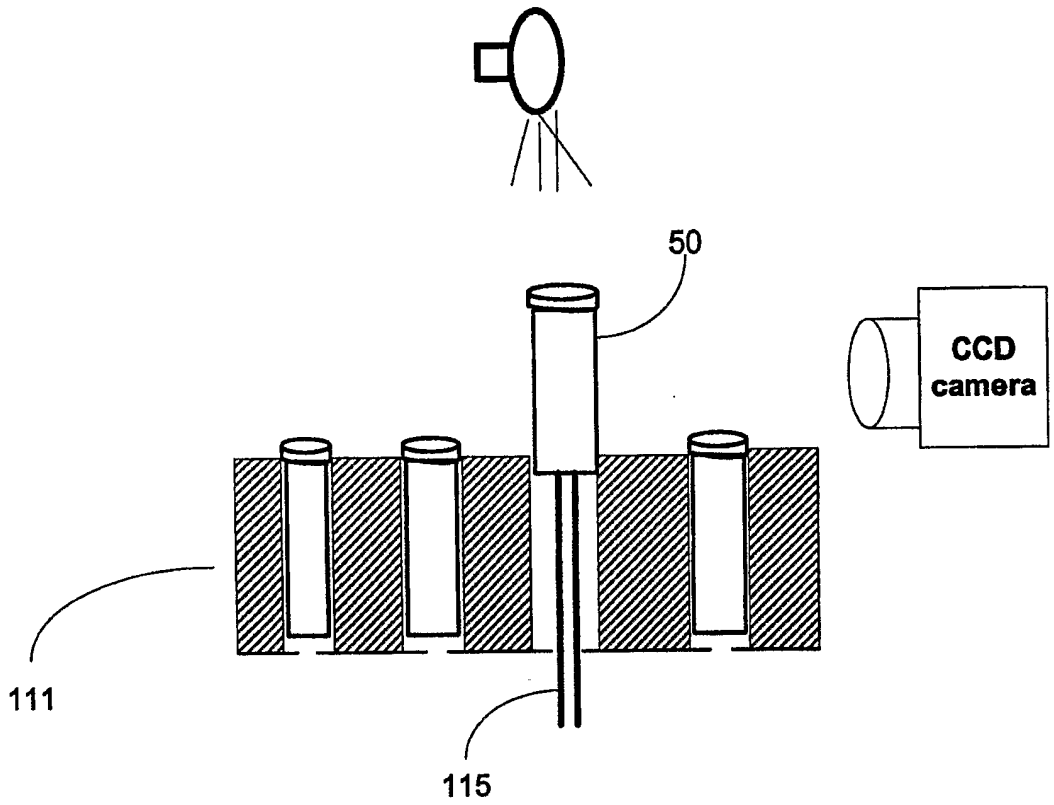


Fig. 7B

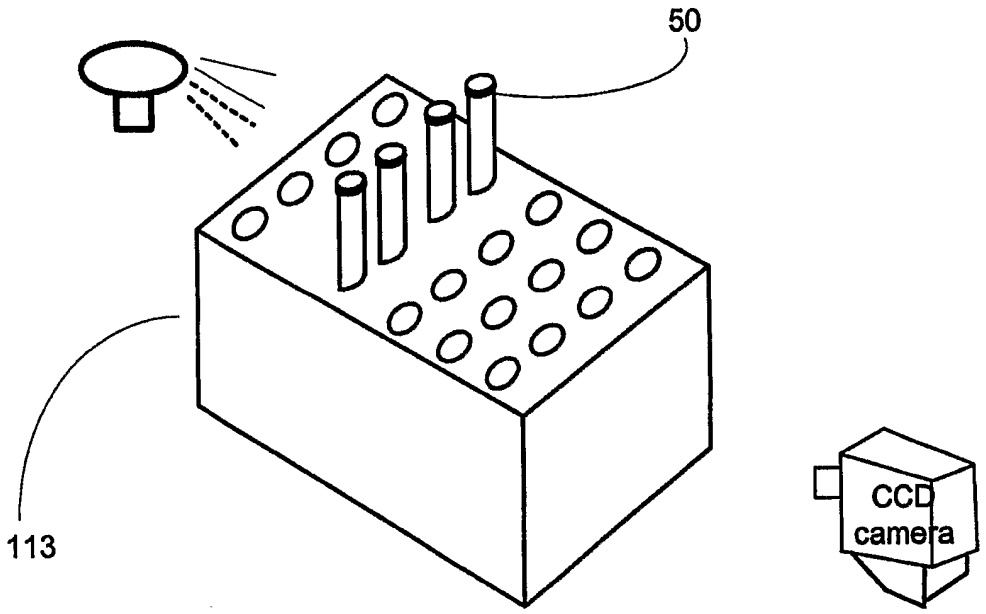


Fig. 7C

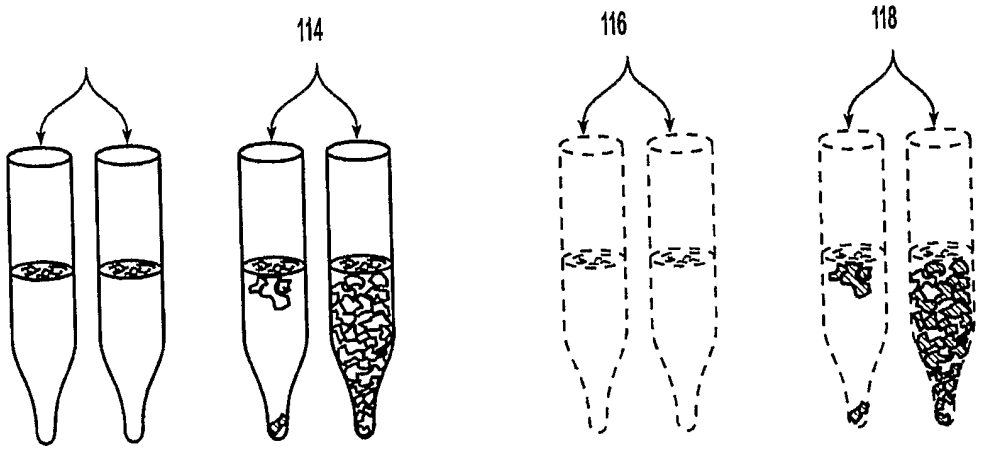

114
116
118

Fig. 8

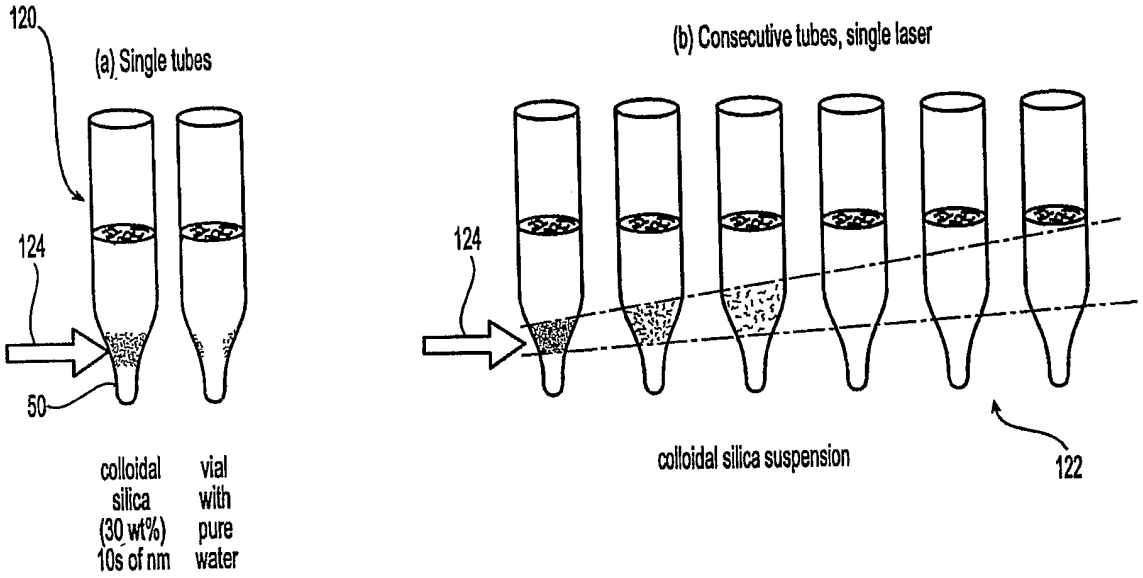

120
(a) Single tubes
124
50
colloidal silica (30 wt%) 10s of nm
vial with pure water
(b) Consecutive tubes, single laser
124
colloidal silica suspension
122

Fig. 9

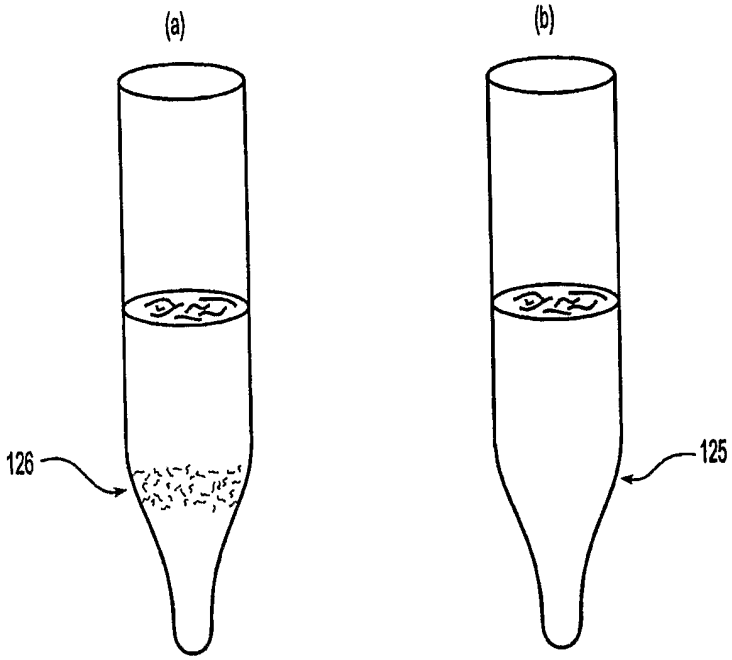


Fig. 10

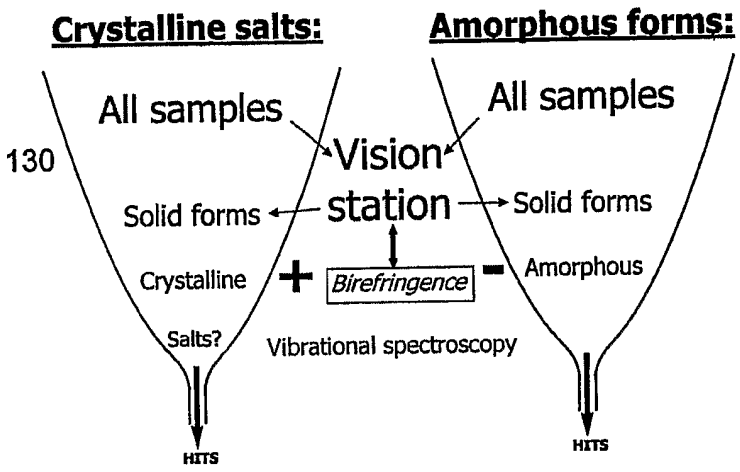

Crystalline salts:
Amorphous forms:
All samples
All samples
130
Vision
station
Solid forms
Solid forms
Crystalline
+
Birefringence
-
Amorphous
Salts?
Vibrational spectroscopy
HITS
HITS

Fig. 11

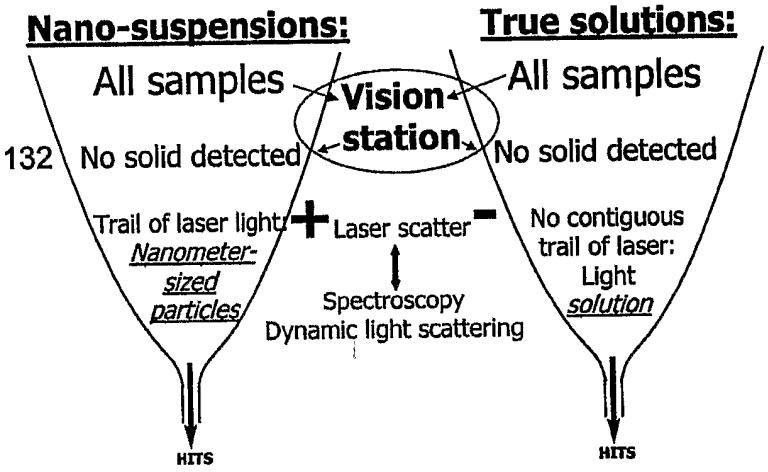

Nano-suspensions:
True solutions:
All samples
Vision
station
All samples
132
No solid detected
No solid detected
Trail of laser light:
Nanometer-sized particles
+
Laser scatter
-
No contiguous trail of laser: Light solution
Spectroscopy
Dynamic light scattering
HITS
HITS

Fig. 12

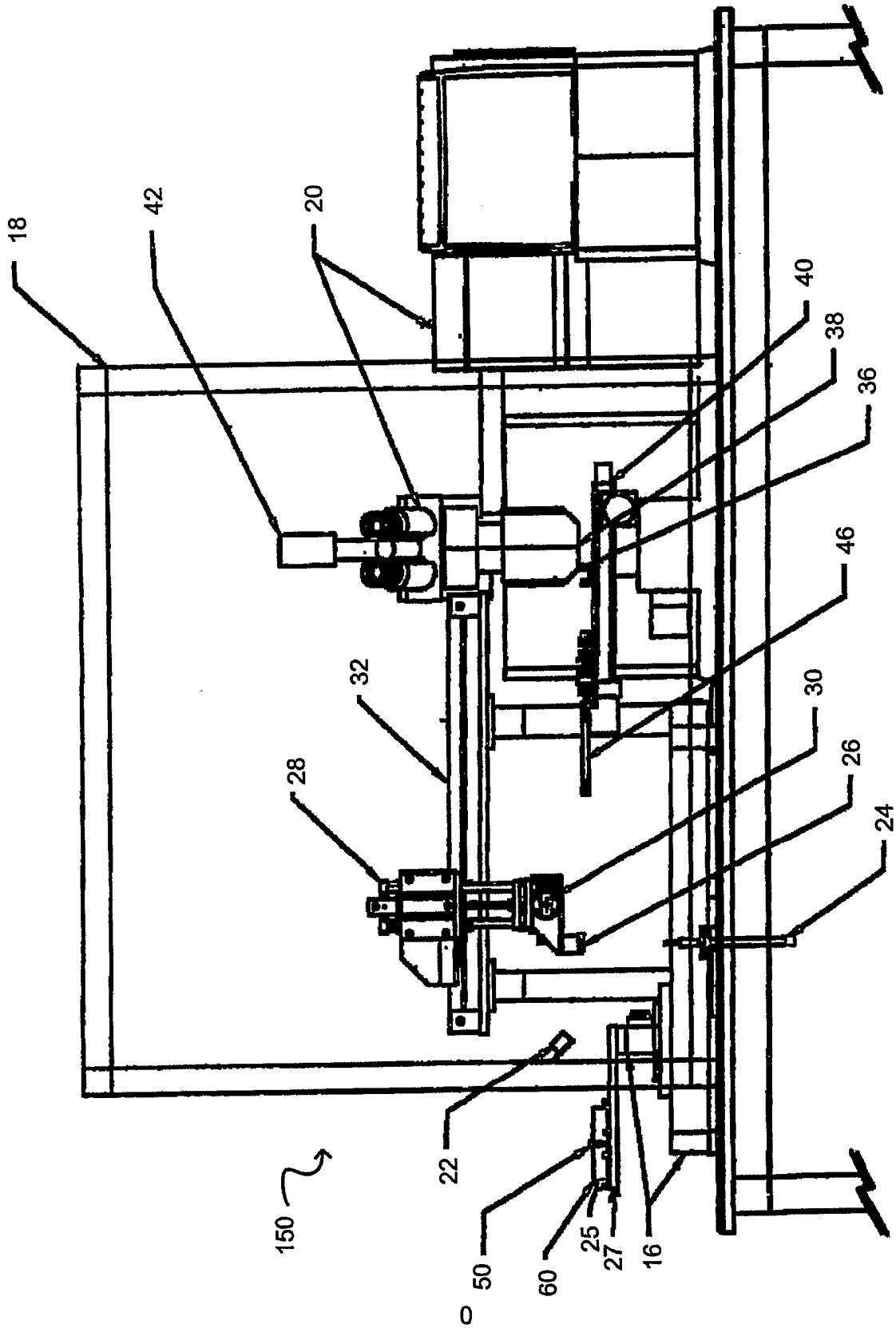


Fig. 13 A

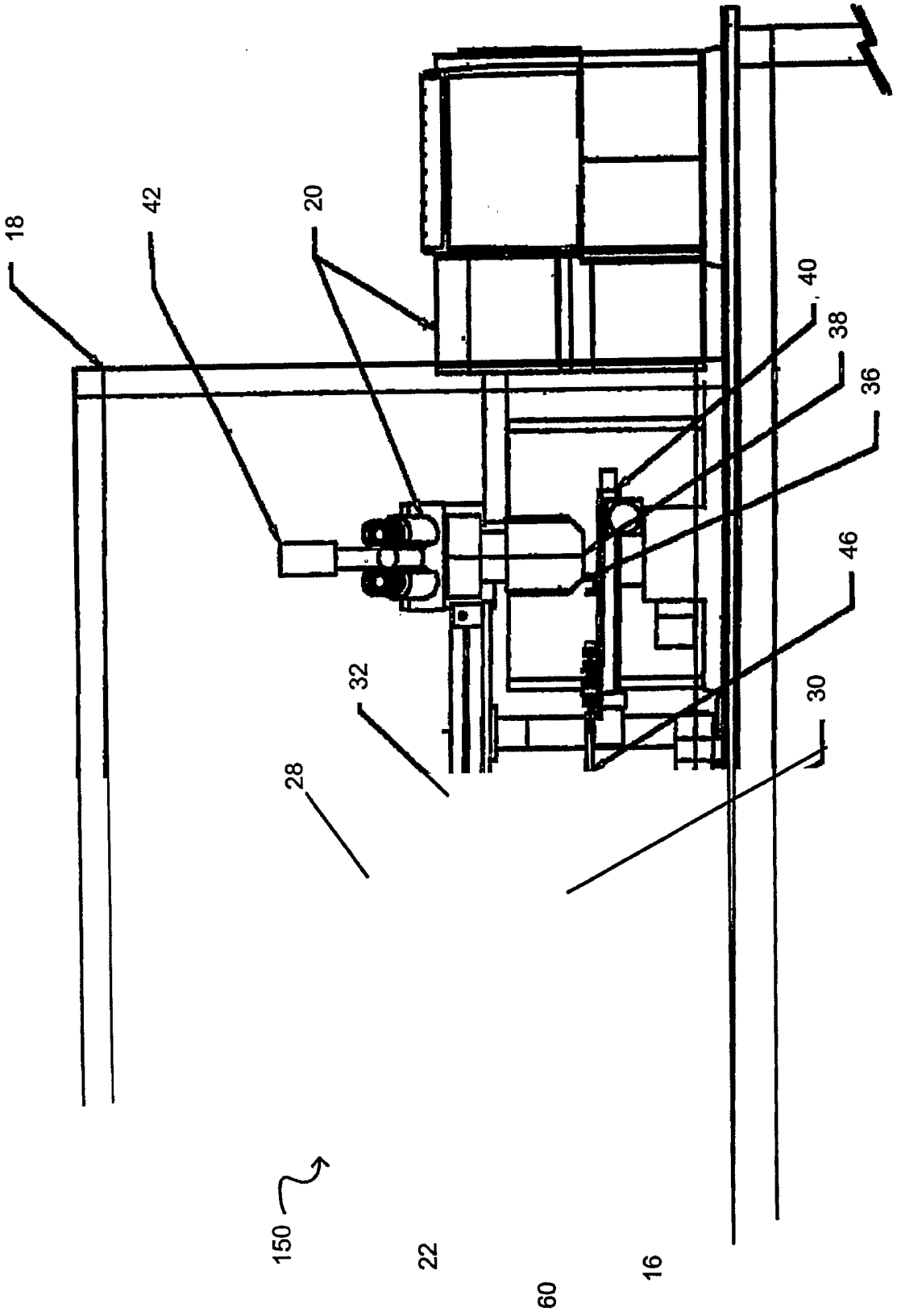


Fig. 13 B

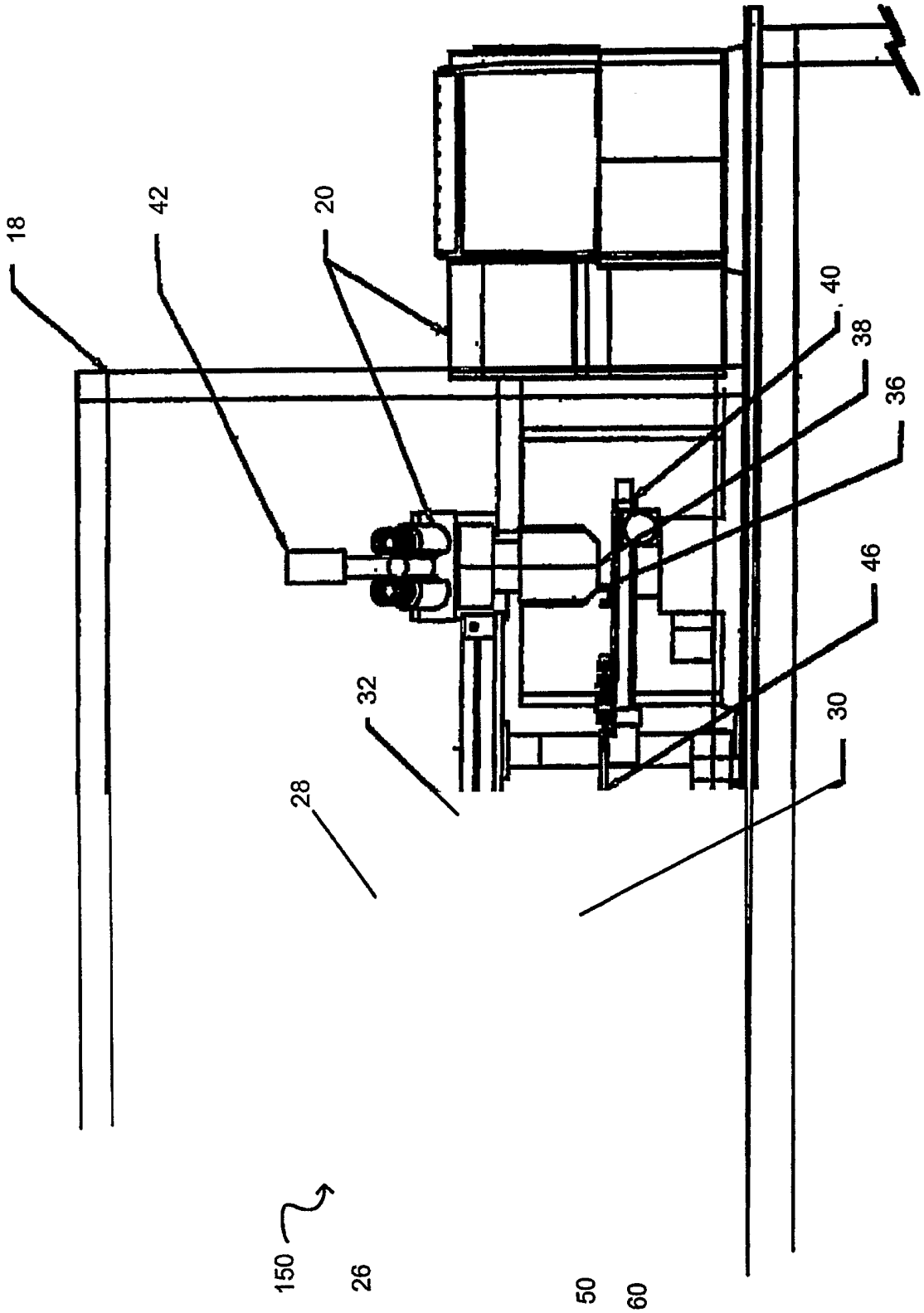


Fig. 13 C

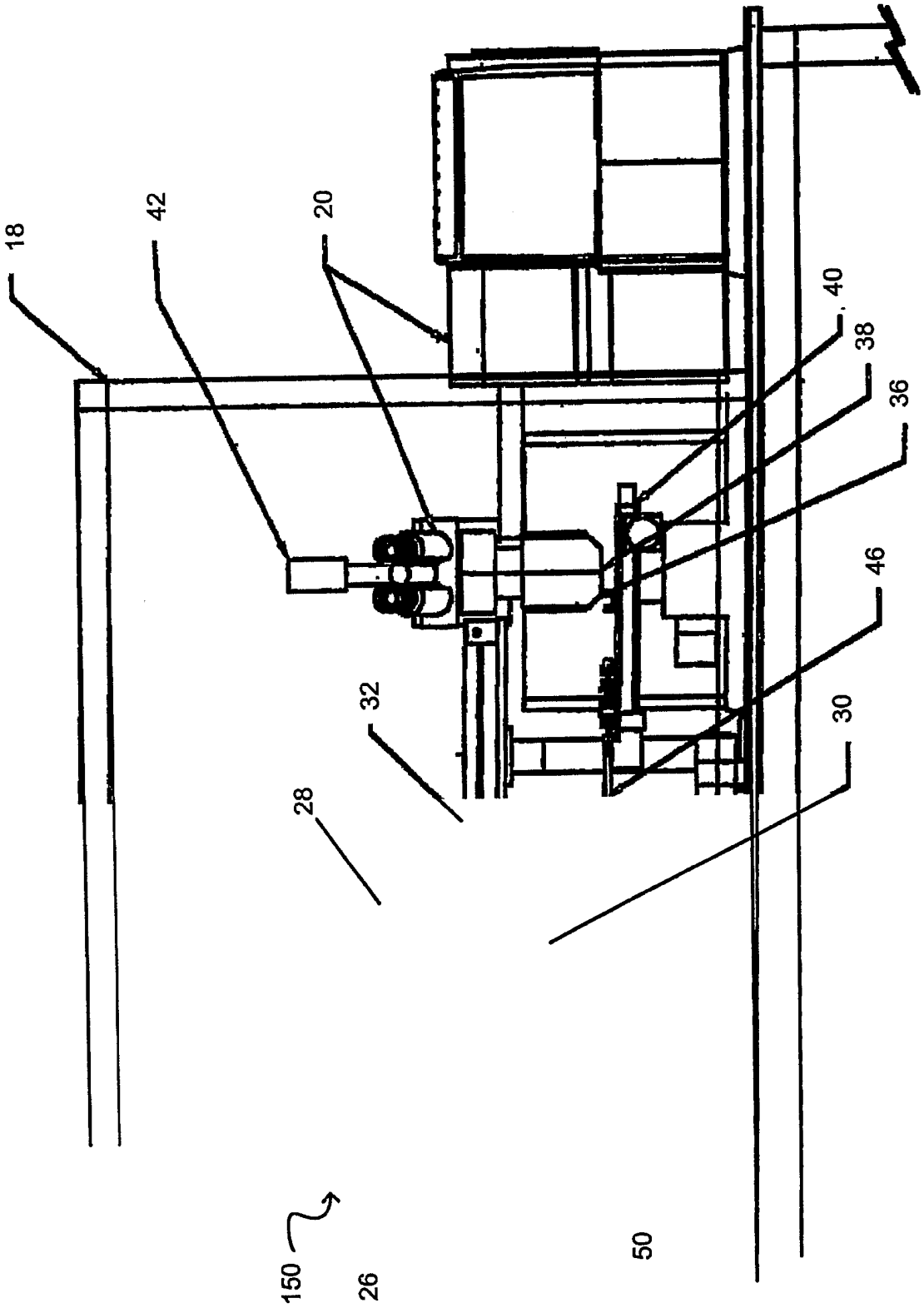


Fig. 13 D

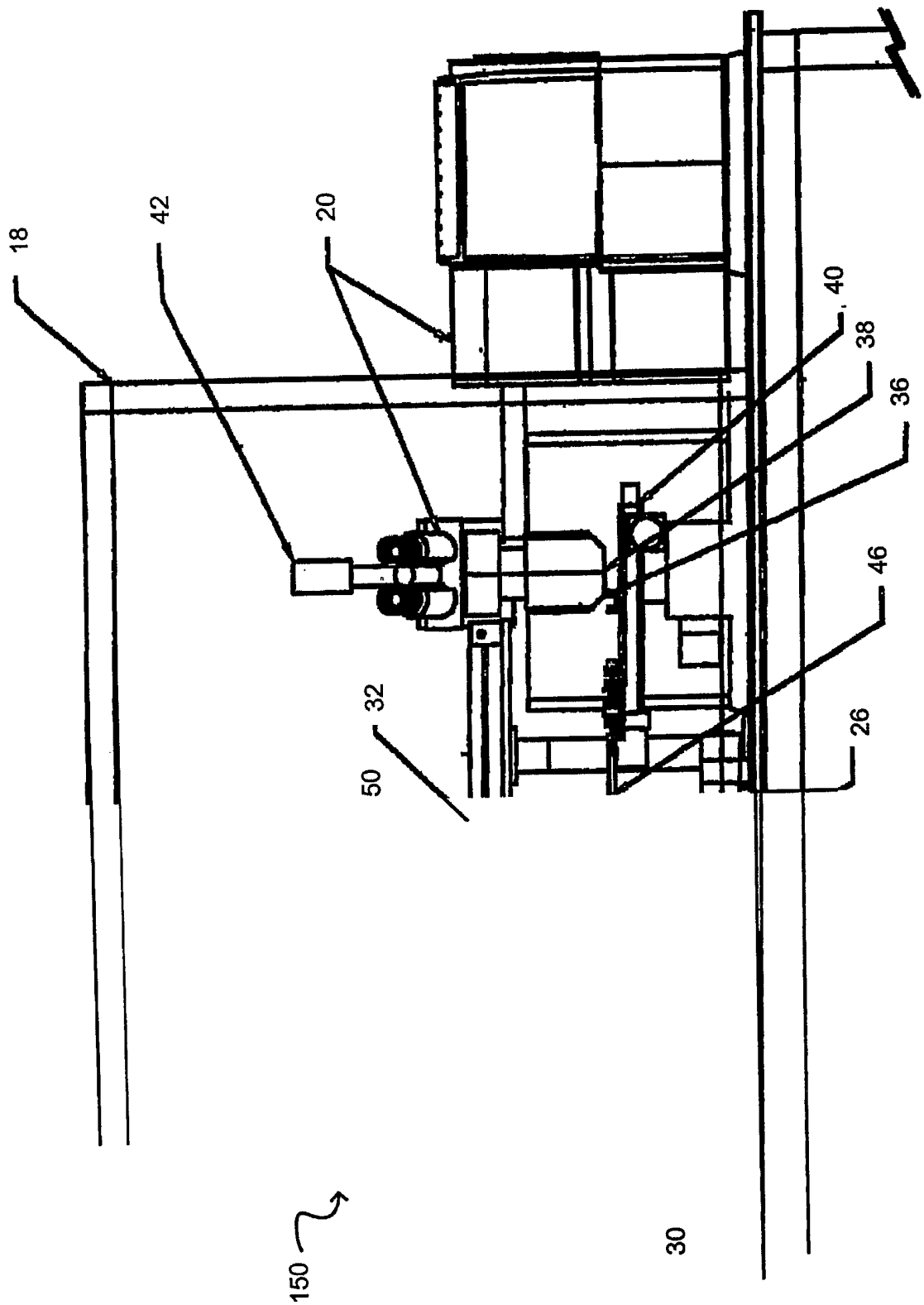


Fig. 13E

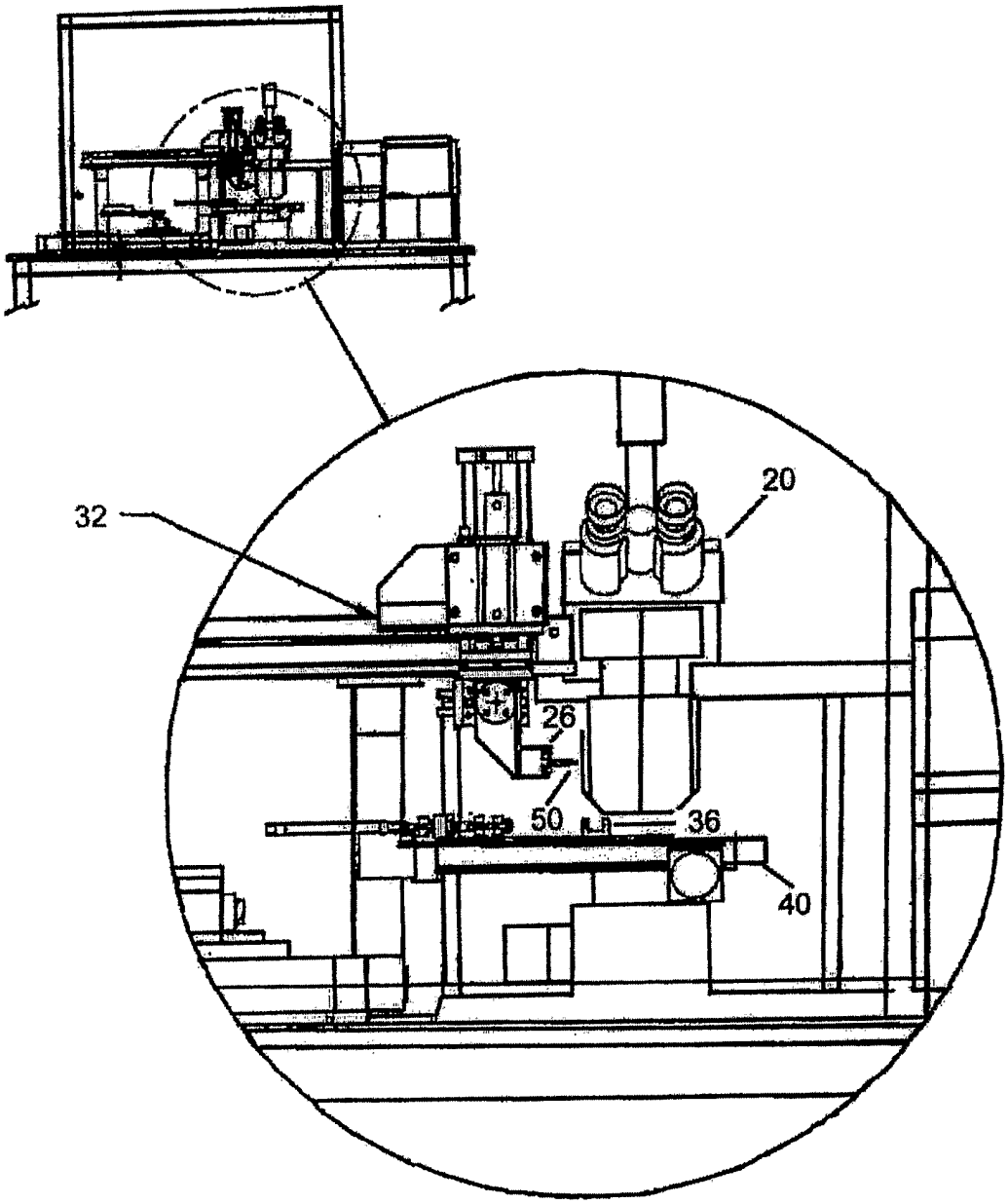


Fig. 13F

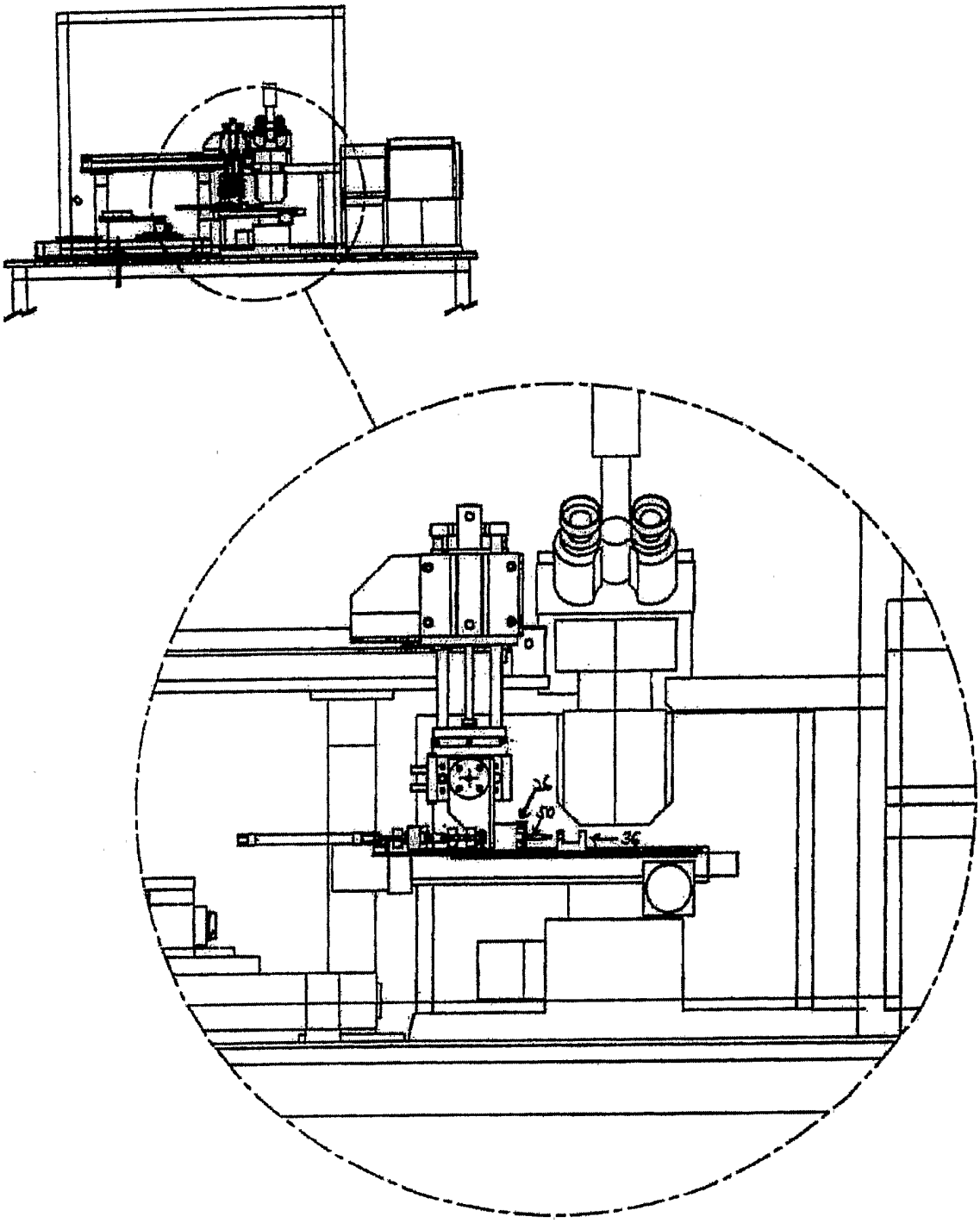


FIG. 13G

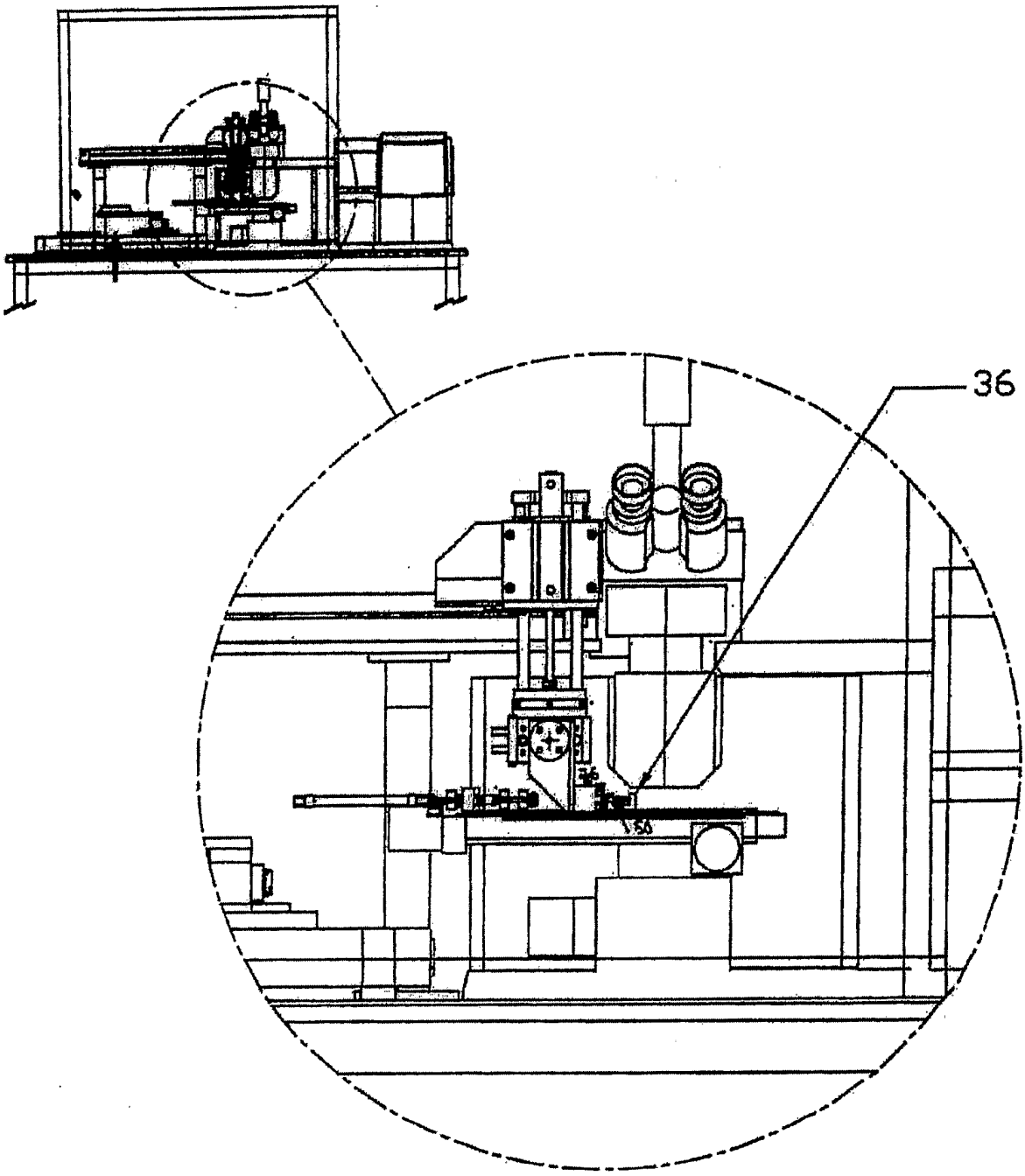


FIG. 13H

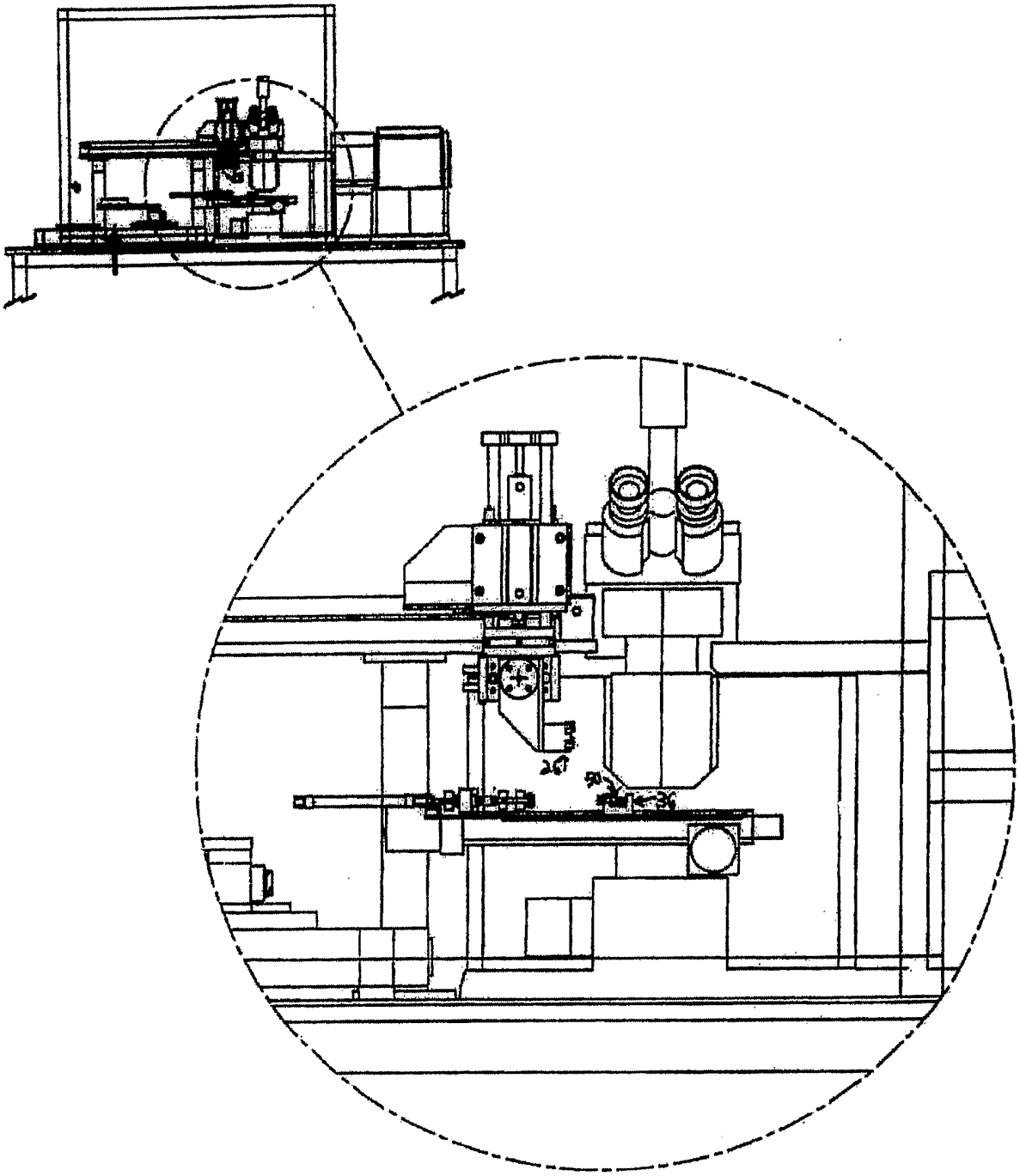


FIG. 13I

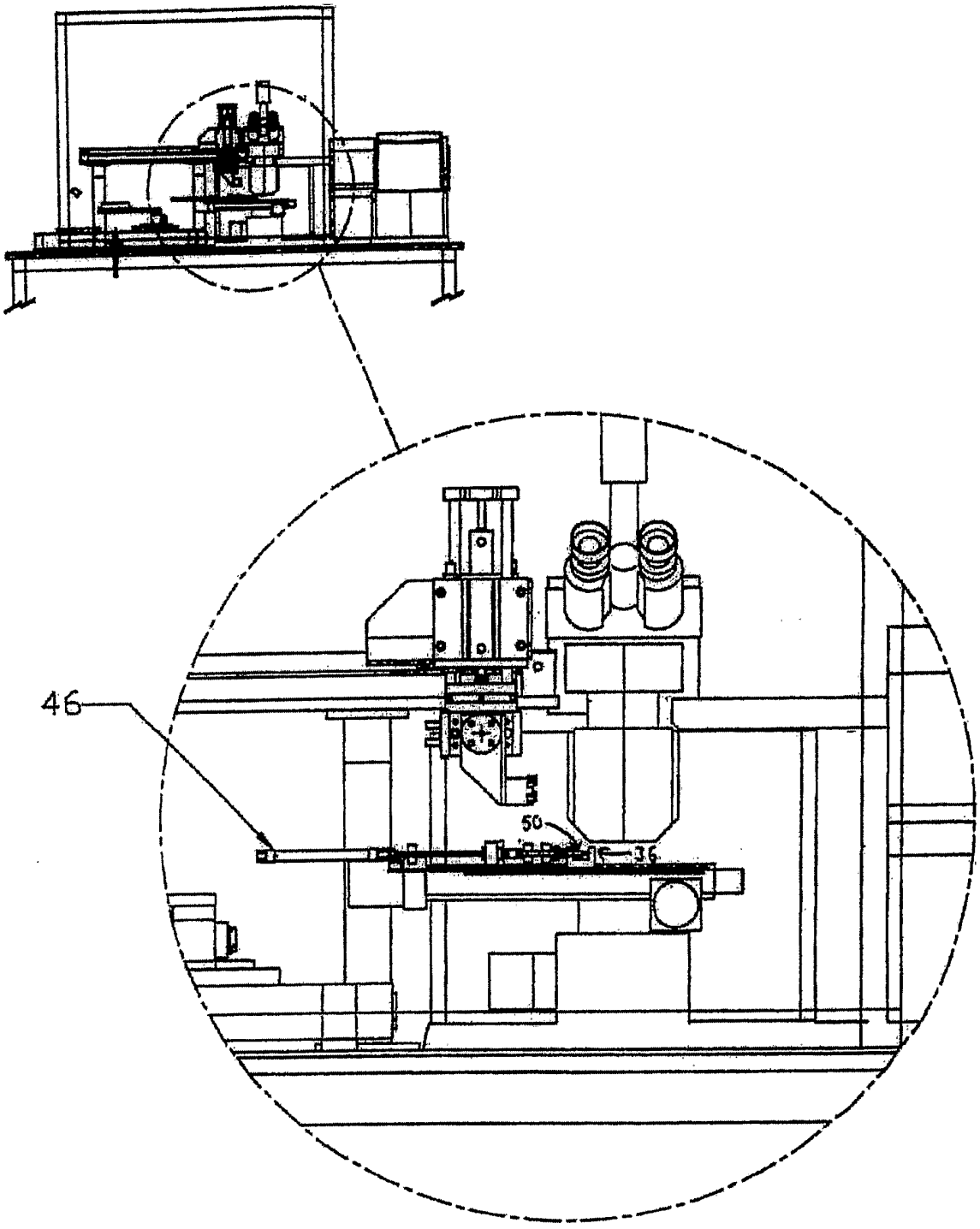


FIG. 13J

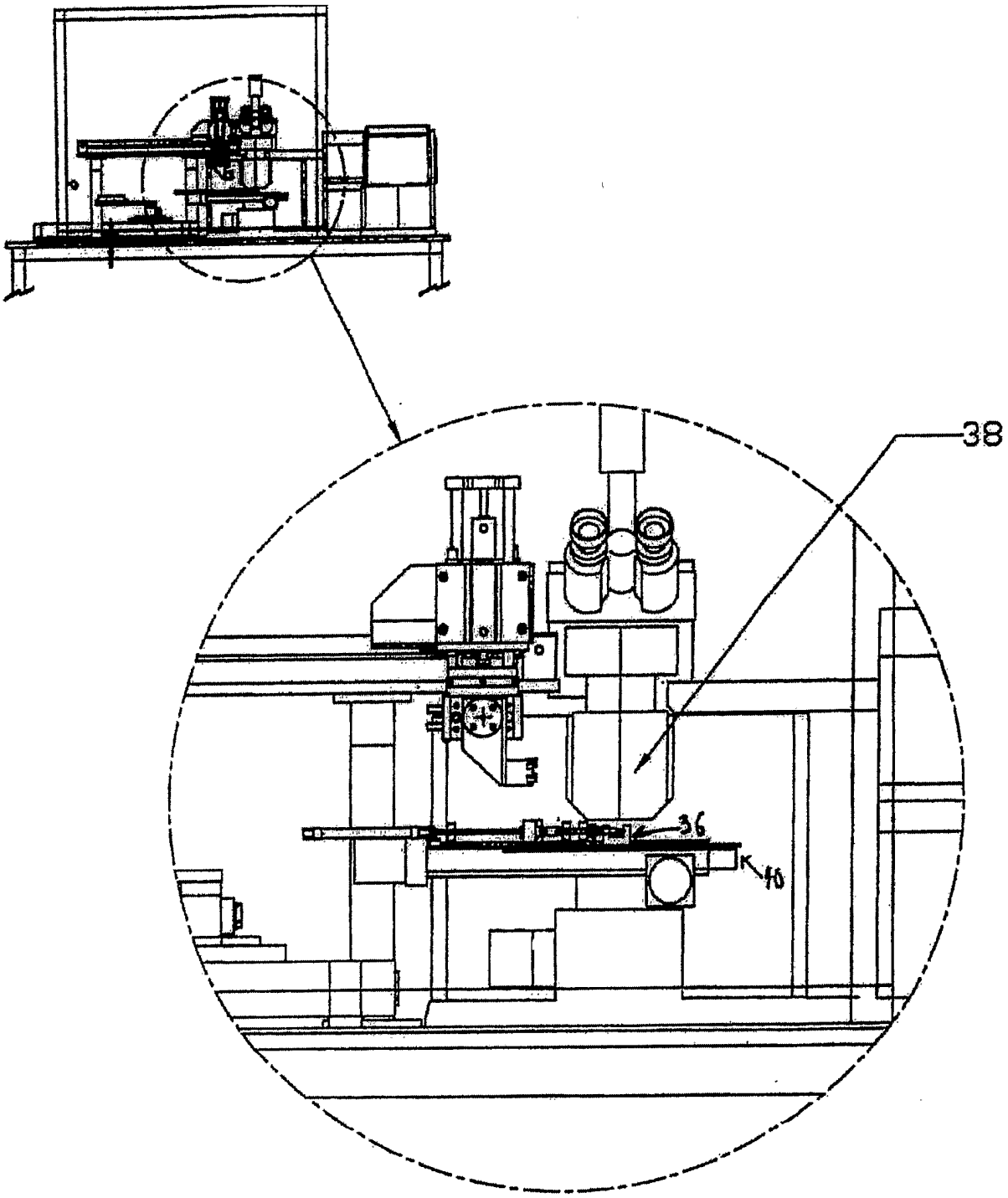


FIG. 13C

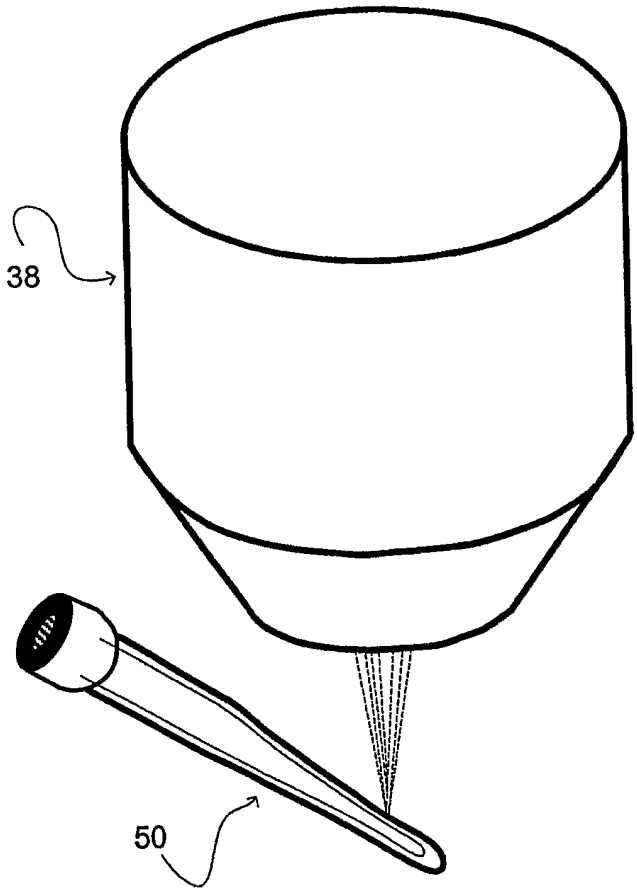


Fig. 14 A

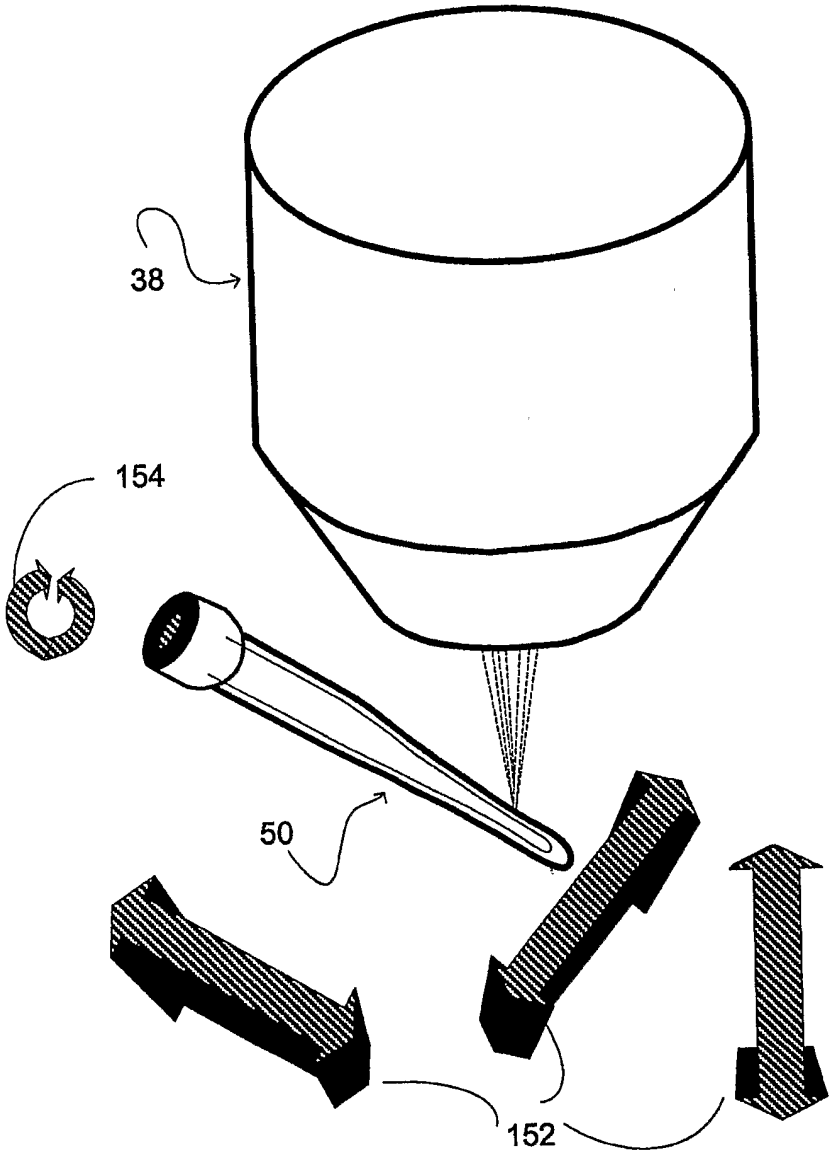


Fig. 14B

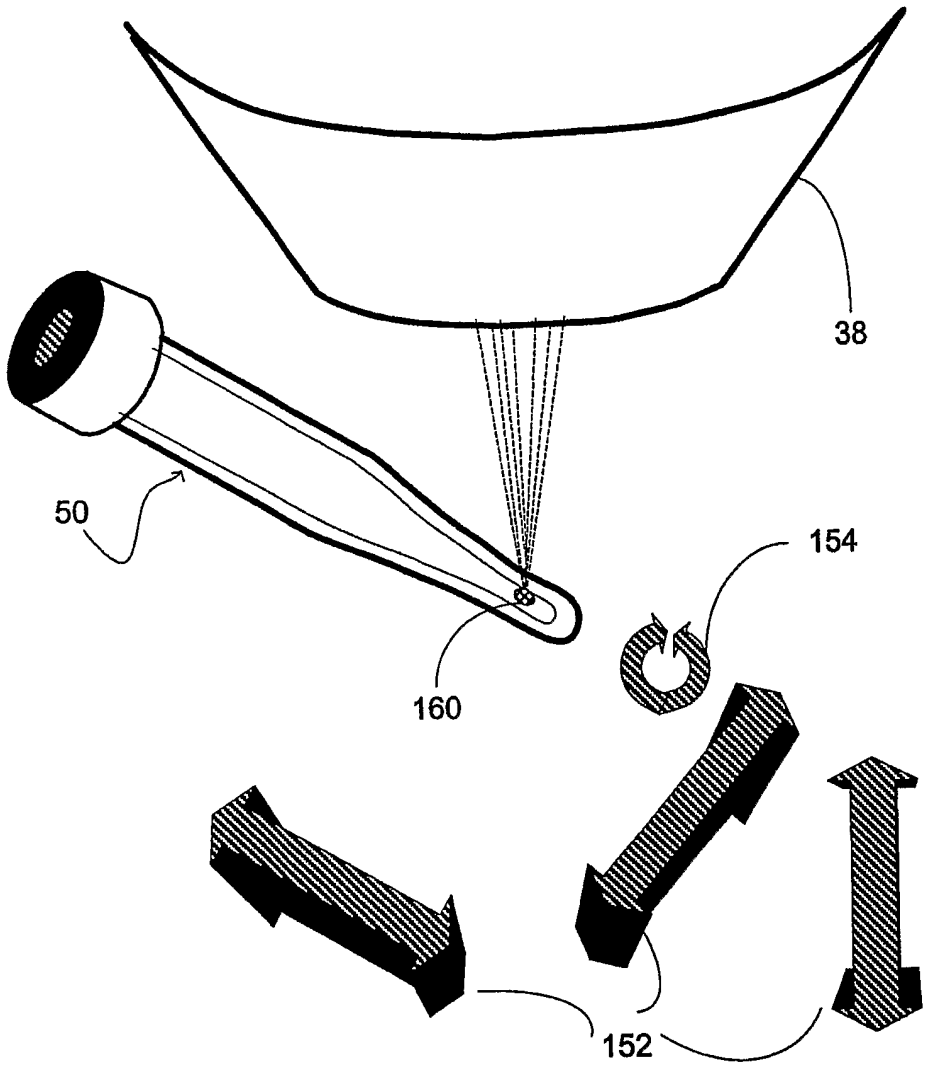


Fig. 14C

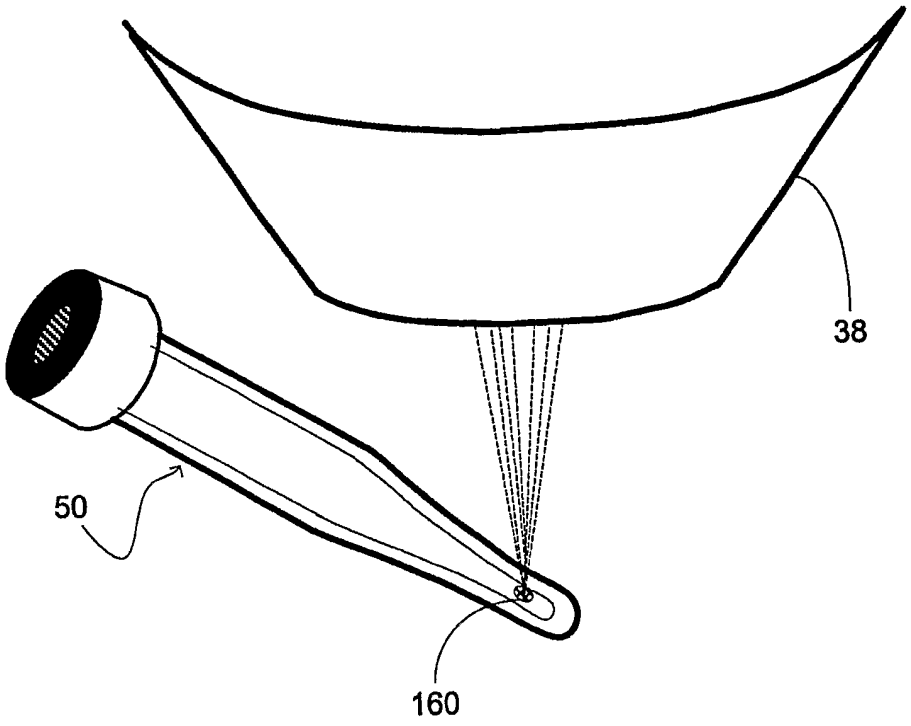


Fig. 14 D

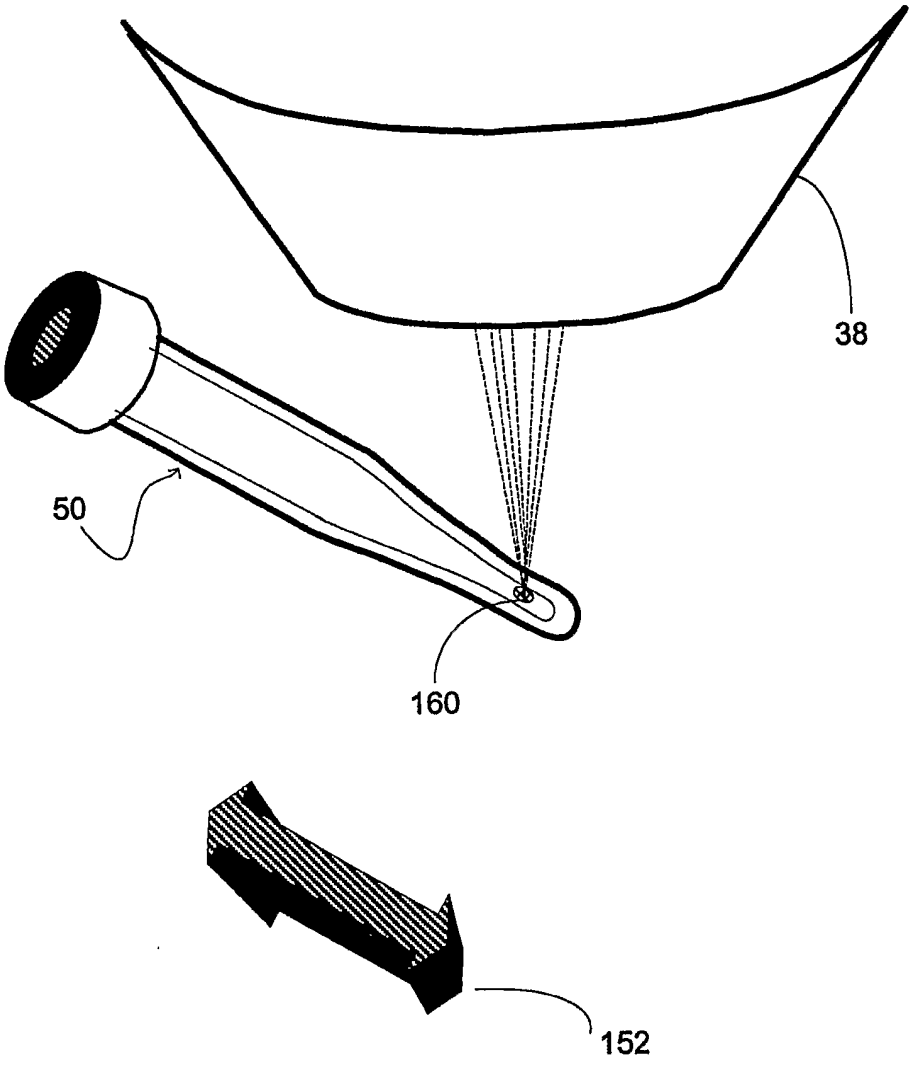


Fig. 14E

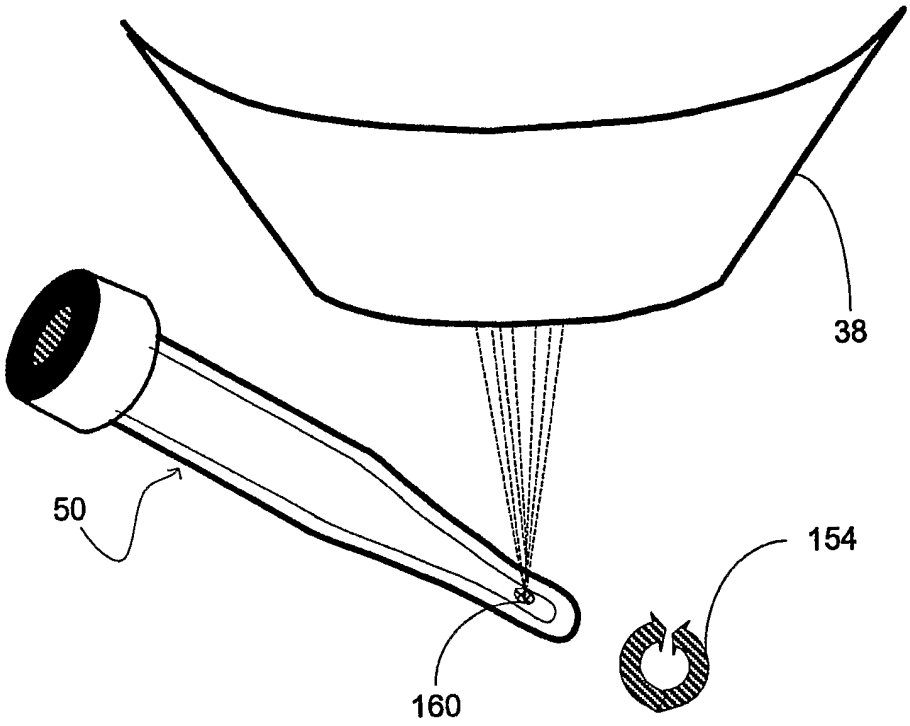


Fig. 14 F

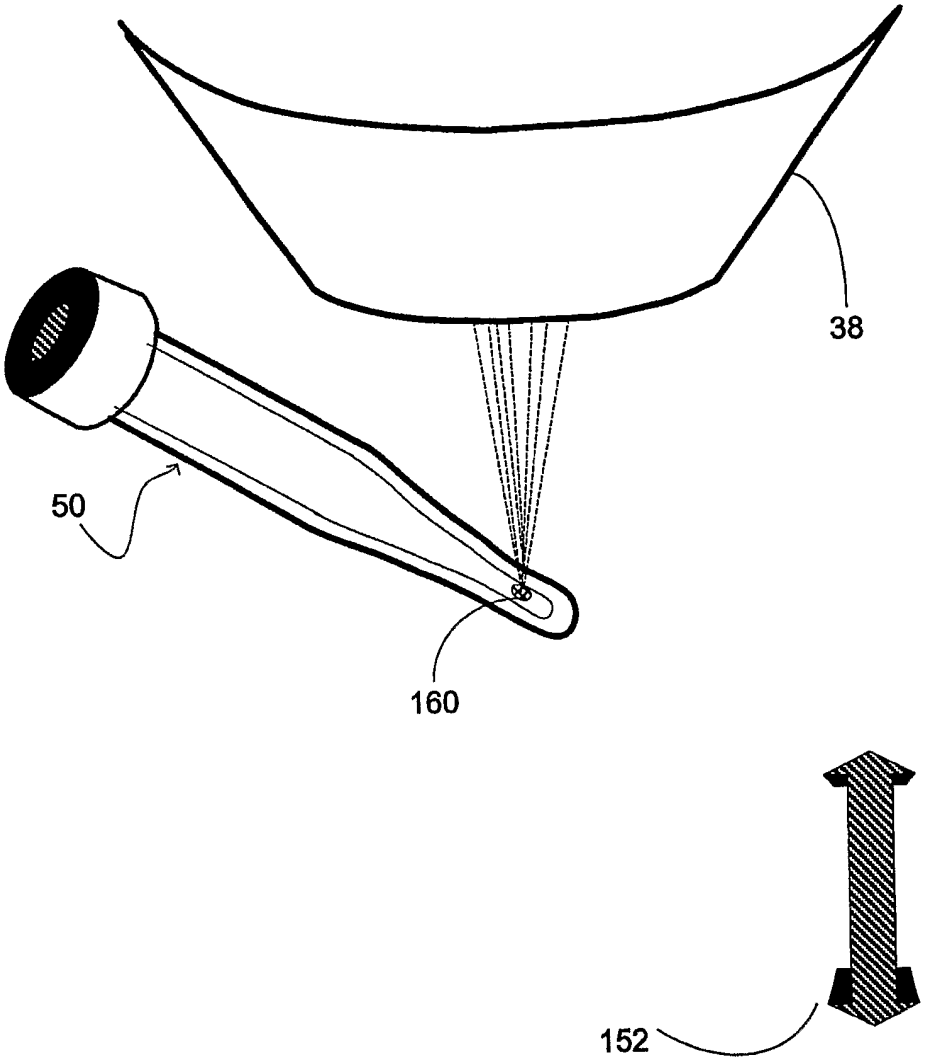


Fig. 14G

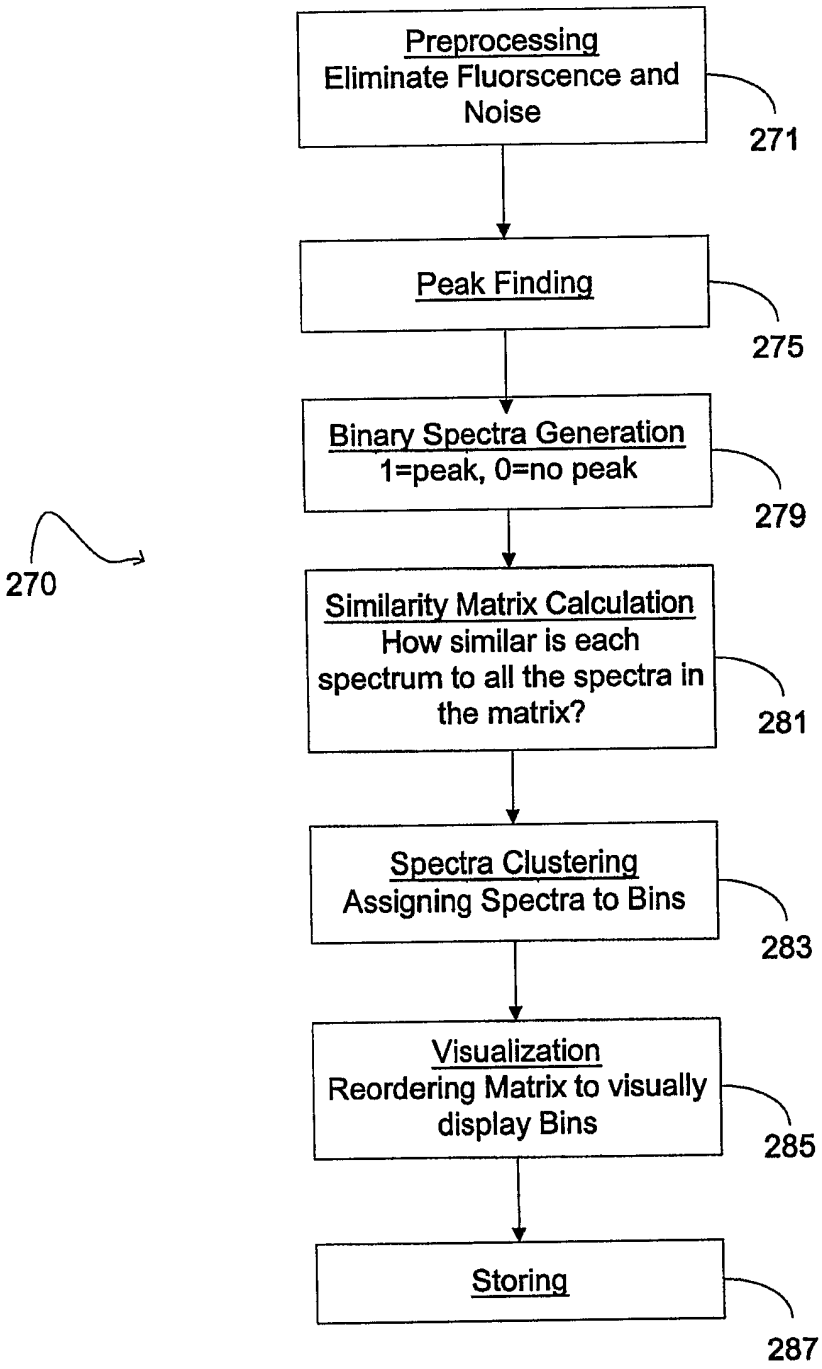

Preprocessing
Eliminate Fluorscence and Noise
271
Peak Finding
275
Binary Spectra Generation
1=peak, 0=no peak
279
270
Similarity Matrix Calculation
How similar is each spectrum to all the spectra in the matrix?
281
Spectra Clustering
Assigning Spectra to Bins
283
Visualization
Reordering Matrix to visually display Bins
285
Storing
287

Fig. 15

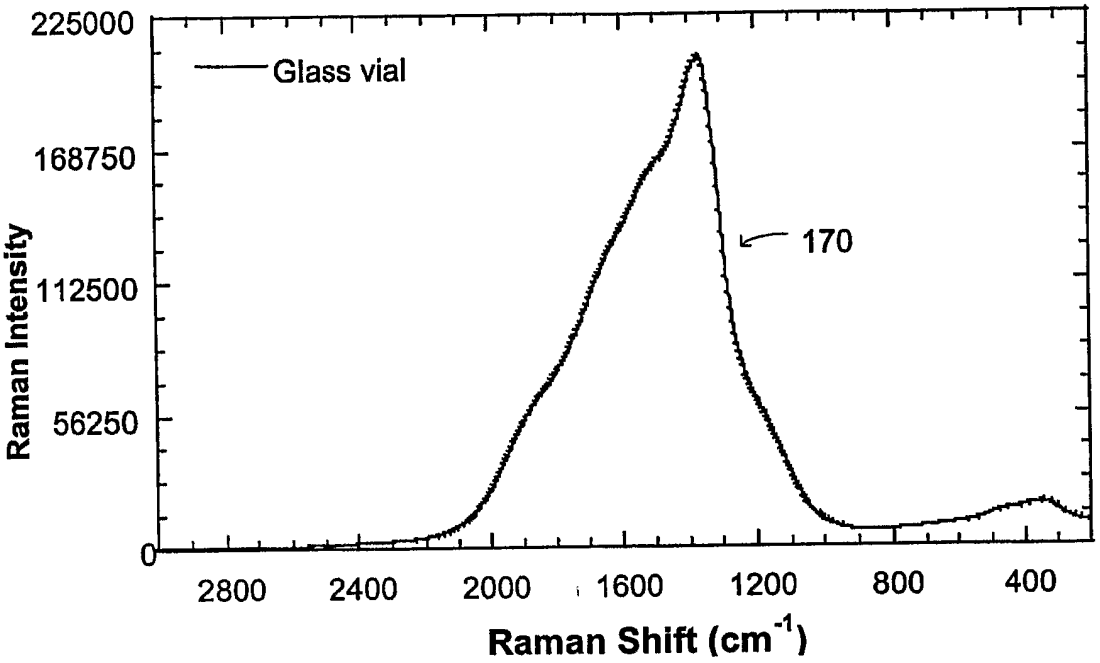


Fig. 16 A

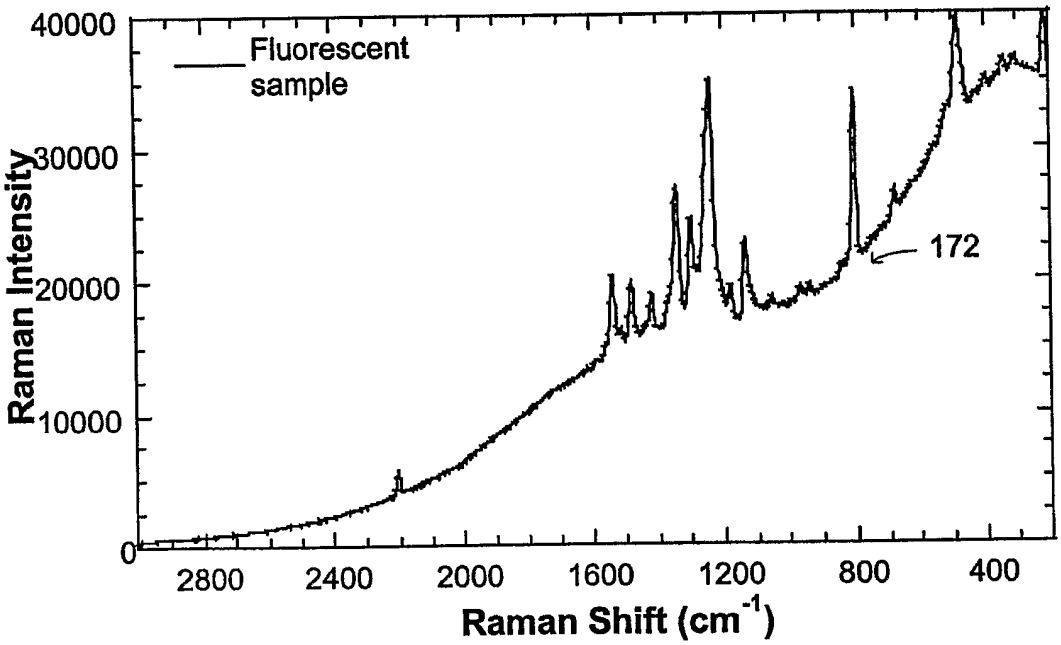


Fig. 16 B

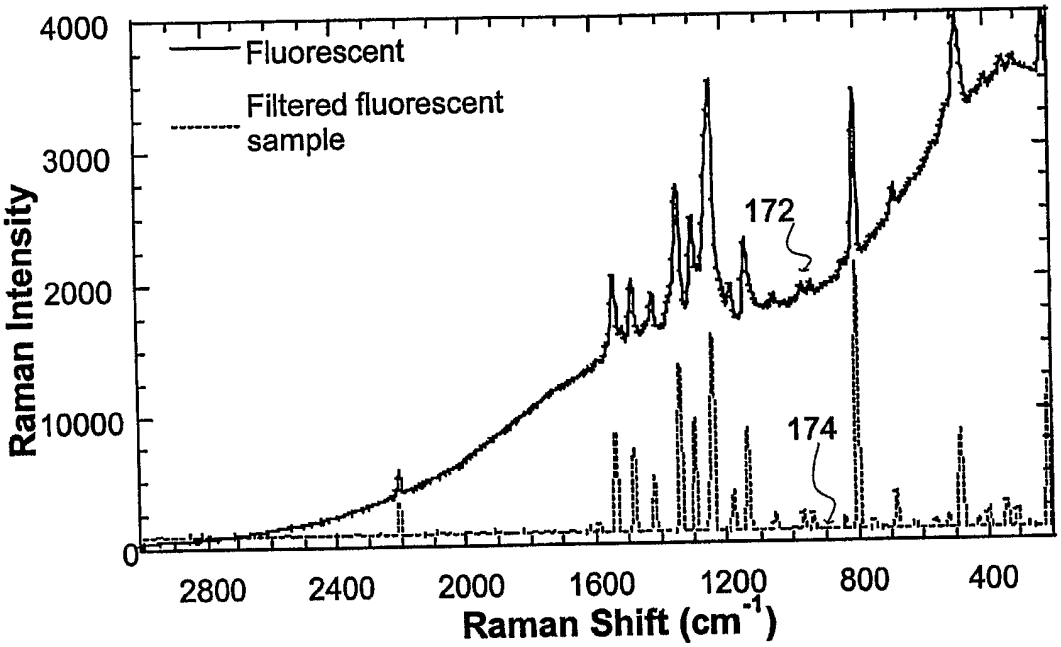

4000
3000
2000
10000
0
Raman Intensity
Fluorescent
Filtered fluorescent sample
172
174
2800
2400
2000
1600
1200
800
400
Raman Shift (cm-1)

Fig. 16 C

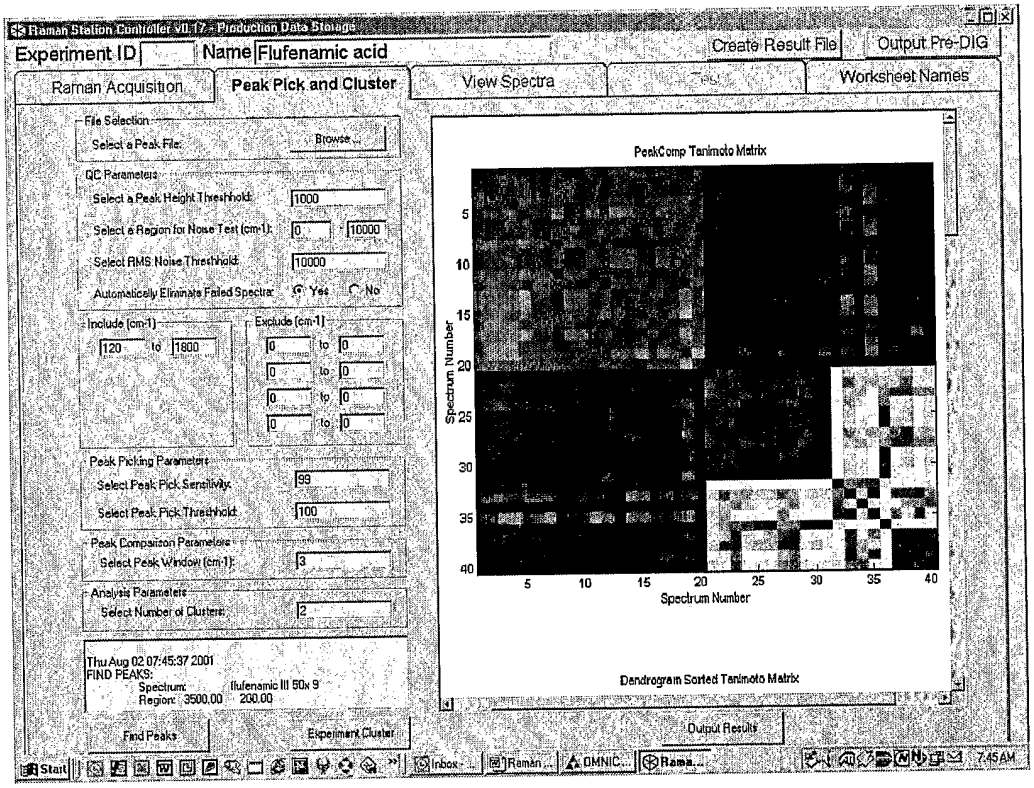

Raman Station Controller v0.17 - Production Data Storage
Experiment ID
Name Flufenamic acid
Create Result File
Output Pre-DIG
Raman Acquisition
Peak Pick and Cluster
View Spectra
Worksheet Names
File Selection
Select a Peak File:
Browse ...
QC Parameters
Select a Peak Height Threshold:
1000
Select a Region for Noise Test (cm-1):
0
10000
Select RMS Noise Threshold:
10000
Automatically Eliminate Failed Spectra:
Yes
No
Include (cm-1)
120
to
1800
Exclude (cm-1)
Peak Picking Parameters
Select Peak Pick Sensitivity:
99
Select Peak Pick Threshold:
100
Peak Comparison Parameters
Select Peak Window (cm-1):
3
Analysis Parameters
Select Number of Clusters:
2
Thu Aug 02 07:45:37 2001
FIND PEAKS:
Spectrum: flufenamic III 50x 9
Region: 3500.00 200.00
Find Peaks
Experiment Cluster
PeakComp Tanimoto Matrix
Spectrum Number
Dendrogram Sorted Tanimoto Matrix
Output Results
Start
Inbox...
Raman...
OMNIC...
7:45 AM

Fig. 17A

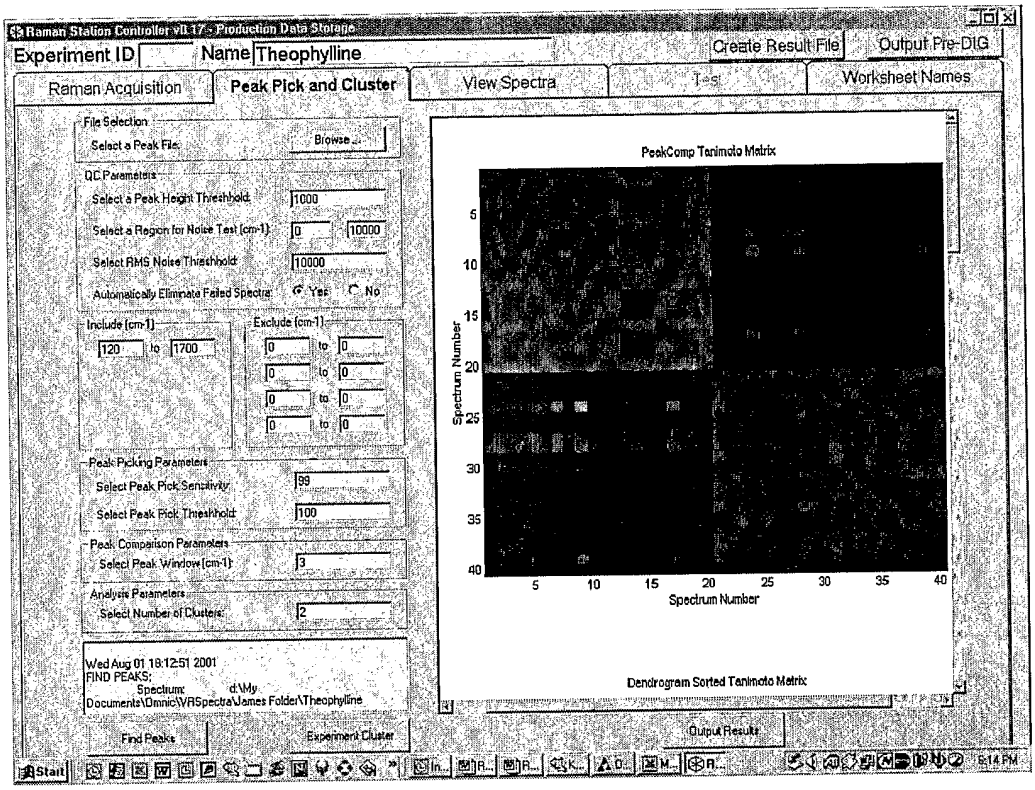

Raman Station Controller v11.17 - Production Data Storage
Experiment ID
Name Theophylline
Create Result File
Output Pre-DIG
Raman Acquisition
Peak Pick and Cluster
View Spectra
Test
Worksheet Names
File Selection
Select a Peak File:
Browse ...
QC Parameters
Select a Peak Height Threshhold:
1000
Select a Region for Noise Test (cm-1):
0
10000
Select RMS Noise Threshhold:
10000
Automatically Eliminate Failed Spectra:
Yes
No
Include (cm-1)
120
to
1700
Exclude (cm-1)
Peak Picking Parameters
Select Peak Pick Sensitivity:
99
Select Peak Pick Threshhold:
100
Peak Comparison Parameters
Select Peak Window (cm-1):
3
Analysis Parameters
Select Number of Clusters:
2
Wed Aug 01 18:12:51 2001
FIND PEAKS:
Spectrum: d:\My Documents\Omnic\VRSpectra\James Folder\Theophylline
Find Peaks
Experiment Cluster
Output Results
PeakComp Tanimoto Matrix
Spectrum Number
Dendrogram Sorted Tanimoto Matrix
Start
6:14 PM

Fig. 17B

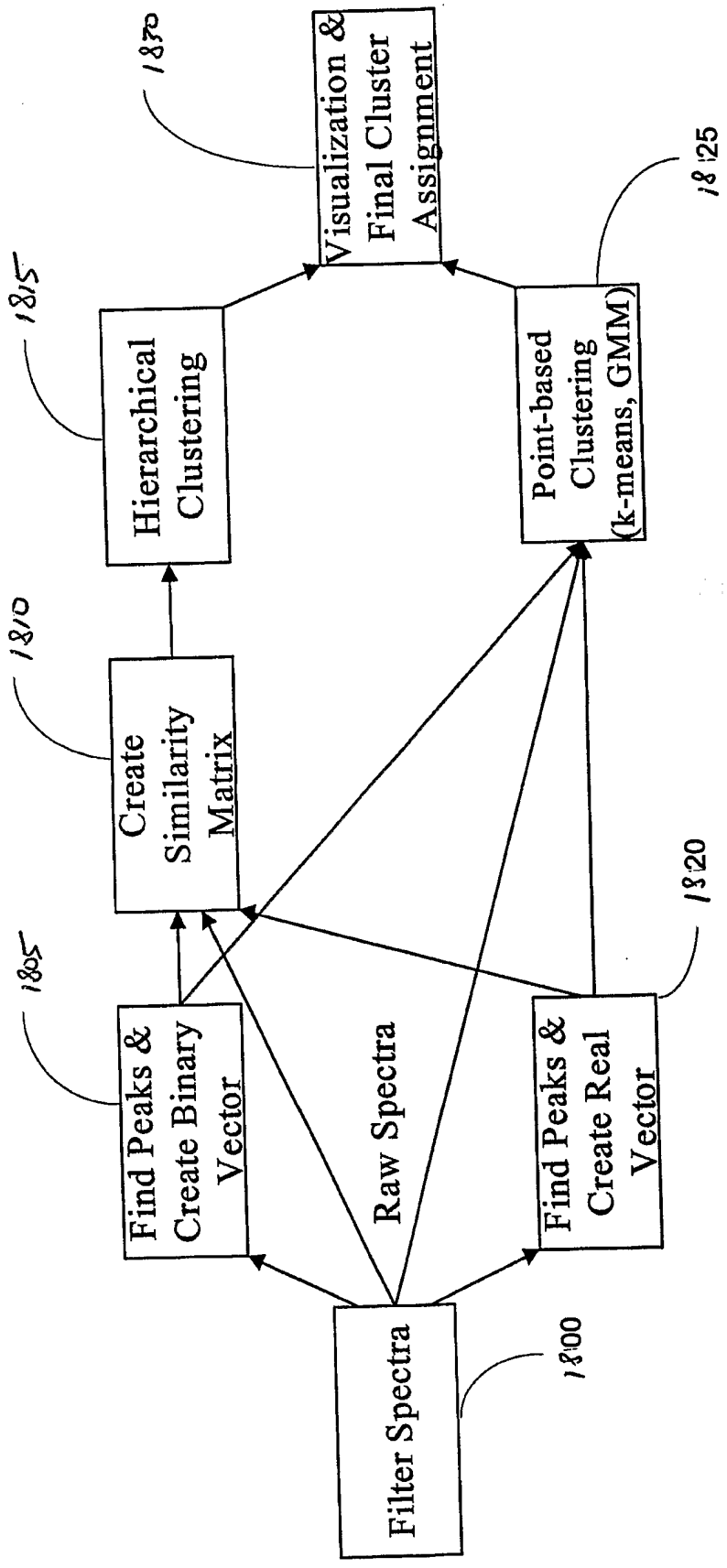

Filter Spectra
1800
Raw Spectra
Find Peaks & Create Binary Vector
1805
Find Peaks & Create Real Vector
1820
Create Similarity Matrix
1810
Hierarchical Clustering
1815
Point-based Clustering (k-means, GMM)
1825
Visualization & Final Cluster Assignment
1830

Fig. 18

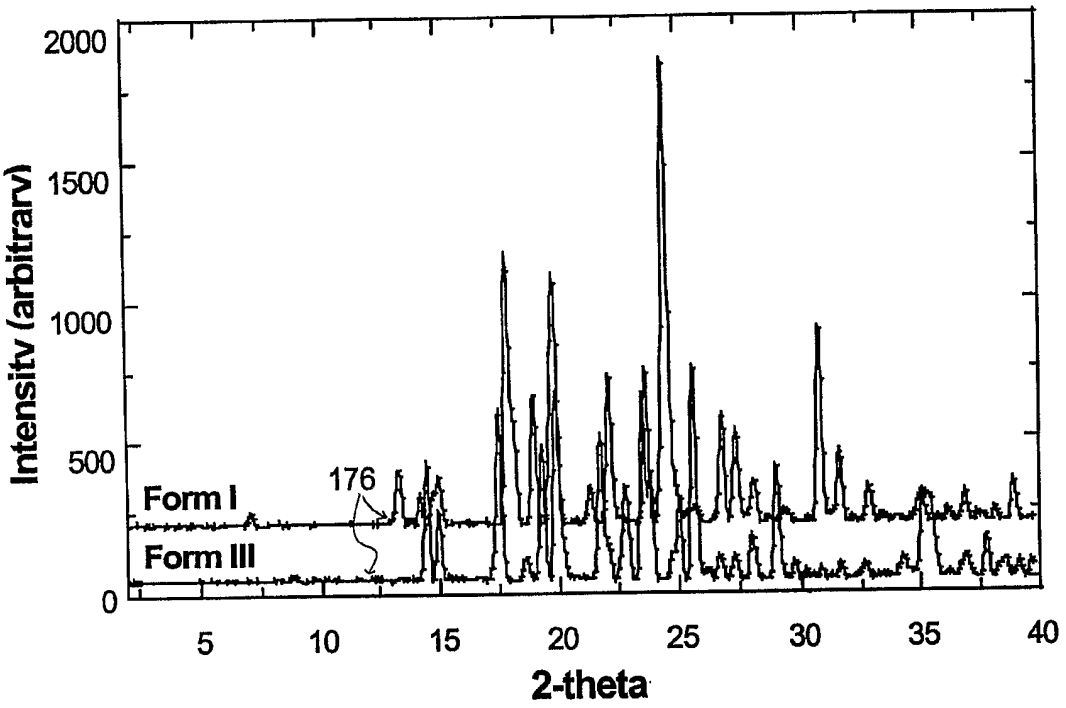


Fig. 19A

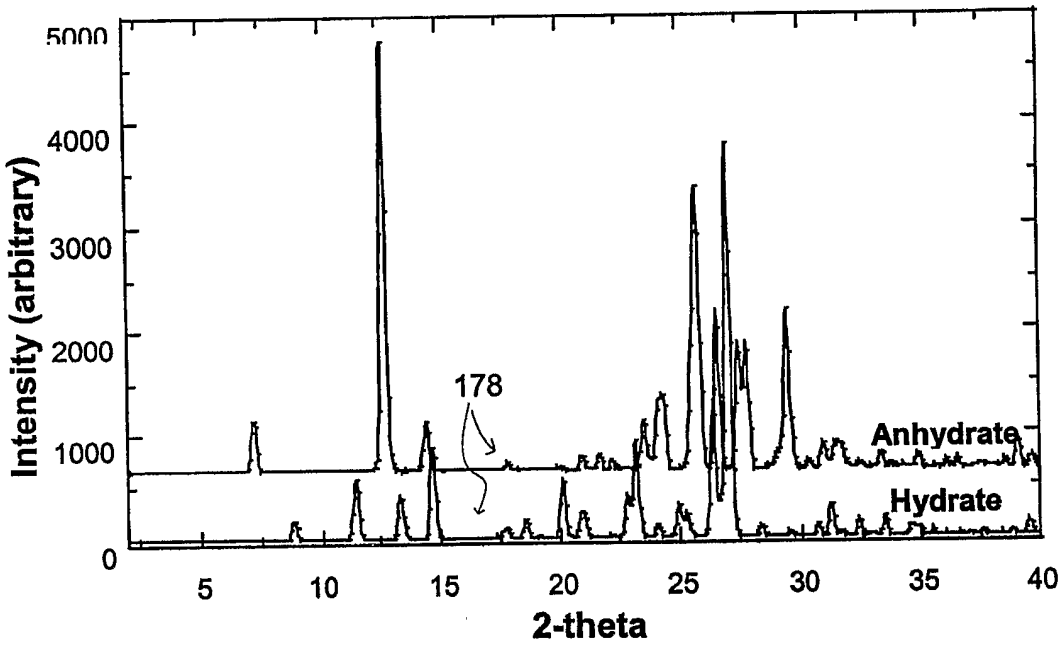


Fig. 19B

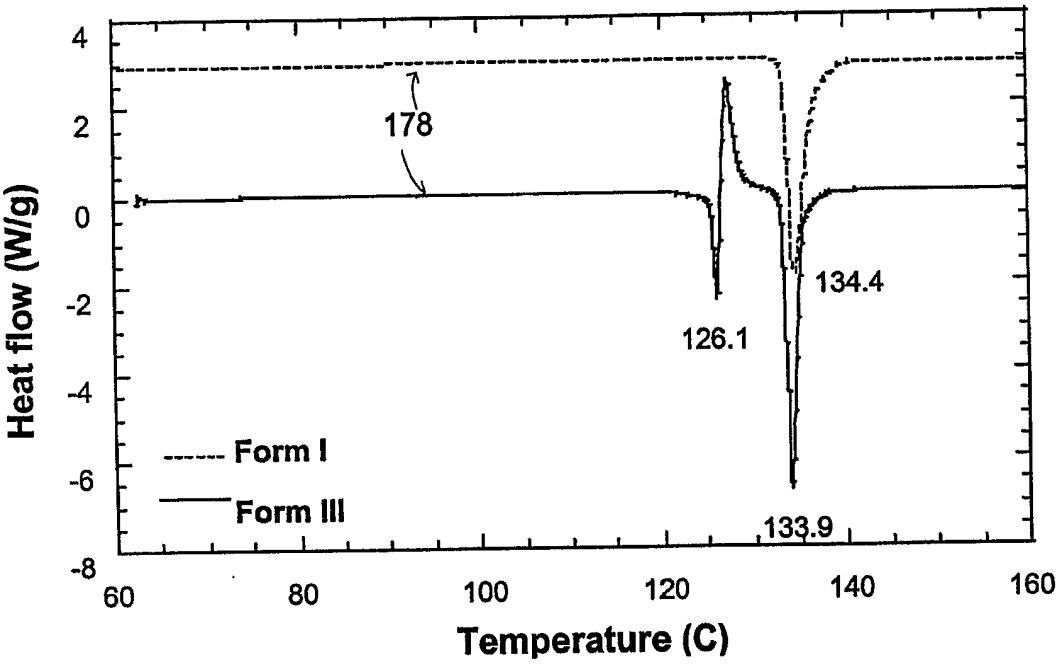


Fig. 20A

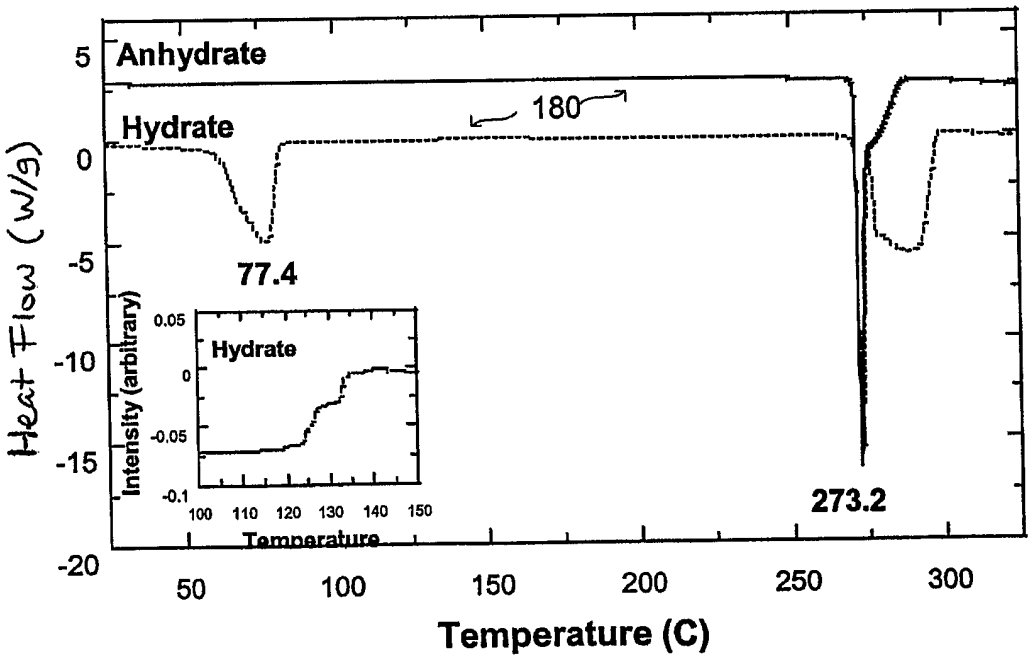


Fig. 20B

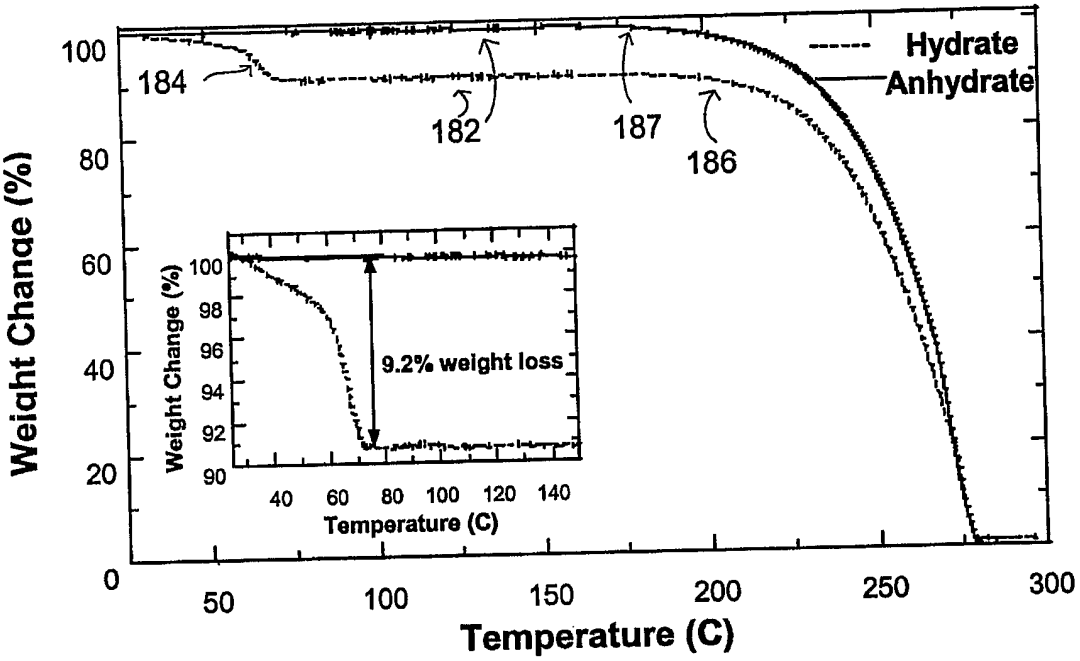

Hydrate
Anhydrate
184
182
187
186
Weight Change (%)
Temperature (C)
9.2% weight loss
100
80
60
40
20
0
50
100
150
200
250
300
98
96
94
92
90
40
60
80
100
120
140

Fig. 21

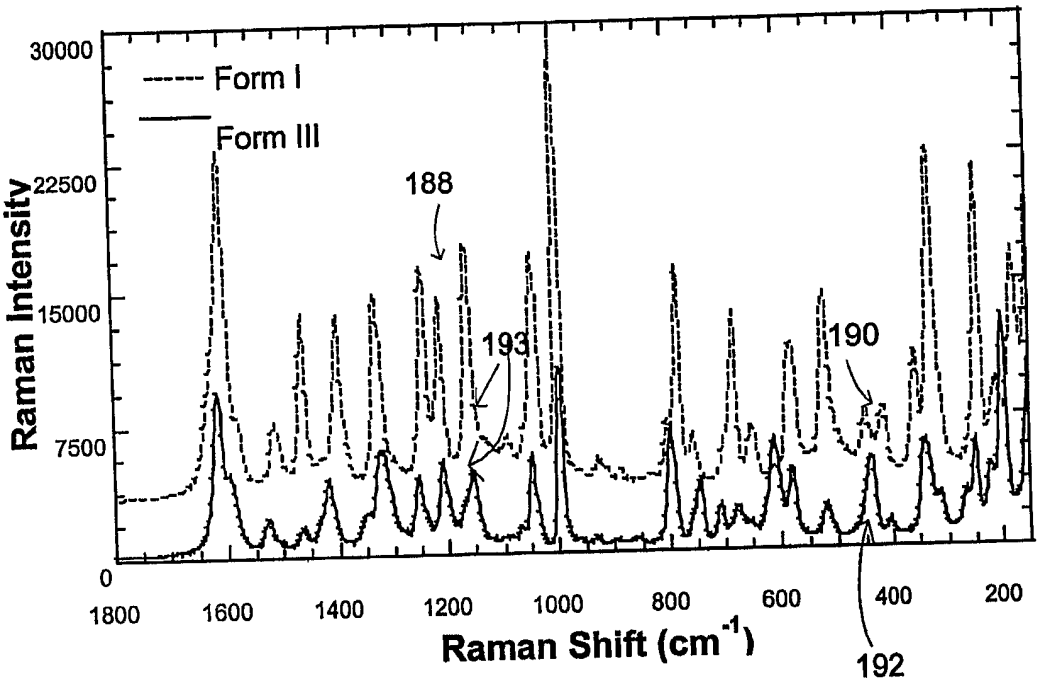

30000
22500
15000
7500
0
Raman Intensity
Form I
Form III
188
193
190
192
1800
1600
1400
1200
1000
800
600
400
200
Raman Shift (cm-1)

Fig. 22 A

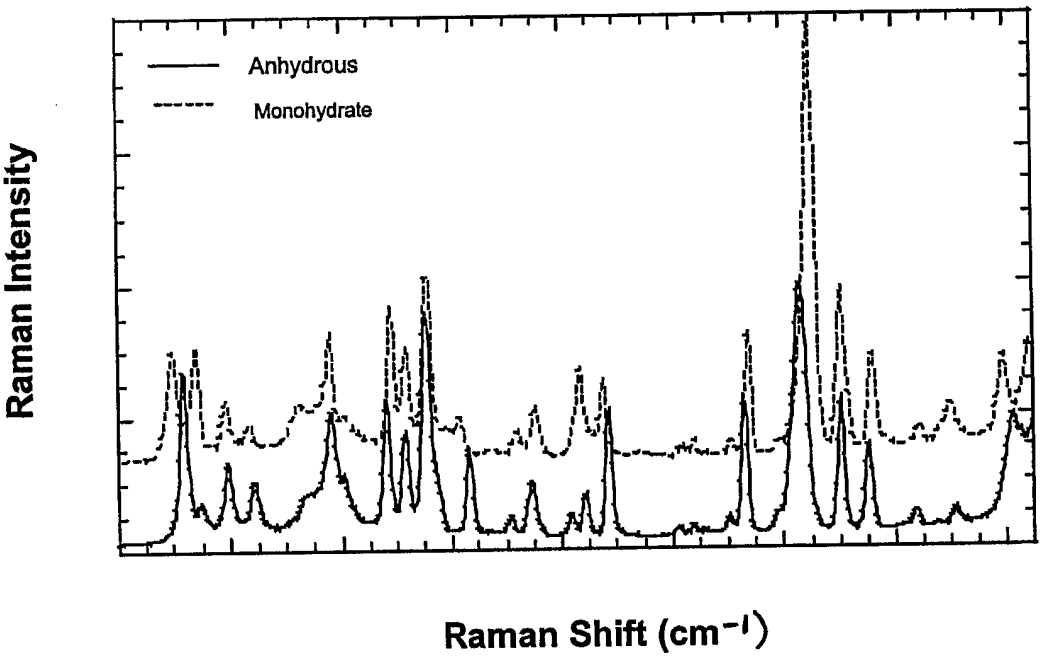


Fig. 22B

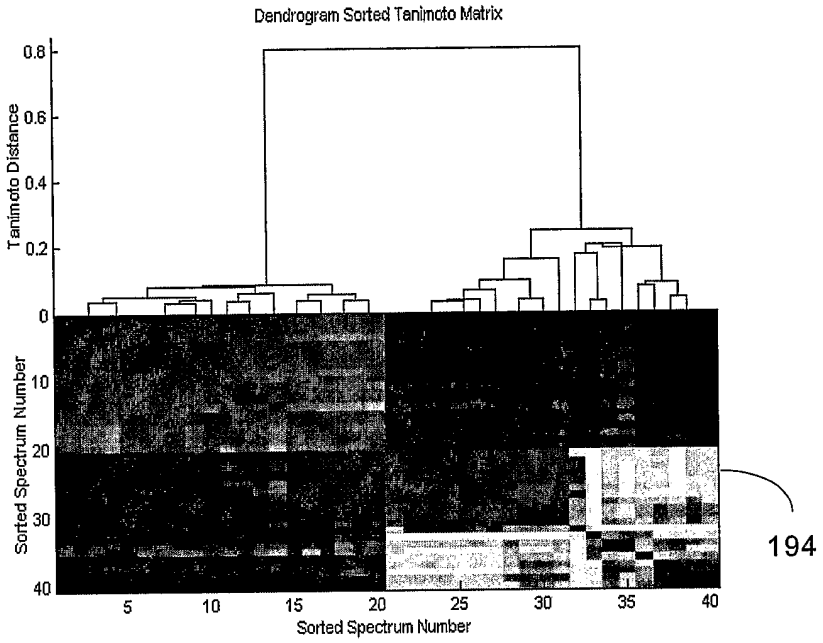

Dendrogram Sorted Tanimoto Matrix
Tanimoto Distance
0.8
0.6
0.4
0.2
0
Sorted Spectrum Number
10
20
30
40
5
10
15
20
25
30
35
40
Sorted Spectrum Number
194

Fig. 23A

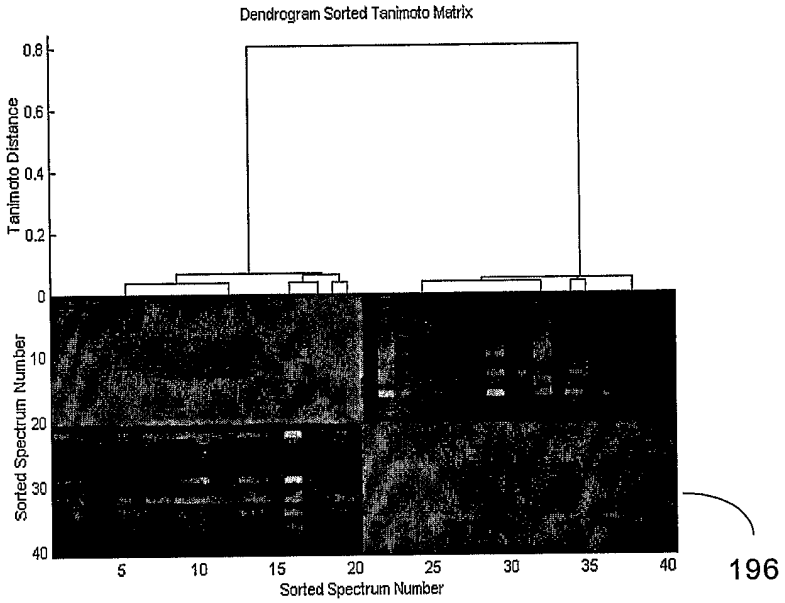

Dendrogram Sorted Tanimoto Matrix
Tanimoto Distance
0.8
0.6
0.4
0.2
0
Sorted Spectrum Number
10
20
30
40
5
10
15
20
25
30
35
40
Sorted Spectrum Number
196

Fig. 23B

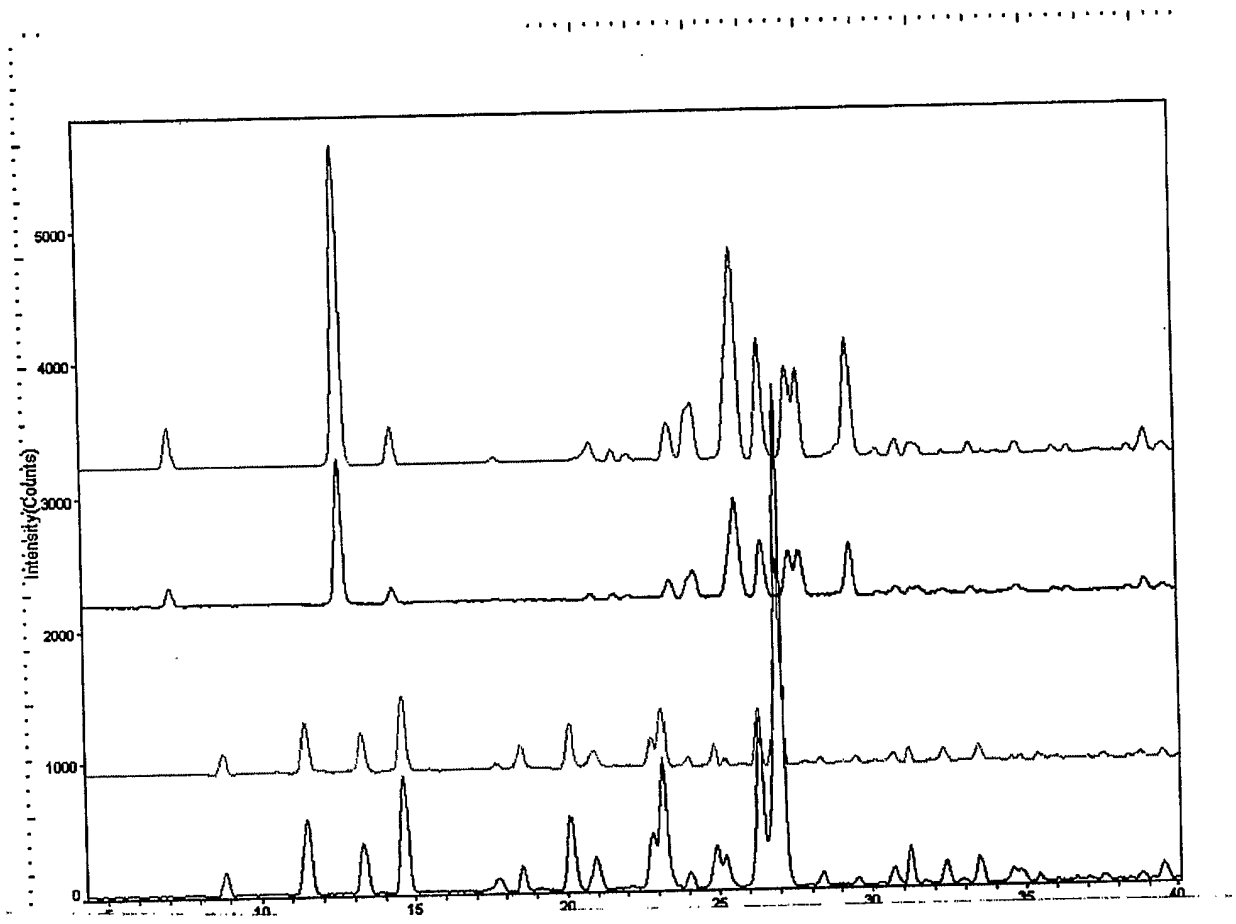

Intensity(Counts)
5000
4000
3000
2000
1000
0
10
15
20
25
30
35
40

Fig. 24A

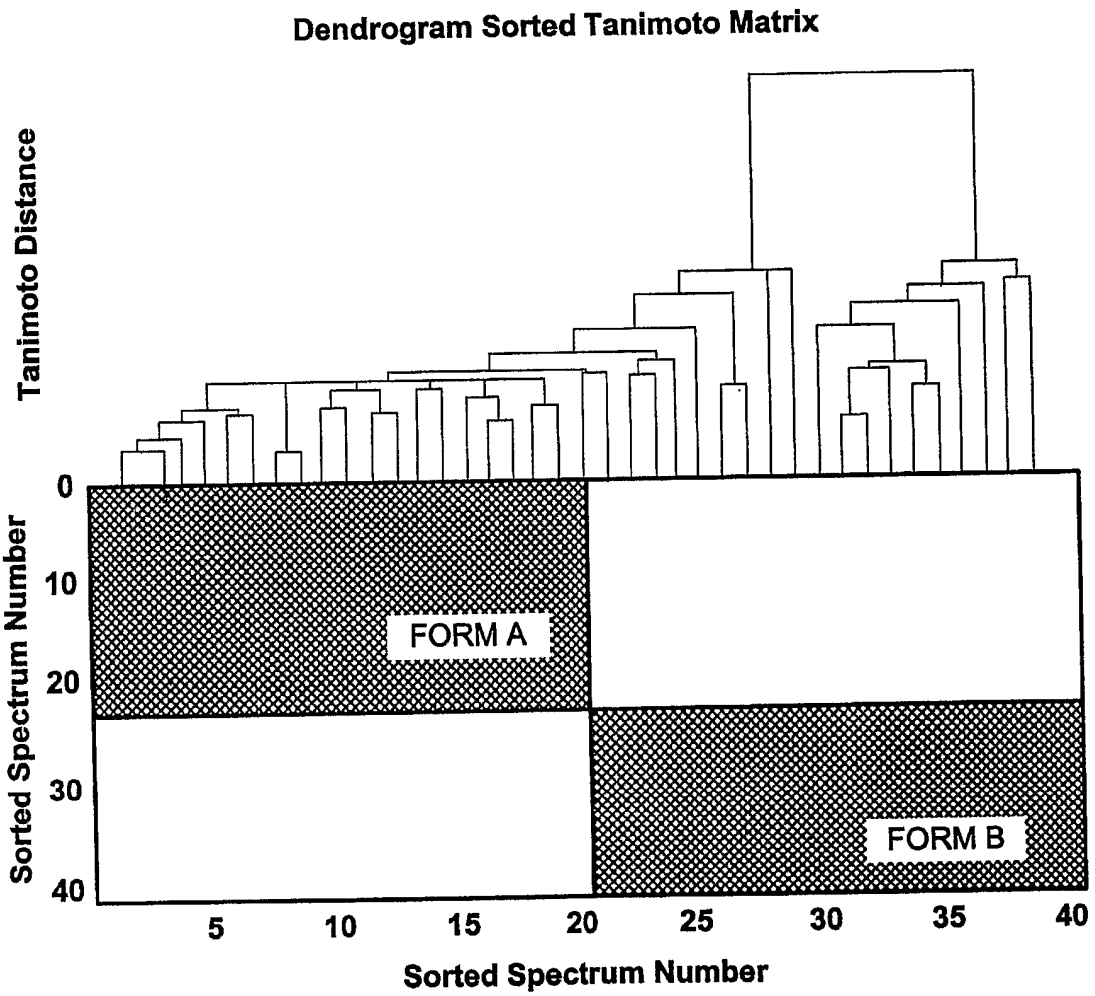

Dendrogram Sorted Tanimoto Matrix
Tanimoto Distance
0
Sorted Spectrum Number
10
20
30
40
FORM A
FORM B
5
10
15
20
25
30
35
40
Sorted Spectrum Number

Fig. 24B