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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 September 2007 (07.09.2007)

PCT

(10) International Publication Number
WO 2007/100748 A2

(51) International Patent Classification:
C12M 3/00 (2006.01)

(21) International Application Number:
PCT/US2007/004881

(22) International Filing Date:
26 February 2007 (26.02.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/776,427 24 February 2006 (24.02.2006) US

(71) Applicant *(for all designated States except US)*: **THE REGENTS OF THE UNIVERSITY OF MICHIGAN** [US/US]; 3003 South State Street, Ann Arbor, MI 48109 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **YOO, Chul** [KR/US]; 3525 Greenbrier Boulevard #18B, Ann Arbor, MI 48105 (US). **ZHENG, Suping** [CN/US]; 1789 Beal Avenue #3, Ann Arbor, MI 48105 (US). **LUBMAN, David** [US/US]; 2012 Shadford Road, Ann Arbor, MI 48104 (US). **PAL, Manoj** [IN/US]; 1929 Plymouth Road, Apt. 5032, Ann Arbor, MI 48105 (US).

(74) Agents: **ARENSON, Tanya, A.** et al.; Medlen & Carroll, LLP, 101 Howard Street, Suite 350, San Francisco, CA 94105 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:
— *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.

WO 2007/100748 A2

(54) Title: PROTEIN SEPARATION AND ANALYSIS

(57) Abstract: The present invention relates to multi-phase protein separation and analysis methods capable of resolving and identifying large numbers of cellular proteins. In particular, the present invention provides systems and methods for high resolution separation and on-plate digestion coupled with mass spectrometry.

PROTEIN SEPARATION AND ANALYSIS

This application claims priority to provisional patent application serial number 60/776,427, filed 2/24/06, which is herein incorporated by reference in its entirety.

The present invention was made with government support under grant No. R01GM49500 from the national institutes of health, grant No. NCC2-1274 from NASA and grant No. RO1CA100104 from the National Cancer institute. The government has certain rights in this invention.

FIELD OF THE INVENTION

The present invention relates to multi-phase protein separation and analysis methods capable of resolving and identifying large numbers of cellular proteins. In particular, the present invention provides systems and methods for high resolution separation and on-plate digestion coupled with mass spectrometry.

BACKGROUND OF THE INVENTION

As the nucleic acid sequences of a number of genomes, including the human genome, become available, there is an increasing need to interpret this wealth of information. While the availability of nucleic acid sequence allows for the prediction and identification of genes, it does not explain the expression patterns of the proteins produced from these genes. The genome does not describe the dynamic processes on the protein level. For example, the identity of genes and the level of gene expression does not represent the amount of active protein in a cell nor does the gene sequence describe post-translational modifications that are essential for the function and activity of proteins. Thus, in parallel with the genome projects there has begun an attempt to understand the proteome (*i.e.*, the quantitative protein expression pattern of a genome under defined conditions) of various cells, tissues, and species. Proteome research seeks to identify targets for drug discovery and development and provide information for diagnostics (*e.g.*, tumor markers).

An important area of research is the study of the protein content of cells (*i.e.*, the identity of and amount of expressed proteins in a cell). This field requires methods that can separate out large numbers of proteins and can do so quantitatively so that changes in expression or structure of proteins can be detected. The method generally used to achieve

such cellular protein separations is 2-D PAGE. This method is capable of resolving hundreds of proteins based upon pI in one dimension and protein size in the second dimension. The proteins separated by this method are visualized using a staining method that can generally be quantified. The result is a 2-dimensional image where the protein map is based on pI and approximate molecular weight. By the use of computer based image analysis techniques, one can search for proteins that are differentially expressed in various cell lines. These methods are used to monitor changes in protein expression that are linked to conditions such as cell transformation and cancer progression, cell aging, the response of cells to environmental insult, and the response of cells to pharmaceutical agents. Once changes in protein expression have been identified, then one can further analyze target proteins to determine their identity and whether they have been altered from their expected structure by sequence changes or post-translational modifications.

Although 2-D PAGE is still widely used for protein analysis, the method has several limitations including the fact that it is labor intensive, time consuming, difficult to automate and often not readily reproducible. In addition, quantitation, especially in differential expression experiments, is often difficult and limited in dynamic range. Also, while the 2-D gel does produce an image of the proteins in the cell, the mass determination is often only accurate to 5-10%, and the method is difficult to interface to mass spectrometric techniques for further analysis.

Another limitation of 2-D PAGE is the amount of protein loaded per gel which is generally below 250 μg. The amount of protein in any given spot may therefore be too low for further analysis. For Coomassie brilliant blue (CBB) stained gels the limit of detection is 100 ng per spot while for silver stained gels the limit of detection is 1 □ 10 ng. Furthermore, proteins that have been isolated in 2-D gels are embedded inside the gel structure and are not free in solution, thus making it difficult to extract the protein for further analysis. Because of these limitations, the art is in need of protein mapping methods that are efficient, automated, and have broader resolution capabilities than presently available technologies.

SUMMARY OF THE INVENTION

The present invention relates to multi-phase protein separation and analysis methods capable of resolving and identifying large numbers of cellular proteins. In particular, the present invention provides systems and methods for high resolution separation and on-plate

digestion coupled with mass spectrometry.

Accordingly, in some embodiments, the present invention provides systems and methods for the separation and analysis for the rapid analysis and identification of proteins. The methods of the present invention are able to rapidly identify large numbers of proteins from a small amount of starting material.

For example, in some embodiments, the present invention provides a system, comprising a reverse phase (e.g., monolithic capillary) HPLC apparatus; and a MALDI mass spectrometry apparatus comprising a MALDI plate configured for on-plate digestion of proteins separated by the reverse phase monolithic capillary HPLC apparatus. In some embodiments, the system further comprises an automated sample handling apparatus that transfers protein samples from the reverse phase monolithic capillary HPLC apparatus to the MALDI mass spectrometry apparatus. In some embodiments, the automated sample handling apparatus is a robot. In some embodiments, the MALDI mass spectrometry apparatus is configured to analyze intact proteins. For example, in some embodiments, the MALDI mass spectrometry apparatus is configured to determine the identity of the proteins separated by the reverse phase monolithic capillary HPLC apparatus. In certain embodiments, the system further comprises software for analyzing data generated by the MALDI mass spectrometry apparatus. In some embodiments, the MALDI mass spectrometry apparatus comprises a MALDI-TOF mass spectrometer. In some embodiments, the system further comprises a non-porous reverse phase HPLC apparatus. In some embodiments, the system further comprises an isoelectic focusing apparatus. In some embodiments, the system further comprises a micro-chromatofocusing apparatus. In some embodiments, the micro-chromatofocusing apparatus is configured for the separation of approximately 1-500, and preferably 5-40 micrograms of protein (e.g., cell extract).

The present invention further provides a method, comprising treating a protein sample with a reverse phase (e.g., monolithic capillary) HPLC apparatus under conditions such that the reverse phase HPLC apparatus separates the protein sample into a plurality of protein fractions; digesting at least a portion of the protein fractions on a MALDI plate to generate digested protein fractions; and analyzing the digested protein fractions with a MALDI mass spectrometer. In some embodiments, the MALDI mass spectrometer is a MALDI-TOF mass spectrometer. In some preferred embodiments, the method is automated. In some embodiments, the protein fractions are transferred from the reverse phase HPLC apparatus to the MALDI plate using an automated sample handling apparatus

(e.g., a robot). In some embodiments, the protein fractions comprise intact proteins. In some embodiments, the analyzing step further comprises determining the identity of at least 5, and preferably at least 25 proteins in the digested protein fractions. In some embodiments, the determining the identity of the digested protein fractions is performed by computer software and a computer processor. In some embodiments, the method further comprises the step of separating said proteins with non-porous reverse phase HPLC prior to separating the proteins with the monolithic reverse phase HPLC. In other embodiments, the method further comprises the step of separating the proteins with isoelectic focusing prior to separating the protein with the monolithic HPLC. In yet other embodiments, the method further comprises separating the proteins with isoelectric focusing followed by non-porous reverse phase HPLC prior to separating the proteins with the monolithic HPLC. In some embodiments, less than 200 ng of total protein is utilized for the monolithic HPLC separation. In some embodiments, prior to the treating step, the protein sample is separated with a micro-chromatofocusing apparatus. In some embodiments, the micro-chromatofocusing apparatus is configured for the separation of approximately 1-500, and more preferably 5-40 micrograms of protein (e.g., cell extract).

Other embodiments of the invention are described in the description and examples below.

DESCRIPTION OF THE FIGURES

Figure 1 shows an experimental scheme of the two-dimensional liquid phase separation followed by on-plate digestion for MALDI-TOF MS analysis for identification of proteins in human breast cancer cell line used in some embodiments of the present invention.

Figure 2 shows MALDI spectra from monolithic capillary HPLC protein collection times of (A) 5.0-5.5 min, (B) 5.5 to 6.0 min, and (C) 6.0-6.5 min.

Figure 3 shows Table 1.

Definitions

To facilitate an understanding of the present invention, a number of terms and phrases are defined below:

As used herein, the term "micro-chromatofocusing apparatus" refers to a chromatofocusing column suitable for the separation of small amounts of protein. For

example, in some embodiments, micro-chromatofusing columns of the present invention are suitable for the separation of less than 5 milligrams, preferably less than 1000 micrograms, even more preferably less than 500 micrograms, yet more preferably less than 100 micrograms, and most preferably less than 50 micrograms of protein (e.g., cell extract). In some particularly preferred embodiments, the micro-chromatofucusing columns of present invention are suitable for the separation of approximately 5-40 micrograms of protein.

As used herein, the term "multiphase protein separation" refers to protein separation comprising at least two separation steps. In some embodiments, multiphase protein separation refers to two or more separation steps that separate proteins based on different physical properties of the protein (*e.g.*, a first step that separates based on protein charge and a second step that separates based on protein hydrophobicity).

As used herein, the term "protein profile maps" refers to representations of the protein content of a sample. For example, "protein profile map" includes 2-dimensional displays of total protein expressed in a given cell. In some embodiments, protein profile maps may also display subsets of total protein in a cell. Protein profile maps may be used for comparing "protein expression patterns" (*e.g.*, the amount and identity of proteins expressed in a sample) between two or more samples. Such comparing find use, for example, in identifying proteins that are present in one sample (*e.g.*, a cancer cell) and not in another (*e.g.*, normal tissue), or are over- or under-expressed in one sample compared to the other.

As used herein, the term "2-dimensional protein map" refers to a "protein profile map" that represents (*e.g.*, on two axis of a graph) two properties of the protein content of a sample (*e.g.*, including but not limited to, hydrophobicity and isoelectric point).

As used herein the term "differential display map" and equivalents "differential display plot" and "differential display image" refer to a "protein profile map" that shows the subtraction of one protein profile map from another protein profile map. A differential display map thus shows the differences in proteins present between two samples. A differential display image may also show differences in the abundance of a protein between the two samples. In some embodiments, multiple colors or color gradients are used to represent proteins from each of the two samples.

As used herein, the term "separating apparatus capable of separating proteins based on a physical property" refers to compositions or systems capable of separating proteins (*e.g.*, at least one protein) from one another based on differences in a physical property

between proteins present in a sample containing two or more protein species. For example, a variety of protein separation columns and compositions are contemplated including, but not limited to ion exclusion, ion exchange, normal/reversed phase partition, size exclusion, ligand exchange, liquid/gel phase isoelectric focusing, and adsorption chromatography. These and other apparatuses are capable of separating proteins from one another based on their size, charge, hydrophobicity, and ligand binding affinity, among other properties. A "liquid phase" separating apparatus is a separating apparatus that utilizes protein samples contained in liquid solution, wherein proteins remain solubilized in liquid phase during separation and wherein the product (*e.g.*, fractions) collected from the apparatus are in the liquid phase. This is in contrast to gel electrophoresis apparatuses, wherein the proteins enter into a gel phase during separation. Liquid phase proteins are much more amenable to recovery/extraction of proteins as compared to gel phase. In some embodiments, liquid phase proteins samples may be used in multi-step (*e.g.*, multiple separation and characterization steps) processes without the need to alter the sample prior to treatment in each subsequent step (*e.g.*, without the need for recovery/extraction and resolubilization of proteins).

As used herein, the term "displaying proteins" refers to a variety of techniques used to interpret the presence of proteins within a protein sample. Displaying includes, but is not limited to, visualizing proteins on a computer display representation, diagram, autoradiographic film, list, table, chart, etc. "Displaying proteins under conditions that first and second physical properties are revealed" refers to displaying proteins (*e.g.*, proteins, or a subset of proteins obtained from a separating apparatus) such that at least two different physical properties of each displayed protein are revealed or detectable. For example, such displays include, but are not limited to, tables including columns describing (*e.g.*, quantitating) the first and second physical property of each protein and two-dimensional displays where each protein is represented by an X,Y locations where the X and Y coordinates are defined by the first and second physical properties, respectively, or vice versa. Such displays also include multi-dimensional displays (*e.g.*, three dimensional displays) that include additional physical properties. In some embodiments, displays are generated by "display software."

As used herein, "characterizing protein samples under conditions such that first and second physical properties are analyzed" refers to the characterization of two or more proteins, wherein two different physical properties are assigned to each analyzed (*e.g.*,

displayed, computed, etc.) protein and wherein a result of the characterization is the categorization (*i.e.*, grouping and/or distinguishing) of the proteins based on these two different physical properties. For example, in some embodiments, two proteins are separated based on isoelectric point and hydrophobicity.

As used herein, the term "comparing first and second physical properties of separated protein samples" refers to the comparison of two or more protein samples (or individual proteins) based on two different physical properties of the proteins within each protein sample. Such comparing includes grouping of proteins in the samples based on the two physical properties and comparing certain groups based on just one of the two physical properties (*i.e.*, the grouping incorporates a comparison of the other physical property).

As used herein, the term "delivery apparatus capable of receiving a separated protein from a separating apparatus" refers to any apparatus (*e.g.*, microtube, trough, chamber, etc.) that receives one or more fractions or protein samples from a protein separating apparatus and delivers them to another apparatus (*e.g.*, another protein separation apparatus, a reaction chamber, a mass spectrometry apparatus, etc.).

As used herein, the term "detection system capable of detecting proteins" refers to any detection apparatus, assay, or system that detects proteins derived from a protein separating apparatus (*e.g.*, proteins in one or more fractions collected from a separating apparatus). Such detection systems may detect properties of the protein itself (*e.g.*, UV spectroscopy) or may detect labels (*e.g.*, fluorescent labels) or other detectable signals associated with the protein. The detection system converts the detected criteria (*e.g.*, absorbance, fluorescence, luminescence etc.) of the protein into a signal that can be processed or stored electronically or through similar means (*e.g.*, detected through the use of a photomultiplier tube or similar system).

As used herein, the term "buffer compatible with an apparatus" and "buffer compatible with mass spectrometry" refer to buffers that are suitable for use in such apparatuses (*e.g.*, protein separation apparatuses) and techniques. A buffer is suitable where the reaction that occurs in the presence of the buffer produces a result consistent with the intended purpose of the apparatus or method. For example, a buffer compatible with a protein separation apparatus solubilizes the protein and allows proteins to be separated and collected from the apparatus. A buffer compatible with mass spectrometry is a buffer that solubilizes the protein or protein fragment and allows for the detection of ions following mass spectrometry. A suitable buffer does not substantially interfere with the apparatus or

method so as to prevent its intended purpose and result (*i.e.*, some interference may be allowed).

As used herein, the term "automated sample handling device" refers to any device capable of transporting a sample (*e.g.*, a separated or un-separated protein sample) between components (*e.g.*, separating apparatus) of an automated method or system (*e.g.*, an automated protein characterization system). An automated sample handling device may comprise physical means for transporting sample (*e.g.*, multiple lines of tubing connected to a multi-channel valve). In some embodiments, an automated sample handling device is connected to a centralized control network. In some embodiments, the automated sample handling device is a robotic device.

As used herein, the term "switchable multi channel valve" refers to a valve that directs the flow of liquid through an automated sample handling device. The valve preferably has a plurality of channels (*e.g.*, 2 or more, and preferably 4 or more, and more preferably, 6 or more). In addition, in some embodiments, flow to individual channels is "switched" on an off. In some embodiments, valve switching is controlled by a centralized control system. A switchable multi-channel valve allows multiple apparatus to be connected to one automated sample handler. For example, sample can first be directed through one apparatus of a system (*e.g.*, a first chromatography apparatus). The sample can then be directed through a different channel of the valve to a second apparatus (*e.g.*, a second chromatography apparatus).

As used herein, the terms "centralized control system" or "centralized control network" refer to information and equipment management systems (*e.g.*, a computer processor and computer memory) operable linked to multiple devices or apparatus (*e.g.*, automated sample handling devices and separating apparatus). In preferred embodiments, the centralized control network is configured to control the operations or the apparatus an device linked to the network. For example, in some embodiments, the centralized control network controls the operation of multiple chromatography apparatus, the transfer of sample between the apparatus, and the analysis and presentation of data.

As used herein, the terms "computer memory" and "computer memory device" refer to any storage media readable by a computer processor. Examples of computer memory include, but are not limited to, RAM, ROM, computer chips, digital video disc (DVDs), compact discs (CDs), hard disk drives (HDD), and magnetic tape.

As used herein, the term "computer readable medium" refers to any device or system

for storing and providing information (*e.g.*, data and instructions) to a computer processor. Examples of computer readable media include, but are not limited to, DVDs, CDs, hard disk drives, magnetic tape and servers for streaming media over networks.

As used herein, the terms "processor" and "central processing unit" or "CPU" are used interchangeably and refers to a device that is able to read a program from a computer memory (*e.g.*, ROM or other computer memory) and perform a set of steps according to the program.

As used herein, the term "hyperlink" refers to a navigational link from one document to another, or from one portion (or component) of a document to another. Typically, a hyperlink is displayed as a highlighted word or phrase that can be selected by clicking on it using a mouse to jump to the associated document or documented portion.

As used herein, the term "display screen" refers to a screen (*e.g.*, a computer monitor) for the visual display of computer generated images. Images are generally displayed by the display screen as a plurality of pixels.

As used herein, the term "computer system" refers to a system comprising a computer processor, computer memory, and a display screen in operable combination. Computer systems may also include computer software.

As used herein, the term "directly feeding" a protein sample from one apparatus to another apparatus refers to the passage of proteins from the first apparatus to the second apparatus without any intervening processing steps. For example, a protein that is directly fed from a protein separating apparatus to a mass spectrometry apparatus does not undergo any intervening digestion steps (*i.e.*, the protein received by the mass spectrometry apparatus is undigested protein).

As used herein, the term "sample" is used in its broadest sense. In one sense it can refer to a cell lysate. In another sense, it is meant to include a specimen or culture obtained from any source, including biological and environmental samples. Biological samples may be obtained from animals (including humans) and encompass fluids, solids, tissues, and gases. Biological samples include blood products (*e.g.*, plasma and serum), saliva, urine, and the like and includes substances from plants and microorganisms. Environmental samples include environmental material such as surface matter, soil, water, and industrial samples. These examples are not to be construed as limiting the sample types applicable to the present invention.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to multi-phase protein separation and analysis methods capable of resolving and identifying large numbers of cellular proteins. In particular, the present invention provides systems and methods for high resolution separation and on-plate digestion coupled with mass spectrometry.

The development of multi-dimensional HPLC separation in proteomics has greatly contributed to simplifying sample purification procedures for the analysis of highly complex biological mixtures. The effluent from the RP-HPLC can be analyzed by on-line ESI-MS, or, alternatively, can be collected for off-line MALDI-based MS analysis. In either case, such mass spectrometric analysis is generally performed using in-solution tryptic digestion and sample purification of the peptide map that results. These extensive off-line sample procedures are a drawback, where enzymatic digestion typically occurs overnight and sample purification with commercially available SPE-based C18 ziptips is tedious and prone to the loss of peptides of extreme hydrophilicity (Li et al., Proteomics 2005, 5, 1460-1471). In addition, loss of peptides due to repetitive sample transfers is inevitable and leads to low sequence coverage for database searching for less confident protein identification. This series of time-consuming and laborious techniques is not suitable for large-scale proteomic study of samples of high complexity where hundreds of unique proteins are to be analyzed.

Thus, the development of rapid and high-throughput techniques to accelerate enzymatic digestion and sample purification for reliable peptide mass fingerprinting (PMF) analysis is needed. Various methods have been proposed to expedite the enzymatic digestion processes, including microwave-assisted digestion (Pramanik et al., Protein Sci 2002, 11, 2676-2687; Lin et al., J Am Soc Mass Spectrom 2005, 16, 581-588; Zhong et al., J Am Soc Mass Spectrom 2005, 16, 471-481) and on-line tryptic digestion (Hedstrom et al., J Chromatogr A 2005, 1080, 117-123; Craft et al., Anal Chem 2005, 77, 2649-2655; Calleri et al., J Proteome Res 2005, 4, 481-490; Slysz et al., Anal Chem 2005, 77, 1572-1579; Calleri et al., J Chromatogr A 2004, 1045, 99-109; Kato et al., Anal Chem 2004, 76, 1896-1902; Samskog et al., J Chromatogr A 2003, 998, 83-91; Slysz et al., Rapid Commun Mass Spectrom 2003, 17, 1044-1050; Peterson et al., J Proteome Res 2002, 1, 563-568; Vecchione et al., Rapid Commun Mass Spectrom 2001, 15, 1383-1390; Marie et al., Anal Chem 2000, 72, 5423-5430). Microwave assisted digestion can accelerate enzymatic digestion into minutes using a very high concentration of acid to assist the hydrolysis of

peptides. An on-line trypsin digestion combined with protein separation was demonstrated (Slysz et al., Anal Chem 2005, 77, 1572-1579), although the application has been limited mainly to standard proteins. Enzymatic digestion on-plate MALDI (Warscheid et al., Proteomics 2004, 4, 2877-2892; Harris et al., Anal Chem 2002, 74, 4410-4416) is one method to significantly reduce the time required for tryptic digestion to obtain high-throughput analysis. The on-plate digestions reported so far, however, were mostly performed on standard protein mixtures or cells from very simple organisms and tandem MALDI MS was often required for protein identification.

Micro-scale HPLC provides rapid separation of proteins; in particular, the monolithic capillary column has drawn much attention in proteomic studies due to its high separation speed, high efficiency, and high recovery (Strancar et al., Adv Biochem Eng Biotechnol 2002, 76, 49-85; Zou et al., J Chromatogr A 2002, 954, 5-32). Moreover, its low flow rate, typically in the range of a few μL/min, is suited to be combined with off-line peak collections for MALDI MS for high throughput analysis. A heated droplet interface (Zhang et al., Anal Chem 2004, 76, 992-1001) and monolithic capillary LC-MALDI (Chen et al., Anal Chem 2005, 77, 2323-2331) have been reported to couple micro-scale HPLC for peptide separations for subsequent tandem MALDI MS analysis.

In experiments conducted during the course of development of the present invention, two dimensional liquid phase separation was coupled with on-plate digestion of proteins for subsequent MALDI MS to identify the proteins in cell lysates from the human breast cancer cell line MCF10A by PMF analysis. It was demonstrated that off-line analysis and on-plate digestion can be achieved for rapid analysis with minimal sample handling. It was shown that high sequence coverage was obtained for close to 40 unique proteins separated by monolith-based HPLC. In addition, the MW values of the proteins identified in this experiment were compared to intact MW values measured using nonporous RP-HPLC separation of the proteins interfaced to ESI-TOF MS as a means to further constrain the MALDI-TOF MS database search.

Experiments conducted during the course of development of the present invention resulted in the development of a method for integrating capillary monolithic RP-HPLC for protein separation with on-probe enzymatic digestion for subsequent MALDI-MS analysis for obtaining protein identifications for human breast cancer cells in a high-throughput manner. The fast protein separation provided by the monolith-based HPLC combined with off-line interfacing with the MALDI-MS is a novel platform for rapid protein identification

with improved sequence coverage. The method is simple and robust, and also effectively minimizes sample loss by avoiding sample transfers and additional sample clean-up procedures. The present work only involved the use of a single stage MS without the need for tandem MALDI-MS analysis due to the use of protein MW to constrain the database search. In other embodiments, tandem MS is used with this approach to further confirm the identifications and for structural analysis. The use of intact protein separation has distinct advantages over total protein digestion of the sample into peptides in that the intact protein method provides protein MW and improved sequence coverage for protein identification.

The methods of the present invention are suitable for the analysis and detection of small amounts of proteins. For example, experiments conducted during the development of the present invention demonstrate that, in some embodiments, 200 ng or less, and preferably 100 ng or less of starting material prior to the monolithic HPLC step is suitable for detection of proteins using a single MALDI-TOF analysis step. In one exemplary embodiment, 120 ng was shown to be suitable for detection of proteins (See e.g., Example 1).

Further experiments conducted during the course of development of the present invention resulted in the development of an automated system for transferring proteins from the HPLC step to the MALDI plate for digestion. In some embodiments, the method utilizes a robotic sample handling module. It is preferred that liquids be delivered through the sample handling device at a flow rate that is sufficient to prevent retention of liquids in tubing used for transport. The use of such systems decrease the analysis time and reduces the chance for contamination.

I. Multi-Phase Separation Techniques

In some embodiments, the present invention provides a multi phase separation method (e.g., a first separation followed by monolithic HPLC). The first dimension separates proteins based on a first physical property. For example, in some embodiments of the present invention proteins are separated by pI using isoelectric focusing in the first dimension (*See e.g.*, Righetti, Laboratory Techniques in Biochemistry and Molecular Biology; Work, T. S.; Burdon, R. H., Elsevier: Amsterdam, p 10 [1983]). However, the first dimension may employ any number of separation techniques including, but not limited to, ion exclusion, ion exchange, normal/reversed phase partition, size exclusion, ligand exchange, liquid/gel phase isoelectric focusing, and adsorption chromatography. In some embodiments (*e.g.*, some automated embodiments), it is preferred that the first dimension be

conducted in the liquid phase to enable products of the separation step to be fed directly into a second liquid phase separation step.

In some embodiments, the second dimension preferably separates proteins using monolithic HPLC and is preferably conducted in the liquid phase. For example, in some embodiments of the present invention proteins are separated by hydrophobicity using monolithic HPLC (*See e.g.*, Strancar et al., Adv Biochem Eng Biotechnol 2002, 76, 49-85; Zou et al., J Chromatogr A 2002, 954, 5-32; Chen et al., Anal Chem 2005, 77, 2323-2331). In other embodiments, proteins are separated in a second dimension using NP-RP-HPLC. In still further embodiments, proteins are separated using both NP-RP-HPLC and monolithic HPLC (e.g., NP-RP-HPLC followed by monolithic HPLC). Having the second and subsequent dimensions conducted in the liquid phase facilitates efficient analysis of the separated proteins and enables products to be fed directly into additional analysis steps (*e.g.*, directly into mass spectrometry analysis).

In some embodiments, the proteins collected from the second or subsequent dimensions are identified using proteolytic enzymes, MALDI-TOF MS and MSFit database searching. Certain preferred embodiments are described in detail below. These illustrative examples are not intended to limit the scope of the invention. For example, although the examples are described using human tissues and samples, the methods and apparatuses of the present invention can be used with any desired protein samples including samples from plants and microorganisms.

Exemplary protein separation and analysis methods suitable for use with the present invention are described in more detail below. One skilled in the relevant arts recognizes that additional methods may be utilized. For example, addition protein separation and analysis methods are described, for example, in U.S. Patent applications 20040010126, 20020039747, 20050230315, 20040033591, 20040214233, 20020098595, 20030064527, and U.S. Patent 6,931,325, each of which are herein incorporated by reference in their entirety.

A. 2D Separation Methods

The following description provides certain preferred embodiments for conducting first dimension (e.g., chromatofocusing) and second dimension (e.g., monolithic HPLC) separation and/or NP-RP-HPLC (second dimension) according to the methods of the present invention.

1. IEF Separation

In some embodiments, the first dimension separation is isoelectric focusing. Proteins are extracted from cells using a lysis buffer. To facilitate an efficient process, this lysis buffer should be compatible with the downstream separation and analysis steps (*e.g.*, HPLC and MALDI-TOF-MS) to allow direct use of the products from each step into subsequent steps. Such a buffer is an important aspect of automating the process. Thus, the preferred buffer should meet two criteria: 1) it solubilizes proteins and 2) it is compatible with each of the steps in the separation/analysis methods. Although the present invention provides suitable buffers for use in the particular method configurations described below, one skilled in the art can determine the suitability of a buffer for any particular configuration by solubilizing protein sample in the buffer. If the buffer solubilizes the protein, the sample is run through the particular configuration of separation and detection methods desired. A positive result is achieved if the final step of the desired configuration produces detectable information (*e.g.*, ions are detected in a mass spectrometry analysis). Alternately, the product of each step in the method can be analyzed to determine the presence of the desired product (*e.g.*, determining whether protein elutes from the separation steps).

After extraction in the lysis buffer, proteins are initially separated in a first dimension. It is preferred that the proteins are isolated in a liquid fraction that is compatible with subsequent HPLC and mass spectrometry steps. In some embodiments, n-octyl ß-D-glucopyranoside (OGl, from Sigma) is used in the buffer. It is contemplated that detergents of the formula n-octyl SUGARpyranoside find use in these embodiments. The lysis buffer utilized was 6M urea, 2M thiourea, 1.0 % n-octyl ß-D-glucopyranoside, 10 mM dithioerythritol and 2.5 % (w/v) carrier ampholytes (3.5 to 10 pI)). After extraction, the supernatant protein solution is loaded to a device that can separate the proteins according to their pI by isoelectric focusing (IEF). In some embodiments, the proteins are solubilized in a running buffer that is compatible with HPLC.

Three exemplary devices that may be used for this step are:

a) Rotofor

This device (Biorad) separates proteins in the liquid phase according to their pI (*See e.g.*, Ayala *et al.*, Appl. Biochem. Biotech. 69:11 [1998]). This device allows for high protein loading and rapid separations that require only four to six hours to perform. Proteins

are harvested into liquid fractions after a 5-hour IEF separation. These liquid fractions are ready for analysis by HPLC. This device can be loaded with up to 1 g of protein.

b) Carrier Ampholyte based slab gel IEF separation with a whole gel eluter

In this case the protein solution is loaded onto a slab gel and the proteins separate in to a series of gel-wide bands containing proteins of the same pI. These proteins are then harvested using a whole gel eluter (WGE, from Biorad). Proteins are then isolated in liquid fractions that are ready for analysis by HPLC. This type of gel can be loaded with up to 20 mg of protein.

c) IPG slab gel IEF separation with a whole gel eluter

Here the proteins are loaded onto a immobiline pI gradient slab gel and separated into a series of gel-wide bands containing proteins of the same pI. These proteins are electro-eluted using the WGE into liquid fractions that are ready for analysis by NP RP HPLC. The IPG gel can be loaded with at least 60 mg of protein.

2. Chromatofocusing

In other embodiments, the first dimension separation is chromatofocusing. In chromatofocusing proteins are eluted from the column according to their pH, either one pH unit or fraction thereof, at a time. Columns for chromatofocusing are commercially available (e.g., Mono P HR 5/20 (Amersham Pharmacia, Uppsala, Sweden)). The column is equilibrated with a first buffer to define the upper pH range of the pH gradient. The proteins are then applied. The second focusing buffer is then applied to elute bound proteins, in the order of their isoelectric (pI) points. The pH of the second buffer is lower, and, defines the lower limit of the pH gradient. The pH gradient is formed as the eluting buffer titrates the buffering groups on the ion-exchanger.

3. Protein Separation by Monolithic HPLC

In some preferred embodiments, proteins are separated using monolithic HPLC chromatography (See e.g., U.S. Patents 6,976,384, 6,956,207, and 6,884,345, each of which is herein incorporated by reference in its entirety). Generally, porous monoliths may be

fabricated by flowing a monomer solution into a channel or conduit, and then activating the monomer solution to initiate polymerization. Various formulations and various activation means may be used. The ratio of monomer to solvent in each formulation may be altered to control the degree of porosity of the resulting monolith. A photoinitiator may be added to a monomer solution to permit activation by means of a lamp or other radiation source. If a lamp or other radiation source is used as the initiator, then photomasks may be employed to localize the formation of monoliths to specific areas within a fluidic separation device, particularly if one or more regions of the device body are substantially optically transmissive. Alternatively, chemical initiation or other initiation means may be used. Numerous recipes for preparing monolithic columns suitable for performing chromatographic techniques are known in the art.

Fréchet in U.S. Pat. Nos. 5,334,310 and 5,453,185, which are each herein incorporated by reference in their entirety, describe the use of a continuous polymer bed formed by in situ polymerization of a monomer solution containing a porogen within a column. Many examples on the use of these continuous or monolithic polymer supports are available in the literature. Liao in U.S. Pat. No. 5,647,979 (herein incorporated by reference) describes a similar use of a continuous polymer bed for reversed-phase chromatography and capillary electrochromatography in capillary columns.

Poly(styrene-co-divinylbenzene) monolithic columns have been successfully used for the fast separation of proteins in the reversed phase mode (Wang et al., J. Anal. Chem. 1993, 65, 2243-2248), and it has recently been shown that polyacrylamide based monoliths can be used for the rapid separation of proteins in the hydrodynamic interaction mode when butyl methacrylate is included in tie polymerization mold (Xie et al., J. Chromatogr. A., 1997, 775, 65-72).

In other embodiments, additional capillary chromatography techniques (e.g., non-porous capillary chromatography) are utilized.

4. Protein Separation by NP-RP-HPLC

In other embodiments, the second dimension separation is non-porous RP HPLC. In some embodiments, NP-RP-HPLC is the only HPLC method utilized. In other embodiments, NP-RP-HPLC is followed by monolithic HPLC. The present invention provides the novel combination of employing non-porous RP packing materials (Eichrom) with another RP HPLC compatible detergent (e.g., n-octyl ß-D-galactopyranoside) to

facilitate the multi-phase separation of the present invention. This detergent is also compatible with mass spectrometry due to its low molecular weight. The use of these types of RP HPLC columns for protein separations as a second dimension separation after IEF in order to obtain a 2-D protein separation is a novel feature of the present invention. These columns are well suited to this task as the non-porous packing they contain provides optimal protein recovery and rapid efficient separations. It should be noted that though several detergents have been mentioned thus far for increasing protein solubility while being compatible with RP HPLC there are many other different low molecular weight non-ionic detergents that could be used for this purpose. Several important features that allow the RP HPLC to work as a second dimension are as follows: The mobile phase should contain a low level of a non-ionic low molecular weight detergent such as n-octyl ß-D-glucopyranoside or n-octyl ß-D-galactopyranoside as these detergents are compatible with RP HPLC and also with later mass spectrometry analyses (unlike many other detergents); the column should be held at a high temperature (around 60°C); and the column should be packed with non-porous silica beads to eliminate problems of protein recovery associated with porous packings.

B. Protein Detection and Identification via Mass Spectrometry

In some embodiments of the present invention, the products of the second separation step are further characterized using mass spectrometry. For example, in some embodiments, the proteins that elute from the HPLC separation are analyzed by mass spectrometry to determine their molecular weight and identity. The present invention is not limited by the nature of the mass spectrometry technique utilized for such analysis. For example, techniques that find use with the present invention include, but are not limited to, ion trap mass spectrometry, ion trap/time-of-flight mass spectrometry, time of flight/time of flight mass spectrometry, quadrupole and triple quadrupole mass spectrometry, Fourier Transform (ICR) mass spectrometry, and magnetic sector mass spectrometry. The following description of mass spectroscopic analysis and 2-D protein display is illustrated with ESI oa TOF mass spectrometry. Those skilled in the art will appreciate the applicability of other mass spectroscopic techniques to such methods.

For this purpose the proteins eluting from the separation can be analyzed simultaneously to determine molecular weight and identity. A fraction of the effluent is used to determine molecular weight by either MALDI-TOF-MS or ESI oa TOF (LCT,

Micromass) (*See e.g.*, U.S. Pat. No. 6,002,127). The remainder of the eluent is used to determine the identity of the proteins via digestion of the proteins and analysis of the peptide mass map fingerprints by either MALDI-TOF-MS or ESI oa TOF. In preferred embodiments, the on-plate digestion technique described above and in Example 1 is utilized. The molecular weight 2-D protein map is matched to the appropriate digest fingerprint by correlating the molecular weight total ion chromatograms (TICs) with the UV-chromatograms and by calculation of the various delay times involved. The UV-chromatograms are automatically labeled with the digest fingerprint fraction number. The resulting molecular weight and digest mass fingerprint data can then be used to search for the protein identity via web-based programs like MSFit (UCSF).

In some embodiments, multiple mass spectrometry (e.g., 2, 3, or more) steps are utilized in the analysis of separated protein fractions. For example, in some embodiments, MALDI-MS/MS is utilized. In other embodiments, MS-MS is utilized.

II. Microscale Analysis

Biological samples often comprise complex mixtures of proteins whose concentrations vary by up to 5 orders of magnitude in a given cell type (and by at least 10 orders of magnitude in serum). Multidimensional liquid-based fractionation strategies allow for distinct protein-containing fractions to be interrogated by a variety of techniques and, importantly, to increase protein load thereby facilitating identification of the lower abundance proteins. Usually the protein separation in the first dimension is based on protein pI, using either chromatofocusing, cation exchange or anion exchange chromatography to resolve the complex mixtures of proteins. For separation in the second dimension, reverse-phase HPLC is typically employed, in which protein separation is based upon protein hydrophobicity.

The amount of protein useful to achieve both high recovery and detection of the lower abundance proteins during a typical multidimensional separation is approximately 5 mg. In some embodiments, this amount of protein is not available for separation.

Experiments conducted during the course of development of embodiments of the present invention resulted in the development of techniques for the first dimension separation of samples with low protein concentrations using Micro-Chromatofocusing (Micro-CF), a weak anion exchange separation technique. Certain embodiments are described below. In some embodiments, the capillary tube column for Micro-CF is

composed of fused silica coated with polyimide (Polymicro Technologies Inc., Phoenix, AZ). The tube dimensions utilized were either 200 μm I.D. x 365μm OD (150 mm length) or 300 μm ID x 437 μm OD (150mm length). The packing material (6-7μm, 300Å silica (Eprogen, Darien, IL)) was homogenized by sonication, and then placed in a pressure vessel. The capillary column was packed by slowly increasing the pressure. After packing, the column was washed with isopropanol. Micro-CF was performed by first equilibrating the column with starting buffer (6M urea, 0.025M Bis-Tris, 0.2% octylglucoside, pH 7.4), and then injecting 5-40 micrograms of the protein sample onto the column. A pH gradient was generated by switching to an elution buffer (6M Urea, 2% poly74 buffer, 0.2% octylglucoside, pH 4.0), using the Ultra-Plus II MD capillary pump (Micro-Tech Scientific, Vista, CA) with the pH being monitored by a flow-thru pH microelectrode (Microelectrodes, Bedford, NH). UV absorbance was monitored at 280nm, using the UV Detector 166 (Beckman-Coulter, Fullerton, CA) equipped with a micro-flow cell. The 5 μl/min flow rate for the Micro-CF was obtained by using a column pre-splitter. The individual fractions were collected at 0.2-0.4 pH unit intervals.

The fractions obtained from Micro-CF were introduced into a monolith column for a reverse phase second dimension separation. The monolith capillary column (360 μm OD x 200μm ID x 60mm L) was prepared using copolymerized styrene and divinylbenzene. The column was heated at 60°C with a column heater. The protein was eluted from the column using a low flow rate (2-3μl/min). The second dimension separation was performed with a two solvent (A: 0.05% Formic acid in deionized water, B: 0.05% Formic acid in acetonitrile) system.

Following the second separation, the eluate from the monolith column can either be spotted directly onto a plate for MALDI TOF-MS or collected for further MS/MS analysis. In order to identify proteins with Mass Spectrometry, protein digestion with either trypsin, or a different proteolytic enzyme, is utilized. In some embodiments, on-plate digestion methods are utilized to mitigate the laborious, time consuming procedure and, in addition, avoids peptide loss, thus facilitating increased protein sequence coverage (See above section on analysis methods).

Combining the three improved methodologies (the two column separations followed by mass spectrometry with on-plate digestion) greatly facilitates the processing of small amounts of sample without losing many of the advantages of regular multidimensional separation techniques.

III. Automation

In some embodiments, all of the above described steps are automated, for example, into one discrete instrument. In one illustrative embodiment, the first dimension is carried out by a Rotofor, with the harvested liquid fractions being directly applied to the second dimension monolithic HPLC apparatus through the appropriate tubing. The products from the second dimension separation are then scanned and the data interpreted and displayed as a 2-D representation using the appropriate computer hardware and software. Alternately, the products from the second dimension fractions are sent through the appropriate microtubing to an on-plate MALDI digestion step, followed by mass spectrometry. See Example 1 below for a description of one exemplary automation device. The resulting data is received and interpreted by a processor. The output data represents any number of desired analyses including, but not limited to, identity of the proteins, mass of the proteins, mass of peptides from protein digests, dimensional displays of the proteins based on any of the detected physical criteria (*e.g.*, size, charge, hydrophobicity, etc.), and the like. In preferred embodiments, the proteins samples are solubilized in a buffer that is compatible with each of the separation and analysis units of the apparatus. Using the automated systems of the present invention provides a protein analysis system that is an order of magnitude less expensive than analogous automation technology for use with 2-D gels (*See e.g.*, Figeys and Aebersold, J. Biomech. Eng. 121:7 [1999]; Yates, J. Mass Spectrom., 33:1 [1998]; and Pinto *et al.*, Electrophoresis 21:181 [2000]).

IV. Software and Data Presentation

The data generated by the above listed techniques may be presented as 2-D images much like the traditional 2-D gel image. In some embodiments, the chromatograms, TICs or integrated and deconvoluted mass spectra are converted to ASCII format and then plotted vertically, using a 256 step gray scale, such that peaks are represented as darkened bands against a white background. In other embodiments, the scale is in a color format. The image generated by this method provides information regarding the pI, hydrophobicity, molecular weight and relative abundance of the proteins separated. Thus the image represents a protein pattern that can be used to locate interesting changes in cellular protein profiles in terms of pI, hydrophobicity, molecular weight and relative abundance. Naturally the image can be adjusted to show a more detailed zoom of a particular region or the more

abundant protein signals can be allowed to saturate thereby showing a clearer image of the less abundant proteins. This information can be used to assess the impact of disease state, pharmaceutical treatment, and environmental conditions. As the image is automatically digitized it may be readily stored and used to analyze the protein profile of the cells in question. Protein bands on the image can be hyper-linked to other experimental results, obtained via analysis of that band, such as peptide mass fingerprints and MSFit search results. Thus all information obtained about a given 2-D image, including detailed mass spectra, data analyses, and complementary experiments (*e.g.*, immuno-affinity and peptide sequencing) can be accessed from the original image.

The data generated by the above-listed techniques may also be presented as a simple read-out. For example, when two or more samples are compared, the data presented may detail the difference or similarities between the samples (*e.g.*, listing only the proteins that differ in identity or abundance between the samples). In this regard, when the differences between samples (*e.g.*, a control sample and an experimental sample) are indicative of a given condition (*e.g.*, cancer cell, toxin exposure, etc.), the read-out may simply indicate the presence or identity of the condition. In one embodiment, the read-out is a simple +/- indication of the presence of particular proteins or expression patterns associated with a specific condition that is to be analyzed.

EXPERIMENTAL

The following examples serve to illustrate certain preferred embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof.

EXAMPLE 1

This Example describes the monolithic HPLC - on plate digestion - MS method of one embodiment of the present invention. The experimental overview is shown in Figure 1. Proteins from MCF10A were first separated using the Rotofor device for pI-based fractionation, where one of the fractions was selected for further separation by monolithic capillary HPLC. The fractions at 30 sec intervals were collected off-line on the MALDI plate pre-coated with trypsin for on-plate digestion and subsequent analysis. The same pH fraction was also analyzed by on-line ESI-TOF MS with nonporous (NPS) C18 packed column to obtain intact protein MW values for comparison to theoretical MW values.

A. METHODS

Materials and Reagents

Ammonium bicarbonate, trifluoroacetic acid (TFA), formic acid, α-cyano-4-hydroxycinnamic acid (α-CHCA), and acetonitrile were purchased from Sigma (St. Louis, MO). A TPCK-modified trypsin (porcine) of sequencing grade was purchased from Promega (Madison, WI). A protein assay kit and bovine serum albumin standard were obtained from Bio-Rad Laboratories (Hercules, CA). DI water was purified by the Milli-Q water filtration system (Millipore, Inc., Bedford, MA). All the reagents were used without further purification.

Sample Preparation and pI-based Separation

The cells used in this work include MCF10A, which is maintained and prepared by the Barbara Ann Karmanos Cancer Institute (Wayne State University, Detroit, MI), as previously described (Santer et al., Breast Cancer Res. Treat. 2001, 65, 101-110). The preparation of cell extracts for liquid-phase IEF and its fractionation using the Mini-Rotofor (Bio-Rad) as well as the protein quantitation based on the Bradford method and pH measurements were performed as described elsewhere (Hamler et al., Proteomics 2004, 4, 562-577).

On-line NPS-RP-HPLC/ESI-TOF MS for Intact MW Analysis

Intact MW analysis was performed as described previously (Hamler et al., supra). Briefly, the NPS-RP HPLC column (33 mm L × 4.6 mm ID) packed with 1.5 μm C18 nonporous ODSIIIE silica beads (Eprogen, Darien, IL) was used for the separation of the MCF10 cell line prefractionated at pH 6.34 at a flow rate of 0.5 mL/min. The HPLC System Gold with UV detector set at 214 nm (Beckman Coulter, Fullerton, CA) was used with the solvents A and B composed of 0.3% formic acid in DI water and acetonitrile, respectively. The column was maintained at 65°C (model 7971 column heater, Jones Chromatography, Resolution Systems, Holland, MI) and the following gradient was used: 5-15% B in 1 min, 15-25% B in 2 min, 25-31% B in 3 min, 31-41% B in 10 min, 41-47% B in 3 min, 47-67% B in 4 min, 67-100% B in 1 min, 100% B for 2 min, and 100-5% B in 1 min. A splitter system was used so that 40% of the eluent from HPLC was delivered on-line to ESITOF MS (LCT, Waters-Micromass, Manchester, U.K.). The desolvation temperature

was set at 300°C and the source temperature at 120°C, where nitrogen gas flow was controlled at 650 L/hr. One mass spectrum was acquired per second and the deconvolution of the combined spectra of the protein was performed by utilizing the MaxEnt1 of MassLynx software version 4.0 (Waters-Micromass).

Monolithic Capillary HPLC Separation and On-MALDI Plate Enzymatic Digestion

The preparation of monolithic capillary columns (360 μm OD × 200 μm ID × 60 mm L) by co-polymerizing styrene and divinylbenzene (PS/DVB) was performed according to procedures described elsewhere (Huber et al., Proteomics 2004, 4, 3909-3920). The Ultra-Plus II MD Capillary Pump module (Micro-Tech Scientific, Vista, CA) was used for the chromatographic separation utilizing a monolithic column. The capillary column was directly mounted to a micro-injector with a 500 nL internal sample loop (Valco Instruments, Houston, TX) by a microtight union (Upchurch Scientific, Oak Harbor, WA). This was connected with fused silica capillary tubing of 20 μm ID with a burnt detection window to the ProteomeLab PA800 Capillary Electrophoresis system (Beckman Coulter, Fullerton, CA) for UV detection. The capillary protein separation was controlled at 60°C with a column heater utilizing a variable autotransformer (Staco Energy Product, Dayton, OH). The flow from the solvent delivery pump was split pre-column in order to produce a flow rate of approximately 2.5 μL/min through a monolithic capillary column. A mobile phase system of two solvents was used, wherein solvents A and B were composed of 0.05% formic acid in DI water and acetonitrile, respectively. A linear gradient of 0 to 100% B in 18 min was applied.

Approximately 120 ng of pre-fractionated MCF10A cell line at pH 6.34 was loaded onto a monolithic capillary column for separation, which was monitored at 214 nm to acquire chromatograms using 32 Karat software version 7.0 (Beckman Coulter). The proteins eluting off the monolithic capillary column were directly deposited onto a 96-well MALDI plate at 30 sec intervals for each spot. The MALDI plate was pre-coated with 0.5 μL of trypsin stock solution of 0.15 μg/μL. Following the sample collection, 0.5 μL of 50 mM ammonium bicarbonate was added to the top layer of each spot and the plate was kept in a humidifier chamber for 5 min at room temperature for digestion. Then, 0.5 μL of 0.1% TFA was added to each spot to stop the digestion, followed by adding 0.5 μL of α-CHCA matrix solution prepared by diluting saturated α-CHCA with 60% acetonitrile/0.1% TFA at a 1:4 ratio. The internal standards, including angiotensin I,

adrenocorticotropic hormone (ACTH) fragment 1-17, and ACTH fragment 18-39, were added so that a final concentration of 50 fmol of each standard was placed in every spot of the MALDI plate.

Automation of Monolithic HPLC - MALDI

The GeSiM Nano-Plotter 2.0 piezoelectric pipetting system was used for automating monolithic HPLC for direct MALDI spotting. It uses micro-machined proprietary piezoelectric tips for dispensing liquids in nano-liter volumes and is a flexible modular instrument used primarily for fabricating protein/DA microarrays on various array surfaces. The instrument is built around a standard robotic platform that is easy to program using the nano-plotter language (NPL) editor and interpreter provided with the Nano-Plotter NP2x PC software for instrument control and hence can be used as a general purpose robot for compatible high precision tasks.

For the current experimental platform, the robotic xyz module or the print head of the instrument was modified by replacing the piezoelectric dispensers with a 50 μm ID fused-silica capillary housing. The hydraulic module of the instrument was not needed since the capillary was attached directly to connecting tubings from the monolithic LC column for protein separation and from the syringe pump used for delivering diluted trypsin solutions, 50 mM ammonium bicarbonate, 0.1% trifluoroacetic acid, and MALDI matrix mixture.

An instrument controller program written in NPL was used to control the print head for precise movements in xyz axes for spotting onto multiple MALDI target plates with control over the previously designated fraction collection/depositions time. Fraction collection was performed in a real-time mode with proper calibration for instrument latency.

MALDI-TOF MS Analysis and Database Searching

The MALDI-TOF MS analysis was performed on a TofSpec2E (Waters-Micromass) equipped with delayed extraction in reflectron mode with positive polarity. A 337 nm Nd:YAG laser was used as the ionization source, where the coarse laser energy was set at 50% and the fine laser energy varied from 20 to 90% at the laser frequency of 5 Hz. The delay time was set at 520 ns, the source voltage at 20 kV, the extraction voltage at 1:1 to the source voltage, the pulse voltage at 2300 V, and the reflectron voltage at 24.5 kV, wherein 15-20 spectra were collected over the m/z range of up to 4000 Da. Each spectrum was

internally calibrated and monoisotopic peptide masses were obtained using MassLynx software version 4.0 (Waters-Micromass) for submission to the MS-Fit search engine using SwissProt database for protein identification. The search was carried out under the species of Homo Sapiens at 50 ppm mass tolerance with no limitation set for the molecular weight and pI. The possible modifications included N-terminal Gln to pyroGlu, oxidation of Met, N-terminal acetylation, and phosphorylation at S, T, and Y.

B. RESULTS AND DISCUSSION

Protein Separation by Monolithic Capillary Column

Proteins pre-fractionated based on pI were separated with a PS/DVB monolithic capillary column using a low flow rate of 2.5 μL/min, which is suitable for directly depositing proteins on the MALDI plate. Only about 120 ng of total proteins were consumed for the entire experiment. The actual separation was completed in approximately 18 min to separate one pH fraction. Compared to the typical 40 min separation time used in packed column capillary HPLC separation, the separation time was reduced significantly by the use of monolith-based micro-scale HPLC separation. Although the 200 μm ID monolithic columns may not be sufficient for high resolution separation of highly complex protein mixtures in human cancer cells, this problem becomes less significant with the detection of protein digests by MALDI-TOF MS, on-plate digestion, and the use of intact MW analysis.

Identification of Proteins by PMF Analysis

Proteins separated by monolithic capillary HPLC were deposited on the trypsinized MALDI-plate and analyzed for the peptide map by MALDI-TOF MS. The proteins identified in this experiment are summarized in Table 1 (Figure 3). All database search results were subjected to manual inspection to consider the following criteria to obtain a confident match: protein hits with the sequence coverage of greater than 20% and the MOWSE score of greater than 103. A total of 37 unique proteins from approximately 120 ng of human breast cancer cell lysates pre-fractionated at pH 6.34 were successfully identified using the monolithic RP-HPLC separation time of 18 min. There are a number of distinct advantages noted from this experiment. Although the pH fractionation used in the work required milligrams of sample, the sample loaded on the monolithic capillary column

was reduced by 50-fold, thus, other methods using microprefractionation can be interfaced to this technique for analysis of small amounts of sample. Also the time of analysis for analyzing large numbers of samples is an issue where typical in-solution digestion proceeds for hours to overnight, while on-probe digestion, providing more surface area for enzymatic reactions, completes in 5 min. In addition, the collection of the proteins, enzymatic digestion, and MS analysis are all integrated into one MALDI plate without any sample transfers in order to avoid unnecessary artificial contamination or sample loss. The high recovery provided by a monolithic column and the ability to minimize sample losses contribute to the high sequence coverage obtained by MALDI-TOF MS (Table 1). Also, several injections with DI water prior to the gradient elution wash away impurities contained in the sample, eliminating the need for laborious and costly sample clean-up procedures typically required for MALDI-MS analysis.

The protein identification in this experiment was constrained by the use of the intact MW value obtained by NPS-RP-HPLC/ESI-TOF MS when compared to the theoretical MW of each protein, also shown in Table 1. This table also indicates that slight differences between experimental and theoretical MW values were observed for several proteins. However, these are mitochondrial precursors that lose transit peptides and are truncated (Neupert, Annu. Rev. Biochem. 1997, 66, 863-917). Given this sequence modification, the experimental and theoretical MW values are closely matched based on the identifications obtained by PMF analysis.

In Table 1, one can observe that unique proteins are usually in each fraction. One can also observe that two or more proteins are often identified from the same spot of the MALDI plate due to the possible co-elution of proteins during the monolithic HPLC separation. For example, in the pH range of 6.34, the separation and PMF analysis indicated that there are three unique proteins identified from each of three MALDI spots corresponding to 5.5 to 6.0 min, 6.5 to 7.0 min, and 20.5 to 21.0 min of protein collection time. Figure 2 shows the presence of the two peptides collected at the interval of 5.5 to 6.0 min, while these were not detected from any of the adjacent spots, implying that proteins were separated with sufficient efficiency to avoid significant overlapping. The database searching unambiguously identified the two peptides originating from 60S ribosomal protein L5. When three consecutive separation intervals were compared, the presence of one of its tryptic peptides, IEGDMIVCAR (69-78; SEQ ID NO:1), was only detected from one spot, but not from any of the neighboring spots, indicating that the protein separation was

sufficient for protein identification by this method.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in the art are intended to be within the scope of the following claims.

CLAIMS

We Claim:

1. A system, comprising

a) a reverse phase monolithic capillary HPLC apparatus; and

b) a MALDI mass spectrometry apparatus comprising a MALDI plate configured for on-plate digestion of proteins separated by said reverse phase monolithic capillary HPLC apparatus.

2. The system of claim 1, further comprising an automated sample handling apparatus that transfers protein samples from said reverse phase monolithic capillary HPLC apparatus to said MALDI mass spectrometry apparatus.

3. The system of claim 2, wherein said automated sample handling apparatus is a robot.

4. The system of claim 1, wherein said MALDI mass spectrometry apparatus is configured to analyze intact proteins.

5. The system of claim 1, wherein said MALDI mass spectrometry apparatus is configured to determine the identity of said proteins separated by said reverse phase monolithic capillary HPLC apparatus.

6. The system of claim 1, wherein said system further comprises software for analyzing data generated by said MALDI mass spectrometry apparatus.

7. The system of claim 1, wherein said MALDI mass spectrometry apparatus comprises a MALDI-TOF mass spectrometer.

8. The system of claim 1, wherein said system further comprising a non-porous reverse phase HPLC apparatus.

9. The system of claim 1, wherein said system further comprising a micro-chromatofocusing apparatus.

10. The system of claim 9, wherein said micro-chromatofocusing apparatus is configured for the separation of approximately 5-40 micrograms of protein.

11. A method, comprising

a) treating a protein sample with a reverse phase monolithic capillary HPLC apparatus under conditions such that said reverse phase monolithic capillary HPLC apparatus separates said protein sample into a plurality of protein fractions;

b) digesting at least a portion of said protein fractions on a MALDI plate to generate digested protein fractions; and

c) analyzing said digested protein fractions with a MALDI mass spectrometer.

12. The method of claim 11, wherein said MALDI mass spectrometer is a MALDI-TOF mass spectrometer.

13. The method of claim 11, wherein said method is automated.

14. The method of claim 13, wherein said protein fractions are transferred from said reverse phase monolithic capillary HPLC apparatus to said MALDI plate using an automated sample handling apparatus.

15. The method of claim 14, wherein said automated sample handling apparatus is a robot.

16. The method of claim 11, wherein said protein fractions comprise intact proteins.

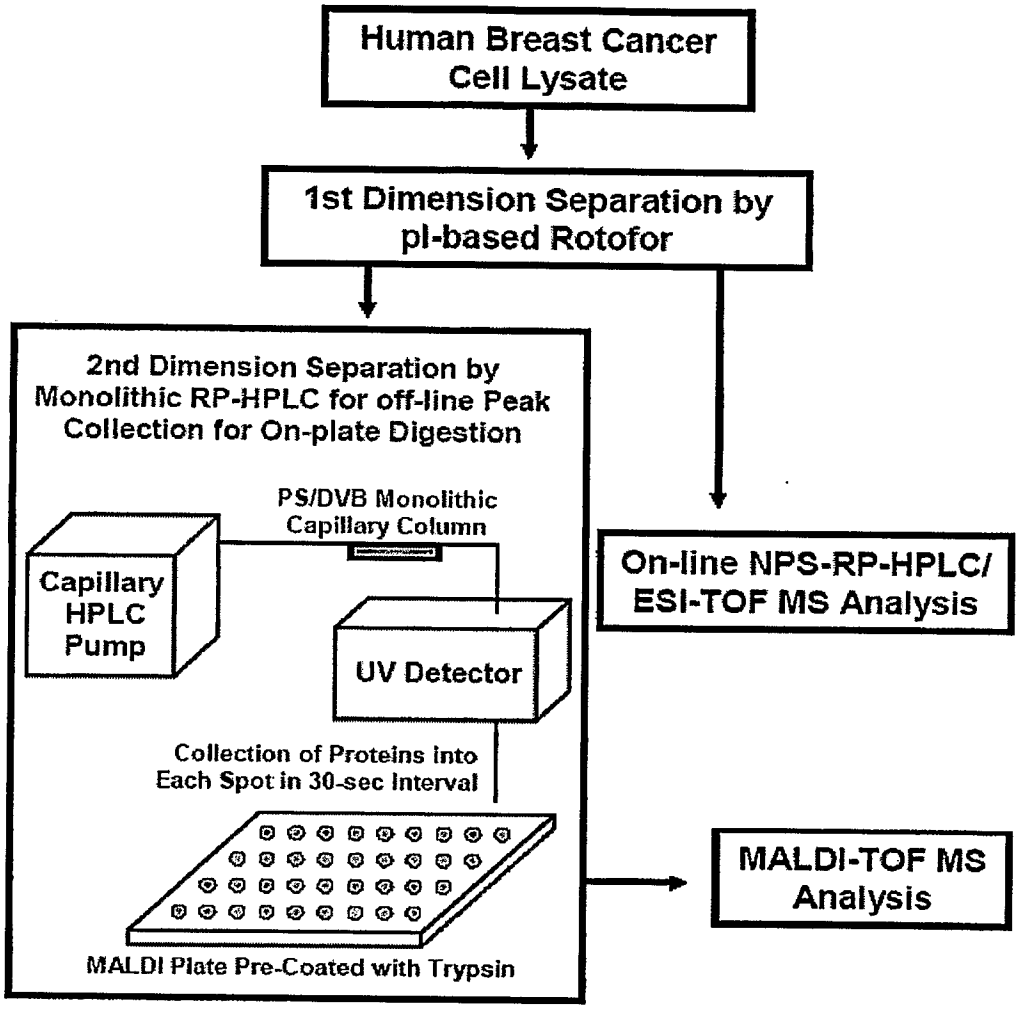
17. The method of claim 11, wherein said analyzing step further comprises determining the identity of at least 5 proteins in said digested protein fractions.

18. The method of claim 11, wherein said separated protein sample comprises less than 200 ng of total protein.

19. The method of claim 11, wherein prior to said treating step, said protein sample is separated with a micro-chromatofocusing apparatus.

20. The method of claim 19, wherein said micro-chromatofocusing apparatus is configured for the separation of approximately 5-40 micrograms of protein.

Figure 1.


Human Breast Cancer Cell Lysate
1st Dimension Separation by pI-based Rotofor
2nd Dimension Separation by Monolithic RP-HPLC for off-line Peak Collection for On-plate Digestion
PS/DVB Monolithic Capillary Column
Capillary HPLC Pump
UV Detector
Collection of Proteins into Each Spot in 30-sec Interval
MALDI Plate Pre-Coated with Trypsin
On-line NPS-RP-HPLC/ ESI-TOF MS Analysis
MALDI-TOF MS Analysis

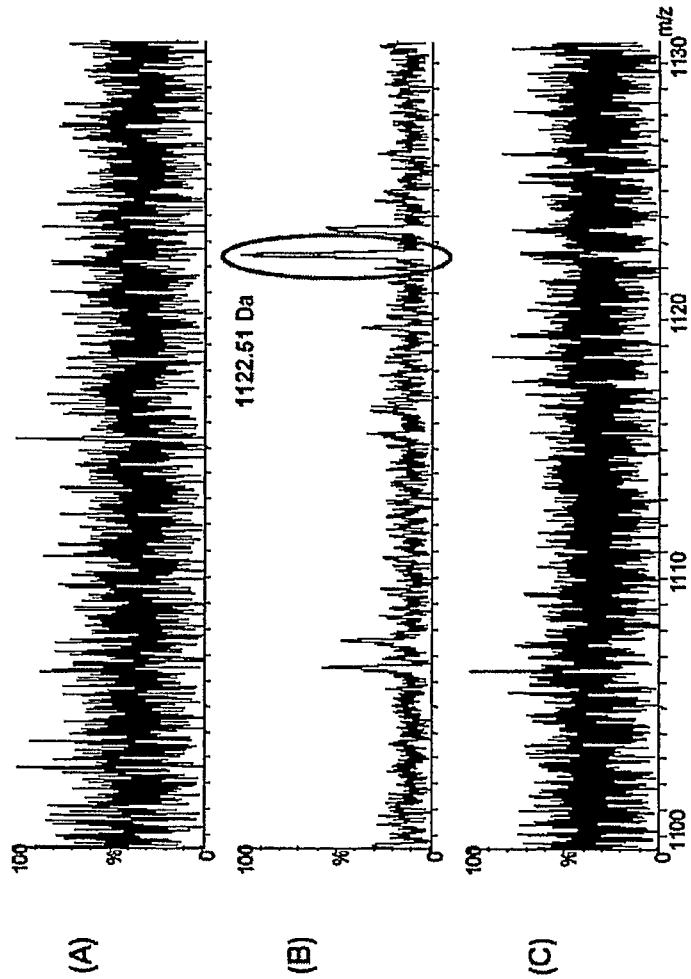


Figure 2.

Figure 3

Table 1. Identification of proteins in MCF10A pre-fractionated at pH 6.34 for monolithic RP-HPLC separation followed by on-probe digestion for MALDI-TOF MS analysis

Collection time (min)	Protein Name	Accession No.	Theoretical		Experimental		Sequence cov. (%)	MOWSE Score
			MW	pI	MW	pI		
5.5-6.0	60S ribosomal protein L5	P46777	34448	9.8	34114	6.34	35	2184
	Small nuclear ribonucleoprotein associated protein B and B'	P14678	24610	11.2	24583	6.34	34	5355
	Glucosamine-fructose-6-phosphate aminotransferase	Q06210	76807	6.7	78816	6.34	43	3712
6.5-7.0	NADH ubiquinone oxidoreductase subunit B14.5b	O95298	14188	9.0	14324	6.34	33	1045
	LIM domain transcription factor LMO4	P61969	17994	8.7	17946	6.34	57	2675
	Ig heavy chain V-III region TEI	P01777	12803	6.9	12808	6.34	57	3625
10.5-11.0	Tyrosine-protein kinase ZAP-70	P43403	69873	7.8	69605	6.34	56	1706
	Alpha-1-antichymotrypsin precursor	P01011	47561	5.3	47573	6.34	49	4122
11.5-12.0	ATP Synthase B chain (truncated)	P24539	28909	9.4	24893	6.34	22	5253
	Succinyl-CoA:3-ketoacid-coenzyme A transferase 1 (truncated)	P55809	56158	7.1	52301	6.34	27	5078
13.0-13.5	Protein disulfide isomerase A3 precursor	P30101	56783	6.0	56664	6.34	60	1.59E+04
13.5-14.0	Lamin A/C	P02545	74140	6.6	74501	6.34	59	5.27E+07

16-16.5	T-complex protein 1, zeta subunit	P40227	58025	6.2	58023	6.34	16	1650
18.0-18.5	Keratin, type II Cytoskeletal 7 (CK 7)	P08729	51418	5.5	51327	6.34	61	2.85E+07
	Ribosomal protein S6 kinase alpha 3	P51812	83737	6.4	83618	6.34	60	2.07E+04
18.5-19.0	Stress-70 protein (truncated)	P38646	73681	5.9	68763	6.34	48	1.94E+04
19.0-19.5	Keratin, type I Cytoskeletal 17 (CK 17)	Q04695	48106	5.0	48830	6.34	47	1.30E+04
	Transitional endoplasmic reticulum ATPase	P55072	89323	5.1	89339	6.34	34	2.71E+04
19.5-20.0	Keratin, type II Cytoskeletal 6D (CK 6D)	P48667	42469	5.3	42625	6.34	53	4.00E+05
20.5-21.0	Keratin, type I Cytoskeletal 19 (CK 19)	P08727	44106	5.0	44012	6.34	76	2.38E+09
	Keratin, type II Cytoskeletal 5 (CK 5)	P13647	62248	8.1	63482	6.34	40	2.78E+05
	Mitochondrial intermediate peptidase (truncated)	Q99797	80612	6.7	77016	6.34	53	1.24E+04
22.5-23.0	L-lactate dehydrogenase A chain	P00338	36689	8.4	36874	6.34	64	1531
	Nucleobindin 1 precursor	Q02818	53852	5.1	53669	6.34	49	1849
23.0-23.5	Tumor-associated calcium signal transducer 1 precursor	P16422	34921	7.4	35030	6.34	34	9381
	Keratin, type II Cytoskeletal 6E (CK 6E)	P16870	60224	8.1	60655	6.34	49	1.04E+06

23.5-24.0	Ribosome biogenesis protein Nop10	Q9NPE3	7706	10.0	7787	6.34	70	4138
	Thymus-specific serine protease precursor	Q9NQE7	55049	8.3	54989	6.34	25	4973
24.5-25.0	TGF-beta receptor type II precursor	P37173	64540	5.6	64508	6.34	46	9147
25.0-25.5	Protein kinase C, theta type	Q04759	81866	7.7	82023	6.34	37	4384
	α-actin, cytoplasmic	P68032	42019	5.2	42055	6.34	27	2369
	γ-actin, cytoplasmic	P63260	41793	5.3	41720	6.34	38	8699
	Retinol-binding protein I	P09455	15850	5.0	15686	6.34	53	3654
25.5-26.0	β-actin, cytoplasmic	P60710	41737	5.3	41712	6.34	56	1.62E+04
26.5-27.0	60 kDa Heat Shock Protein (truncated)	P10809	61055	5.7	57973	6.34	63	4.00E+05
28.5-29.0	Pyruvate dehydrogenase protein X component (truncated)	O00330	54123	8.8	47818	6.34	47	4167
29.5-30.0	T-complex protein 1, theta subunit	P50990	59621	5.4	59256	6.34	59	4845