Available online at www.sciencedirect.com
ScienceDirect
Journal of Chromatography B, 849 (2007) 43–52
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Review

Higher dimensional (Hi-D) separation strategies dramatically improve the potential for cancer biomarker detection in serum and plasma☆

Seth A. Hoffman, Won-A. Joo, Lynn A. Echan, David W. Speicher*

The Wistar Institute, 3601 Spruce Street, Philadelphia, PA 19104, USA

Received 11 July 2006; accepted 27 October 2006
Available online 30 November 2006

Abstract

The plasma proteome has a wide dynamic range of protein concentrations and is dominated by a few highly abundant proteins. Discovery of novel cancer biomarkers using proteomics is particularly challenging because specific biomarkers are expected to be low abundance proteins with normal blood concentrations of low nanograms per milliliter or less. Conventional, one- and two-dimensional proteomic methods including 2D PAGE, 2D DIGE, LC–MS/MS, and LC/LC–MS/MS do not have the capacity to consistently detect many proteins in this range. In contrast, new higher dimensional (Hi-D) separation strategies, utilizing more than two dimensions of fractionation, can profile the low abundance proteome.

Keywords: Cancer; Proteomics; Plasma proteome; Serum proteome; Blood proteome; Multidimensional separation; Biomarkers; Protein separation; Peptide separation

Contents

1. Introduction 44
   1.1. Specific new cancer diagnostic tests could dramatically improve morbidity and mortality 44
   1.2. Current clinically utilized cancer tests 45
   1.3. Plasma is a promising source for new disease biomarkers 45
2. One- and two-dimensional protein profiling strategies are unlikely to discover new specific cancer biomarkers 46
   2.1. The plasma proteome is highly complex 46
   2.2. Numerous abundant plasma proteins are affected by the acute phase response and are frequently observed as highly non-specific changes 46
   2.3. Specific cancer biomarkers are likely to be low abundance plasma proteins 47
   2.4. One- and two-dimensional proteome separation strategies do not have adequate detection dynamic ranges for discovering low abundance cancer biomarkers 47
3. Higher dimensional separation strategies evaluate the greatest number of candidate biomarkers 47
   3.1. Initial steps to reduce plasma's complexity 47
      3.1.1. Abundant protein depletion 47

Abbreviations: 1D SDS-PAGE, one-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis; 2D DIGE, two-dimensional difference gel electrophoresis; AMT, accurate mass and time tag; DRE, digital rectal examination; ESI, electrospray ionization; FDA, federal drug administration; FTICR MS, fourier transform ion cyclotron resonance mass spectrometry; Hi-D, higher dimensional; HPLC, high pressure liquid chromatography; HUPO, Human Proteome Organization; ICAT, isotope-coded affinity tags; IEF, isoelectrofocusing; IPI, International Protein Index; LC, liquid chromatography; MARS, multiple affinity removal system; MS, mass spectroscopy; MudPit, Multidimensional Protein Identification Technology; PSA, prostate specific antigen; RP, reverse phase; SELDI-MS, surface enhanced laser desorption ionization mass spec; v/v, volume/volume

☆ This paper is part of a special volume entitiled "Analytical Tools for Proteomics" guest edited by Erich Heftmann.

* Corresponding author. Fax: +1 215 898 0664.
E-mail address: speicher@wistar.org (D.W. Speicher).

1570-0232/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2006.10.069

3.1.2. Solid-phase ligand libraries 49
3.1.3. Sub-proteome capture enrichment: hydrazide chemistry for glycopeptide capture and thiol-affinity resins for cysteinyl peptide capture 49
4. Specific Hi-D approaches show substantial promise for detection of low abundance plasma proteins 50
4.1. A four-dimensional strategy 50
4.2. Intact Protein Analysis System 50
4.3. Three-dimensional separation using immunoaffinity depletion and hydrazide chemistry glycoprotein capture 50
4.4. Using a Hi-D separation strategy to create an Accurate Mass/Time Tags Database 50
5. Conclusion 51
Acknowledgments 51
References 51

1. Introduction

The handful of plasma cancer biomarkers that are currently used clinically were discovered via conventional biomedical research approaches. While they have some utility in a variety of situations including screening, diagnosis, staging, targeting therapy, monitoring therapy, and monitoring for disease recurrence, all of these biomarkers lack sufficient sensitivity and specificity for routine screening of either the general population or high risk groups. Proteomic techniques have so far not been able to contribute useful new biomarkers to this set. Interestingly, despite a steadily increasing number of publications discussing potentially useful biomarkers [1], the number of FDA approvals for biomarker tests has been steadily decreasing [2]. Candidate biomarkers must leap many hurdles before gaining clinical acceptance. They must be validated in large patient cohorts, achieve FDA approval, satisfy cost-benefit ratios, and finally achieve recognition amongst the clinicians who order the test [1].

The lack of new biomarkers discovered by proteomics is primarily due to proteomic technology limitations relative to the complexity of the problem. Plasma and serum proteomes contain a large number of proteins spanning a wide dynamic range of concentrations. Most proteins are extensively and heterogeneously modified, and a few very abundant proteins limit the amount of plasma that can be separated and analyzed [2,3]. These characteristics are driving the development of new proteomic methodologies for the study of plasma [4]. However, in the first decade of the proteomics era, 2D PAGE or one- and two-dimensional non-gel protein profiling methods have usually been used to attempt to discover plasma cancer biomarkers. These methods are unable to effectively mine the low abundance plasma proteome (<100 ng/ml) with the detected proteins limited primarily to high (0.1–40 mg/ml) and medium abundance (0.1–100 μg/ml) proteins. Although cancer-related changes to some high and medium abundance plasma proteins do occur and can be detected by these methods, such changes are usually involved in the acute phase response (see below), which is not specific to cancers. Hence, these proteomic methods have a very low probability of discovering proteins that will achieve validation as specific plasma cancer biomarkers. Proteomics discovery of biomarkers for other diseases face the same challenges posed by the high complexity of plasma proteins, although for some other diseases, specific biomarkers may not be as restricted to low abundance proteins.

Recently, several higher dimensional (Hi-D) separation strategies, which involve more than two protein/peptide separation dimensions, have been developed. This is usually achieved by adding one or more orthogonal protein separations prior to LC–MS/MS or LC/LC–MS/MS analysis using high speed, high sensitivity mass spectrometers. These Hi-D separation methods can usually detect a substantial portion of the low abundance plasma proteome, and therefore represent the most promising strategies for discovery of novel specific cancer biomarkers with high potential for achieving clinical utility.

1.1. Specific new cancer diagnostic tests could dramatically improve morbidity and mortality

Cancer is the second leading cause of death in the United States. It was responsible for 22.9% of all deaths in 2001, and is the leading cause of death for persons under the age of 65. Age-adjusted cancer death rates have not changed significantly over the past 25 years [5]. However, a patient's probability of cancer survival is strongly correlated with the cancer's stage at diagnosis. Highly effective, accurate screening tests that could detect most cancers at an early stage, before metastasis has occurred, should dramatically reduce cancer death rates if high-risk groups are assayed at appropriate intervals.

A cancer screen is judged on its ability to correctly diagnose the presence of cancer. This ability is described by the test's sensitivity and specificity. Sensitivity is defined as the probability that a test will be positive given a patient with the condition. Specificity is defined as the probability that a test will be negative given a patient without a condition. A test with poor specificity will often be positive even if the disease is not present, i.e. a false positive result.

Few good molecular cancer screens exist because adequate sensitivities and specificities are difficult to achieve. Furthermore, the predictive value of a test depends on the disease's prevalence in the population. This makes identification of appropriate screens for low-prevalence cancers extremely challenging because even specificities as high as 99% will result in high numbers of false positive tests when the general population is screened. For example, a test with 100% sensitivity (detects all true cases) and 99% specificity for ovarian cancer, which has a prevalence of about 5/10,000 women, would be positive in approximately 105 women per 10,000 women screened. This would include the five actual cancer cases, but also would

include approximately 100 false positive results that lead to substantial unnecessary anxiety, as well as expensive follow up testing and procedures that carry additional health risks.

Achieving high specificity for blood tests is complicated by the fact that protein concentrations within non-diseased subjects vary substantially. These ranges in the normal population are referred to as "reference ranges", which often substantially overlap the concentration ranges of diseased subjects. Attempts to increase a test's specificity by widening the reference range so that only the most extreme test results lay outside this range decreases the test's sensitivity. Determining an optimal cut-off value in protein concentration between normal and abnormal ranges can be quite difficult. Often substantial disputes ensue regarding the appropriate balance between reducing false positives and increasing false negatives.

1.2. Current clinically utilized cancer tests

Most common cancer screening tests currently utilized clinically are morphological rather than molecular assays, e.g. mammography for breast cancer, the Papanicolaou (Pap) smear for cervical cancer, and colonoscopy for colon cancer. Prostate specific antigen (PSA) is the only serum/plasma protein biomarker that has undergone extensive clinical scrutiny as an initial cancer detection method and is relatively widely used. However, there is general agreement that even this test has a suboptimal, high false positive rate. Only one in four men with slightly elevated PSA levels and normal digital rectal exam (DRE) will have prostate cancer [6]. Additionally, the test often over-diagnoses because it correctly detects slow growing cancers that are not life-threatening and do not require clinical treatment. The over-diagnosis rate using PSA is estimated to be 29% for white men and 44% for black men [7]. While currently available data suggest that mass PSA screening combined with DRE may decrease prostate cancer specific mortality, results from important randomized controlled trials, which are expected to be completed in a few years, are needed to unambiguously determine the impact of this assay on mortality reduction [6].

Other blood protein biomarkers routinely used in clinical practice today are not considered specific enough for screening large at-risk populations and instead are usually used to: (a) diagnose and stage disease after other signs and symptoms of cancer become apparent, (b) target therapy, or (c) monitor for disease response to therapy and for post-therapy recurrence. Protein biomarkers could also be combined with other tests or clinical parameters to stratify patients with regard to risk of having cancer previous to determining who should progress to further, more expensive or invasive screening testing.

The current clinically utilized blood protein biomarkers were discovered many years ago using conventional biomedical research approaches, e.g. PSA in 1971, *Human chorionic gonadotropin* (hCG) in 1927, *Alpha fetoprotein* (AFP) in 1956, *Carcinoembryonic antigen* (CEA) in 1955, *Cancer antigen*-125 (CA 125) in 1981, and *Cancer antigen* 15-3 (CA 15-3) in 1984 [8]. These proteins are low abundance in blood of normal individuals and are often secreted proteins (Table 1).

Table 1
Commonly used cancer protein biomarkers are low abundance in normal human plasma/sera

Plasma cancer biomarker	Reference range[a] (ng/ml)	Difference from albumin
Alpha fetoprotein[b]	0–15 (s/p)	10^6
Human chorionic gonadotropin-B[b]	0–0.1[c] (s)	10^8
Carcinoembryonic antigen	0–3 (s)	10^7
Thyroglobulin[b] [50]	1.8–68 (s/p)	10^6
Prostate-specific antigen[b]	0–4 (s/p)	10^7
HER2/NEU [51]	5.1–13.5 (s)	10^6

[a] Reference values published by a Northeastern Tertiary Care University Hospital's specialty laboratory assayed in serum (s), plasma (p) or both (s/p) unless otherwise noted [52].
[b] Secreted protein.
[c] From [53].

1.3. Plasma is a promising source for new disease biomarkers

While many tissues and biological fluids, including tumor tissues and tumor cell lines, have been used as specimens in proteomic studies [9], plasma and serum are particularly promising sources for cancer biomarkers. Blood collection is minimally invasive, can be readily performed at remote locations, and assays are routine and relatively economical. Furthermore, large archives of blood exist for initial discovery of new biomarkers. Since most cells in the body are thought to leak and secrete proteins into the plasma [2], the plasma proteome may reflect the health status of every organ and tissue in the body. Of course plasma and serum proteomes are thought to be highly complex at least in part due to this shedding of proteins and protein fragments by diverse tissues.

So far it is generally thought that only a modest portion of the normal plasma proteome has been defined. Until very recently, most plasma or serum proteome studies identified <500 proteins, regardless of method [2,10,11]. About 2 years ago, a non-redundant set of 1173 plasma proteins was produced by combining datasets from several different laboratories [12]. At about the same time, an individual study reported either about 800 or almost 1700 protein identifications depending upon the analysis criteria used [13]. Very recently, a more comprehensive plasma proteome was produced as part of the Human Proteome Organization's (HUPO) Plasma Proteome Project's pilot study [14]. LC–MS/MS data from a total of 57 serum or plasma proteome analyses from 18 participating labs worldwide was compiled and analyzed in this study. A unique feature of the HUPO study was that all datasets were reanalyzed using consistent analysis methods, a single version of the IPI human database, and uniform statistical criteria for all peptide identifications. A composite database of merged results was created [14–16] and placed on the web [17]. It represents the most comprehensive plasma/serum proteome currently available, with a core dataset of 3020 proteins identified by at least two different "high confident" peptides, and a more rigorously defined subset of 889 proteins identified with a 95% confidence limit [16,17].

Plasma is the non-cellular liquid component of unclotted blood, and serum is the non-cellular liquid component of blood remaining after coagulation, which involves activation of a protease cascade. These proteases and the associated incubation and handling steps further increase sample complexity through incidental proteolysis of non-coagulation proteins. Both types of samples have been used in past proteome studies, although recently the Specimens Committee of the HUPO plasma proteome project recommended using EDTA-plasma and proposed general guidelines for pre-analytical sample handling and storage [18].

2. One- and two-dimensional protein profiling strategies are unlikely to discover new specific cancer biomarkers

2.1. The plasma proteome is highly complex

The plasma proteome presents a number of challenges that obligate high analytical capacity of proteomic methods designed to discover low abundance plasma cancer biomarkers. In 2002, Anderson and Anderson estimated the dynamic range of blood proteins by comparing clinical reference intervals for a set of plasma proteins derived from validated diagnostic assays. Comparing the most abundant protein (albumin at 35–50 mg/ml) with the least abundant protein (interleukin 6 at 0–0.005 ng/ml), they demonstrated that a concentration range of greater than 10^{10} is represented by known blood proteins [2]. As detection sensitivities increase we are likely to discover important disease related proteins in the sub-pg/ml range, which means that plasma proteome profiling methods should ultimately be able to handle dynamic ranges substantially wider than 10^{10}.

The total protein content of plasma in a normal individual generally varies between 60 and 80 mg/ml, and is dominated by albumin and a modest number of additional highly abundant proteins that severely limit the volume of plasma that can be separated and analyzed by most proteomic methods. One study estimated that only 22 high abundant proteins comprise about 99% of plasma protein content [3]. Exactly which proteins make up the top 99% of the plasma proteome is somewhat ambiguous and may vary between subjects due to the wide concentration ranges of many plasma proteins in the normal population. The remaining 1% of the plasma proteome, comprised of medium and low abundance proteins, presumably contains thousands of plasma proteins, although this has not yet been explicitly and convincingly demonstrated. This 1% of the total protein content has an estimated dynamic concentration range of greater than 10^7-fold, and all specific cancer biomarkers are expected to be in this sub-proteome.

Analysis of plasma proteins is further complicated by extensive molecular heterogeneity. Many plasma proteins are present in blood in highly heterogeneous forms. Isoforms often span multiple pI's and molecular weights, and obscure detection of less abundant species, especially when methods capable of detecting molecular heterogeneity such as 2D PAGE are used (Fig. 1). These isoforms include: heterogeneous post-translational modifications, especially glycosylation, proteolytic fragments, and other physiological and artifactual forms such as oxidative modifications.

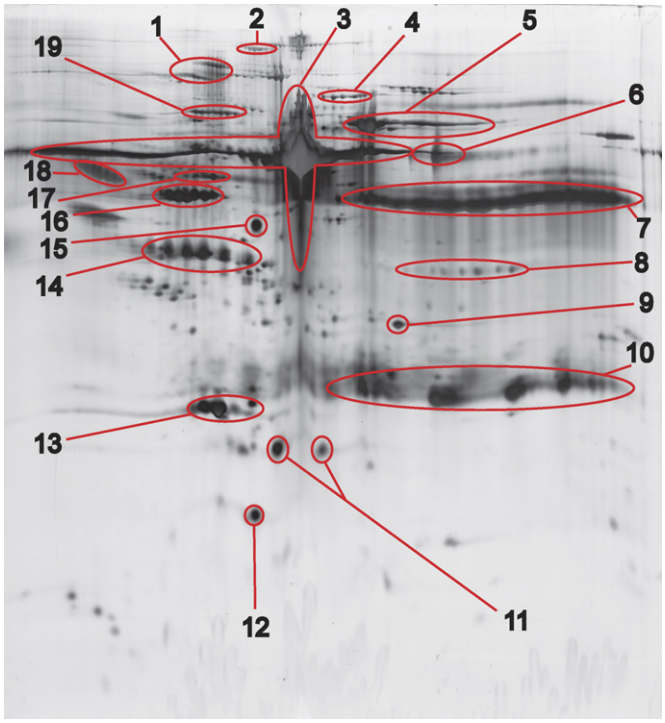

Fig. 1. Plasma proteome on 2D SDS-PAGE. One hundred micrograms of unfractionated serum run on a 2D gel (18 cm × 18 cm, pH 3–10 IPG, 10% Tris/Tricine, silver stain). Protein spots known to be among the most abundant serum proteins (encircled) were named via correlation with 2D images from the on-line SWISS-2D PAGE map selection tool [54]. Overloading of the gel is evident by the negative staining of albumin as well as extensive streaking of the most abundant serum proteins in both dimensions. Despite the heavy load, few medium abundance and no low abundance proteins are detectable. The estimated dynamic range for this 2DE is about 10^2. Identified proteins are: (1) ceruloplasmin, (2) α-2 macroglobulin, (3) albumin, (4) complement factor B, (5) transferrin, (6) complement c3, (7) and (8) immunoglobulin heavy chain γ, (9) c4 complement, (10) immunoglobulin light chain, (11) and (14) haptoglobin, (12) prealbumin, (13) apolipoprotein A, (15) fibrinogen, (16) α-1 antitrypsin, (17) α-1 antichymotrypsin, (18) immunoglobulin heavy chain α, (19) secretory immunoglobulin chain α.

2.2. Numerous abundant plasma proteins are affected by the acute phase response and are frequently observed as highly non-specific changes

The acute phase is the body's standard response to a wide variety of inflammatory or environmental insults, including infection, trauma, surgery, burns, cancer, and even psychiatric disease or psychological distress. The term "acute" phase is a misnomer. The response occurs in both acute conditions, which develop rapidly, and chronic conditions like cancer, which develop slowly over time. The response is characterized by a variety of clinical phenomena. Important to proteomic studies, it involves substantial changes in the concentrations of numerous plasma proteins, termed "acute phase proteins" [19]. These proteins are not useful cancer biomarkers because changes in their plasma concentrations could be due to a

number of heterogeneous diseases or environmental insults and therefore lack specificity for cancer.

Most high abundance plasma proteins are involved in the acute phase response. Of the 22 highest abundance plasma proteins, at least 10 proteins vary by at least 25% in the acute phase response, including: albumin, transferrin, fibrinogen, alpha-1-antitrypsin, C3, haptoglobin, ceruloplasmin, C4, factor B, and C9 [19,3]. At least 7 of the remaining 12 high abundance proteins have also been reported to change concentrations in the acute phase response, including: prealbumin [20], alpha-1-acid glycoprotein, alpha-2-macroglobulin [21], apolipoprotein A-1 [22], apolipoprotein B [23], lipoprotein (a) [24], and C8 [25,3].

2.3. Specific cancer biomarkers are likely to be low abundance plasma proteins

To develop specific plasma cancer biomarker tests, it is important to discover proteins that are either shed exclusively by the tumor, or shed at low levels by the specific non-malignant tissue and at a higher level by the tumor. PSA is an example of the latter situation. In support of this argument, surveys of the plasma proteome suggest that tissue-specific proteins are usually present in plasma as low abundance proteins [2]. Furthermore, all of the current, clinically utilized cancer plasma biomarkers are low abundance plasma proteins (<100 ng/ml) that are usually less than one millionth the concentration of albumin (Table 1). Based on these precedents, specific cancer biomarkers will almost certainly be low abundance plasma proteins, and as we develop the ability to detect extremely low abundance proteins (sub-pg/ml), we may find that such proteins may be the most specific cancer biomarkers.

2.4. One- and two-dimensional proteome separation strategies do not have adequate detection dynamic ranges for discovering low abundance cancer biomarkers

Most 1D and 2D proteomics strategies utilize either protein (e.g. SELDI MS, 2D PAGE, 2D DIGE) or peptide separations (e.g. LC–MS/MS, LC/LC–MS/MS). Such analyses are now relatively routine, and while some of these methods, e.g. LC/LC–MS/MS, can be at least partially automated, throughput remains low for most methods, with the possible exception of SELDI-MS and related approaches. Other relatively high throughput methods are antibody arrays or similar, specific ligand affinity-arrays. But these methods are not discovery based, i.e. the only proteins that can be detected and quantitated are those known proteins targeted by the available antibodies or similar affinity ligand. Antibody arrays are further limited by the relatively small number of suitable antibodies currently available.

The most important limitation of current 1D and 2D methods is that they have relatively narrow detection dynamic ranges. For example, 2D PAGE and 2D DIGE have dynamic ranges of about 10^2 to 10^4, depending on the detection method used. Hence, 2D PAGE of neat plasma is dominated by plasma's abundant proteins and detects no low abundance protein spots (Fig. 1). Even after depletion of six abundant proteins and loading 10–20 times more plasma volume relative to before depletion, no low abundance proteins could be detected on large, high-resolution 2D gels using a sensitive silver stain [26]. LC/LC–MS/MS methods have a better chance of detecting a few low abundance proteins, but in general all 1D and 2D strategies predominantly detect high and medium abundance plasma proteins and their proteolytic fragments [14]. Hence, it is not surprising that when these approaches are used to profile cancer-related plasma proteome changes, the feature changes usually observed are non-specific acute phase reaction proteins. Re-identifying such protein changes, as has typically occurred in SELDI-MS based cancer biomarker discovery studies, is very unlikely to lead to identification of novel specific cancer biomarkers.

3. Higher dimensional separation strategies evaluate the greatest number of candidate biomarkers

Recently, a number of research groups have developed separation strategies that incorporate more than two tandem separation methods. These higher dimensional (Hi-D) strategies invariably add protein separation steps, because the LC/LC–MS/MS approach has already maximally exploited available physical differences of tryptic peptides. Abundant protein depletion using immunoaffinity resins to remove low probability, obscuring species is usually the first fractionation step in most Hi-D strategies. As expected, the throughput of most current Hi-D strategies is even lower than for 1D and 2D methods due to the large number of fractions created for subsequent analysis by LC–MS/MS. This limits these approaches to the types of studies that require the analysis of only a few samples. Hi-D strategies are promising approaches because they show enhanced capacities to detect low abundance plasma proteins.

3.1. Initial steps to reduce plasma's complexity

3.1.1. Abundant protein depletion

Early protein depletion methods included Cibacron blue, a chlorotriazine dye with high affinity for albumin [27], and Protein A/G, which depletes immunoglobulins [28]. Cibacron blue columns showed low specificity with removal of significant amounts of non-targeted proteins in the depleted fraction. Also, these dye-based columns and Protein A/G systems incompletely removed the targeted proteins. This was a serious limitation because even if 90% of albumin was removed, the approximately 4 mg/ml of albumin that remained was still one of the most abundant proteins [26].

More recently, polyclonal antibody-based depletion columns such as the multiple affinity removal system (MARS) from Agilent (Wilmington, DE) became available commercially. Over the past several years, a number of similar products that deplete different numbers of abundant proteins have become available from multiple suppliers. In general, the antibodies used in these affinity columns are polyclonal, which unlike monoclonal antibodies have the potential for variability in affinity and specificity over the long term as antibodies are produced in different hosts. However, polyclonal antibodies are still preferred over monoclonal

antibodies because they are much more likely to recognize and deplete most forms of the targeted abundant proteins, including molecules with different post-translational modifications and proteolytic fragments [29,30].

The MARS "Top-6" is an HPLC column that effectively removes the six most abundant plasma proteins efficiently, including: albumin, transferrin, haptoglobin, alpha-1-antitrypsin, IgG, and IgA, which comprise about 85% of the total plasma protein content, with minimal non-targeted protein depletion. This enables substantially higher volumes of plasma to be separated in subsequent fractionation steps and increases the number of proteins that can be detected by most analysis methods. However, the next most abundant proteins quickly limit even higher protein loads and it quickly becomes clear that depletion of even more abundant proteins would be advantageous [26].

A subsequent commercial immunodepletion column used immunoglobulin yolk antibodies (IgY) in a "MIXED12" column (GenWay Biotech, San Diego, CA), to deplete plasma or serum of 12 abundant proteins. The column is reported by the manufacturer to remove 95–99% of these proteins effectively with high specificity. On 2D PAGE, spots previously hidden by the abundant proteins became apparent but were not further identified [31]. Hence, these new spots may simply be minor forms of high and medium abundance proteins as was observed with Top-6 depleted samples [26].

In late 2005, we beta tested a Top 20 depletion column (the Sigma ProteoPrep-20) that was commercialized in early 2006 (Fig. 2). This column removes 20 high abundant plasma/serum proteins; albumin, transferrin, fibrinogen, haptoglobin, alpha-2-macroglobulin, alpha-1-antitrypsin, complement C3, complement C4, complement C1q, IgGs, IgAs, IgDs, IgMs, apolipoprotein A1, apolipoprotein A2, apolipoprotein B-100, acid-1-glycoprotein, ceruloplasmin, prealbumin, and plasminogen. This depletion column is substantially superior to columns that deplete fewer proteins because it further reduces the total plasma protein content and allows up to 100–200-fold more plasma volume to be loaded into downstream separations and analyses after depletion, thereby greatly enhancing detection of low abundance proteins. Most targeted abundant proteins were extensively depleted (>98%). As a bonus, 26 other abundant proteins were partially removed, which either shared high sequence homology with a targeted protein or were known to exist in tight complexes with a targeted protein. Only eight medium-to-low abundance proteins were detected in the bound fraction, but these proteins were still predominantly in the unbound (depleted) fraction. The partial depletion of a few non-targeted proteins is not surprising and is not a significant problem. Of course, if one wishes to ensure that no potential biomarkers are overlooked, one could analyze both bound and unbound fractions in downstream analyses [32].

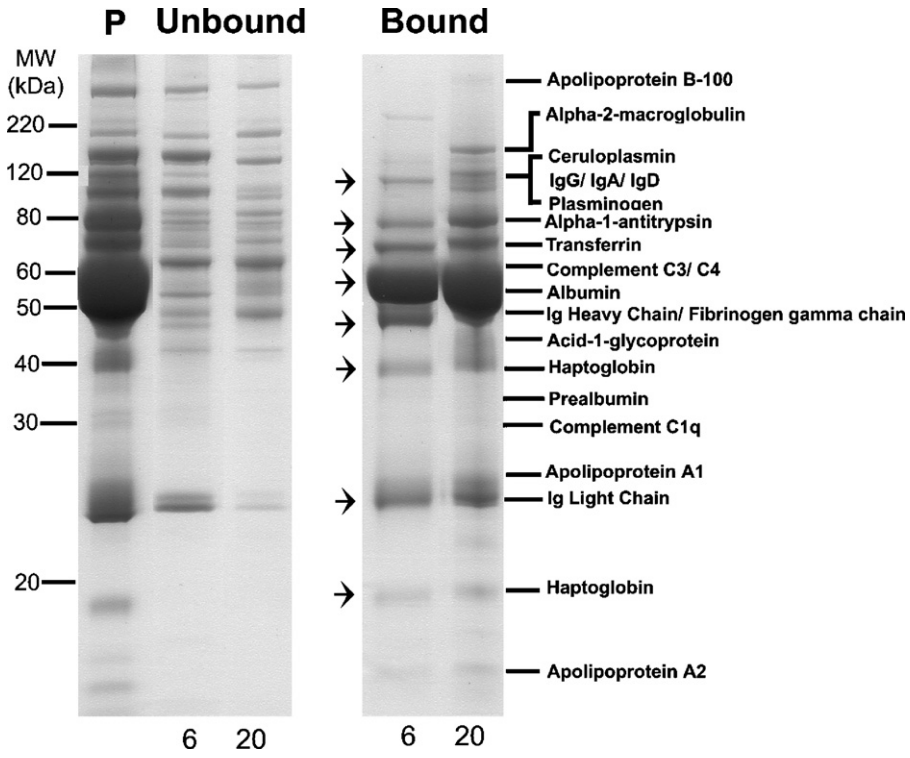

Fig. 2. Comparing different depletion columns. Human plasma (50 μl) was depleted using a Multiple Affinity Removal System HPLC column (Top 6, 4.6 mm × 100 mm) from Agilent. Albumin, IgG, transferrin, IgA, α-1-antitrypsin and haptoglobin (indicated as arrows) were effectively depleted using this column; however, other major proteins still limit sample loads and detection of low abundance proteins. In a similar experiment, human plasma (100 μl) was depleted using a large Sigma ProteoPrep-20 Immunodepletion spin column (Top 20 depletion). The positions of most prominent isoforms of the depleted proteins are indicated. Samples were separated on 12% Bis-Tris 1D gels and stained with colloidal Coomassie blue; all samples loaded onto the gel were volume normalized to the undepleted sample (25 μg).

3.1.2. Solid-phase ligand libraries

A recently developed alternative approach to abundant protein reduction uses solid phase random ligand libraries generated by classical combinatorial synthesis. A large number of different ligands are produced that will presumably bind to specific proteins in the plasma. Such ligand library columns with equal amounts of all ligands have been proposed to bind similar amounts of both abundant proteins and low abundance proteins, thereby greatly enriching low abundance proteins [33,34]. Analysis of serum passed through a solid phase ligand library and sequentially eluted showed altered patterns compared with 2D PAGE of neat serum, but only a few new protein spots were detected and these spots were not identified [35]. At this point it is not clear whether solid phase ligand libraries substantially increase the detection of low abundance plasma proteins. More importantly, it is not clear that such strategies can be used to discover quantitative changes in different plasma samples. Further development of this technique is needed before its utility for biomarker discovery can be fully assessed.

3.1.3. Sub-proteome capture enrichment: hydrazide chemistry for glycopeptide capture and thiol-affinity resins for cysteinyl peptide capture

Protein glycosylation is a common post-translational modification and alterations in protein glycosylation often correlate with cancer. Hence, isolation of this subset of plasma proteins may facilitate detection of cancer biomarkers. Indeed, many of the current clinical protein biomarkers are glycosylated. A recently developed Hi-D approach specifically isolates and analyzes *N*-linked glycopeptides. Carbohydrate moieties on plasma proteins are oxidized, and the glycoproteins are covalently linked to a hydrazide resin via their oxidized carbohydrates. Non-glycosylated proteins do not bind to the column and are washed through. Bound proteins are digested with trypsin and non-covalently linked peptides are removed. *N*-linked tryptic peptides are specifically released by digestion with *N*-glycosidase [36].

Application of this technique to mouse serum followed by LC–MS/MS identified 93 glycoproteins in an initial study. Of these, nine were known to have concentrations in the low μg/ml range in human serum. Peak intensities for these peptides were on average 900-fold greater than noise and the authors estimated this system should have the capability to detect proteins on the order of ng/ml [37].

Another affinity method targets cysteine-containing peptides [38,39]. Cysteines typically constitute approximately 1.7% of amino acids in proteomes, and therefore enriching peptides with cysteines provides a substantial simplification of complex peptide mixtures. In a recently published study, a thiol-specific resin was used to enrich cysteine-containing peptides from a mammary epithelial cell proteome before fractionation by strong cation exchange and identification by LC–MS/MS. A number of low abundance proteins were detected in the cysteinyl-enriched fraction, which were not identified in non-enriched fractions [40].

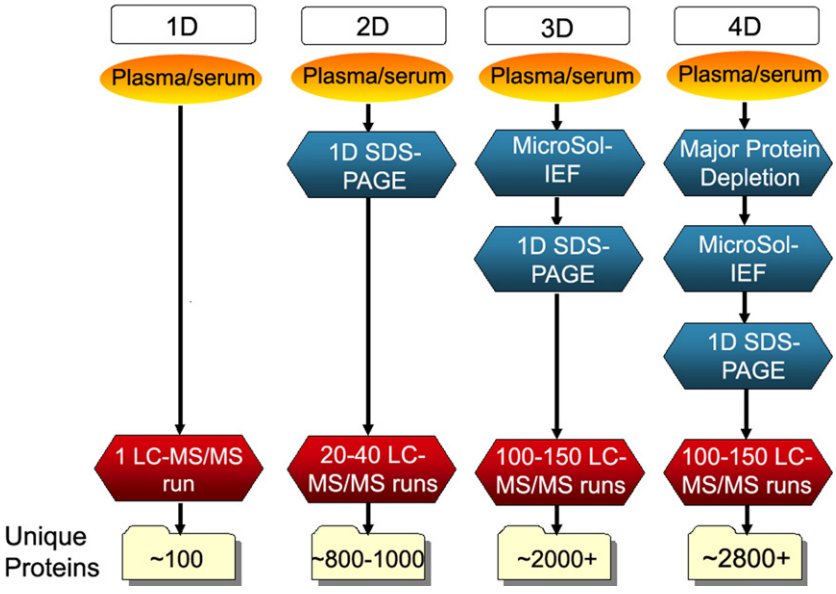

Fig. 3. Schematic illustrating impact of higher dimensional separation strategies on depth of plasma proteome analysis. Hi-D separations increase total numbers of LC–MS/MS runs and thereby reduce throughput while substantially increasing the number of identified proteins. Protein separation steps (blue boxes) and the reverse phase peptide separation common to all methods (red boxes) are shown. Total numbers of LC–MS/MS runs that might be used with the different schemes are shown, although the number of fractions utilized at the MicroSol-IEF and SDS-PAGE steps can be readily adjusted up or down to increase resolution or to increase throughput, respectively. The estimated approximate number of unique plasma proteins typically identified using a hybrid linear ion trap mass spectrometer (Thermo LTQ FT) are illustrated.

4. Specific Hi-D approaches show substantial promise for detection of low abundance plasma proteins

4.1. A four-dimensional strategy

We recently developed an integrated powerful 4D protein profiling method for in depth analysis of serum and plasma proteomes. This method combines three protein separations with sensitive LC–MS/MS analysis of tryptic peptide mixtures. Major protein depletion is performed with a MARS "Top-6" polyclonal immunoaffinity column or in more recent studies, using the ProteoPrep-20 "Top-20" column. This is followed by micro-scale solution IEF (MicroSol IEF) using a ZOOM-IEF fractionator (Invitrogen, Carlsbad, CA) [41,42]. Subsequent fractions are run on 1D SDS-PAGE gels. Lanes from each fraction are divided into uniform slices, or pixels, and in-gel trypsin digestion is performed on all pixels. Peptide digests are then analyzed by nano-LC–MS/MS using a high sensitivity linear ion trap mass spectrometer [43].

Fig. 3 illustrates the impact of adding one or more protein separation methods prior to LC–MS/MS analysis using a high speed, high sensitivity linear ion trap mass spectrometer. Addition of the protein separation steps increases the number of LC–MS/MS runs while substantially increasing the number of total proteins and total number of low abundance proteins that are identified.

Key advantages of the 4D strategy include: (1) the three protein separations are orthogonal and very high resolution with minimal distribution of specific proteins between fractions, which reduces fraction complexity far more than lower resolution methods; (2) the number of fractions generated in early steps is low, which minimizes the number of parallel separations required in subsequent steps; and (3) although the total number of fractions per proteome is high in later steps, these steps (trypsin digestion and LC–MS/MS analysis) have been automated.

Our 4D method has proven to be quite powerful when applied to human serum and plasma [26,43]. As mentioned above, we participated in the pilot phase of the HUPO Plasma Proteome Project and applied our 4D method to several HUPO samples. Initially we analyzed a plasma sample using the 4D method with Top 6 depletion (the Top 20 column was not available at that time) and the LC–MS/MS step was performed using a 3D ion trap (LCQ Deca XP+, Thermo Electron, San Jose, CA). We detected 575 proteins using HUPO criteria for "high confidence" assignments. Our subsequent analysis of a HUPO serum sample utilized a faster, more sensitive LTQ linear ion trap mass spectrometer (Thermo Electron, San Jose, CA), which dramatically expanded the proteins detected to 2890 proteins using the HUPO criteria. This number of proteins was much larger than those obtained in any other dataset from other participating labs regardless of the type of mass spectrometer used (several used Thermo LTQs) or the analysis methods, which included Multidimensional Protein Identification Technology (MudPIT), Isotope-Coded Affinity Tags (ICAT), accurate mass and time tags, etc. [14]. The most compelling indication of the power of our 4D method came from comparison of our HUPO serum dataset to quantitation of known proteins in the same samples by Haab et al. [44]. In our HUPO serum/LTQ dataset, we identified 14 of a total of 20 proteins in the 1–100 ng/ml range and 3 of 23 proteins in the pg/ml range. It is estimated that the optimized method allows a detection dynamic range of about 10^9.

4.2. Intact Protein Analysis System

The Intact Protein Analysis System (IPAS) is a three-dimensional intact protein analysis strategy with quantitative comparison capability. Case and control samples are first immunodepleted of abundant proteins using an Agilent MARS Top-6 depletion column. Paired case and control samples are labeled with sample-specific Cy dyes. (Cy dyes are fluorescent tags often used to differentiate a protein's sample of origin when viewed on a single output.) After Cy dye labeling the samples are subsequently mixed. Proteins are then separated in three dimensions using: (1) Rotofor liquid-based IEF (Bio-Rad, Hercules, CA), (2) reverse phase HPLC separation of proteins, and (3) 1D SDS-PAGE. Images of the 1D gels are scanned and analyzed with spot analysis software to identify significant Cy dye ratio changes between case and control samples. Spots of interest are subjected to in-gel trypsin digestion and MS/MS for identification [45,46]. In a study profiling the plasma proteome for changes that occur with acute graft-versus-host-disease, this method identified 75 proteins in the micromolar (approximately μg/ml) to femtomolar (approximately fg/ml) range exhibiting quantitative differences between pre- and post-graft-versus-host-disease samples [45]. The extent of low abundance coverage is somewhat difficult to assess because reference concentrations for the identified proteins were not provided, and only proteins exhibiting changes in abundance were identified in this method.

4.3. Three-dimensional separation using immunoaffinity depletion and hydrazide chemistry glycoprotein capture

In this approach, plasma was subjected to "Top-6" abundant protein depletion followed by the *N*-linked glycoprotein capture method described in Section 3.1.3. After trypsin digestion, peptide samples were analyzed by strong cation exchange chromatography followed by reverse-phase capillary LC–MS/MS. A total of 303 non-redundant *N*-glycoproteins were confidently identified, including proteins with reported plasma concentrations in the ng/ml to pg/ml range. Eighteen of these proteins have approximate concentrations of <100 ng/ml based on previously published concentrations. The authors estimate this Hi-D approach has a detection dynamic range of 10^8 [47].

4.4. Using a Hi-D separation strategy to create an Accurate Mass/Time Tags Database

The Accurate Mass and Time (AMT) Tag approach is a high throughput 1D nanoLC–MS method utilizing high performance reverse phase columns and an FT ICR mass spectrometer [48]. To identify a feature, its mass and elution times are compared

with a database of known proteins' masses and corresponding elution times. Development of such a database involves low throughput nano-LC MS/MS methods coupled with prior protein fractionation.

Two Hi-D separation methods were recently used in parallel to create a comprehensive reference AMT database for plasma from trauma patients [49]. Plasma from trauma patients was pooled and subjected to "MIXED12" abundant protein depletion (GenWay, San Diego, CA), and two aliquots of depleted plasma were created. One aliquot was subjected to trypsin digestion and cysteinyl enrichment on a thiol affinity resin. The resultant cysteinyl and non-cysteinyl peptide fractions were each further fractionated by strong cation exchange chromatography, and then identified via LC–MS/MS. In addition, the second aliquot of depleted plasma was subjected to *N*-glycopeptide enrichment on a hydrazide resin. Carbohydrate moieties on plasma proteins were oxidized, and the glycoproteins were covalently linked to a hydrazide resin via their oxidized carbohydrates. Non-glycosylated proteins did not bind to the column and were washed through. Bound proteins were digested with trypsin and non-covalently linked peptides were removed. *N*-linked tryptic peptides were specifically released by digestion with *N*-glycosidase. The *N*-glycopeptides and non-glycopeptide fractions were each further fractionated by strong cation exchange chromatography, and then identified via LC–MS/MS. This Hi-D method was able to identify 3654 proteins when all peptide populations were combined (cysteinyl, non-cysteinyl, *N*-glycosylated, non-glycosylated). Confident identification of low abundance proteins was exemplified by detection of 78 "classic" cytokines and cytokine receptors and 136 human cell differentiation molecules. Six low abundance proteins related to the inflammatory response were assayed by ELISA and found to exist at concentrations ranging from 0.5 to 20 ng/ml in the trauma subjects. The authors estimate an overall detection dynamic range of 10^7 [49].

5. Conclusion

Proteomics has so far failed to discover new plasma cancer biomarkers with clinical utility presumably because the complexity of the blood proteome substantially exceeds the analytical capacity of conventional protein profiling methods. These conventional proteomics methods utilize only one or two separation dimensions of either peptides or proteins and primarily rediscover cancer-related changes in high abundance plasma proteins associated with the well known and non-specific acute-phase reaction. Recently developed Hi-D separation strategies include one or more protein separations followed by trypsin digestion and subsequent analysis of resulting peptide mixtures by either LC–MS/MS or LC/LC–MS/MS. These Hi-D methods can more effectively overcome the complexity of the plasma proteome and can detect a substantial number of low abundance proteins (<100 ng/ml), which is the concentration range where specific cancer biomarkers are expected. Hence, Hi-D strategies have the greatest potential for discovery of the next generation of plasma cancer biomarkers.

Acknowledgments

This work was supported by NIH grant CA77048 and grant SAP4100020718 from the PA Department of Health to D.W.S and a Post-doctoral fellowship from the Korea Science & Engineering Foundation (KOSEF) to W.J. This work was also supported by institutional grants to the Wistar Institute, including a National Cancer Institute Cancer Core Grant (CA10815), and the Commonwealth Universal Research Enhancement Program, Pennsylvania Department of Health.

References

[1] J.A. Ludwig, J.N. Weinstein, Nat. Rev. Cancer 5 (2005) 845.
[2] N.L. Anderson, N.G. Anderson, Mol. Cell. Proteom. 1 (2002) 845.
[3] R.S. Tirumalai, K.C. Chan, D.A. Prieto, H.J. Issaq, T.P. Conrads, T.D. Veenstra, Mol. Cell. Proteom. 2 (2003) 1096.
[4] J.J. Jacobs, J.N. Adkins, W.-J. Qian, T. Liu, Y. Shen, D.G. Camp II, R.D. Smith, J. Proteom. Res. (2005) 1073.
[5] A. Jemal, R.C. Tiwari, T. Murray, A. Ghafoor, A. Samuels, E. Ward, E.J. Feuer, M.J. Thun, CA Cancer J. Clin. (2004) 8.
[6] H.-P. Schmid, W. Riesen, L. Prikler, Crit. Rev. Oncol. Hematol. (2004) 71.
[7] R. Etzioni, D.F. Penson, J.M. Legler, D. Di Tommaso, R. Boer, P.H. Gann, E.J. Feuer, J. Natl. Cancer. Inst. (2002) 981.
[8] J.-M. Bidart, F. Thuillier, C. Augereau, J. Chalas, A. Daver, N. Jacob, F. Labrousse, H. Voitot, Clin. Chem. (1999) 1695.
[9] S. Aldred, M.M. Grant, H.R. Griffiths, Clin. Biochem. (2004) 943.
[10] L. Anderson, J. Physiol. (2005) 23.
[11] J.N. Adkins, S.M. Varnum, K.J. Auberry, R.J. Moore, N.H. Angell, R.D. Smith, D.L. Springer, J.G. Pounds, Mol. Cell. Proteom. (2002) 947.
[12] N.L. Anderson, M. Polanski, R. Pieper, T. Gatlin, R.S. Tirumalai, T.P. Conrads, T.D. Veenstra, J.N. Adkins, J.G. Pounds, R. Fagan, A. Lobley, Mol. Cell. Proteom. (2004) 311.
[13] Y. Shen, J.M. Jacobs, D.G. Camp 2nd, R. Fang, R.J. Moore, R.D. Smith, W. Xiao, R.W. Davis, R.G. Tompkins, Anal. Chem. (2004) 1134.
[14] G.S. Omenn, D.J. States, M. Adamski, T.W. Blackwell, R. Menon, H. Hermjakob, R. Apweiler, B.B. Haab, R.J. Simpson, J.S. Eddes, E.A. Kapp, R.L. Moritz, D.W. Chan, A.J. Rai, A. Admon, R. Aebersold, J. Eng, W.S. Hancock, S.A. Hefta, H. Meyer, Y.K. Paik, J.S. Yoo, P. Ping, J. Pounds, J. Adkins, X. Qian, R. Wang, V. Wasinger, C.Y. Wu, X. Zhao, R. Zeng, A. Archakov, A. Tsugita, I. Beer, A. Pandey, M. Pisano, P. Andrews, H. Tammen, D.W. Speicher, S.M. Hanash, Proteomics 5 (2005) 3226.
[15] M. Adamski, T. Blackwell, R. Menon, L. Martens, H. Hermjakob, C. Taylor, G.S. Omenn, D.J. States, Proteomics (2005) 3246.
[16] D.J. States, G.S. Omenn, T.W. Blackwell, F. Damian, J. Eng, D.W. Speicher, S.M. Hanash, Nat. Biotech. (2006) 333.
[17] Human Proteome Project—Plasma Proteome Project. Available at: www.bioinformatics.med.umich.edu/hupo/ppp. Accessed July 6, 2006.
[18] A.J. Rai, C.A. Gelfand, B.C. Haywood, D.J. Warunek, J. Yi, M.D. Schuchard, R.J. Mehigh, S.L. Cockrill, G.B.I. Scott, H. Tammen, P. Schulz-Knappe, D.W. Speicher, F. Vitzthum, B.B. Haab, G. Siest, D.W. Chan, Proteomics (2005) 3262.
[19] C. Gabay, I. Kushner, N. Engl. J. Med. (1999) 448.
[20] G. Biolo, G. Toigo, B. Ciocchi, R. Situlin, F. Iscra, A. Gullo, G. Guarnieri, Nutrition (1997) 52S.
[21] H. Moshage, J. Pathol. (1997) 257.
[22] A. Chait, C.Y. Han, J.F. Oram, J.W. Heinecke, J. Lipid Res. (2005) 389.
[23] G. Lippi, V. Braga, S. Adami, G. Guidi, Clin. Chim. Acta (1998) 79.
[24] G. Chimienti, F. Aquilino, M.T. Rotelli, F. Russo, L. Lupo, G. Pepe, Br. J. Surg. (2006) 347.
[25] L. Brio, G. Domjan, A. Falus, L. Jakab, K. Cseh, L. Kalabay, G. Tarkovacs, J. Tresch, E. Malle, J. Kramer, Z. Prohaszka, J. Jako, G. Fust, A. Csaszar, Eur. J. Clin. Invest. (1998) 679.
[26] L.A. Echan, H.Y. Tang, N. Ali-Khan, K. Lee, D.W. Speicher, Proteomics 5 (2005) 3292.

[27] J. Travis, J. Bowen, D. Tewksbury, D. Johnson, R. Pannell, Biochem. J. (1976) 301.
[28] N.I. Govorukhina, A. Keizer-Gunnink, A.G.J. van der Zee, S. de Jong, H.W.A. de Bruijn, R. Bischoff, J. Chromatogr. A (2003) 171.
[29] L.F. Steel, M.G. Trotter, P.B. Nakajima, T.S. Mattu, G. Gonye, T. Block, Mol. Cell. Proteom. (2003) 262.
[30] R. Pieper, Q. Su, C.L. Gatlin, S.-T. Huang, N.L. Anderson, S. Steiner, Proteomics (2003) 422.
[31] L. Huang, G. Harvie, J.S. Feitelson, K. Gramatikoff, D.A. Herold, D.L. Allen, R. Amunngama, R.A. Hagler, M.R. Pisano, W.-W. Zhang, X. Fang, Proteomics (2005) 3314.
[32] L.A. Echan, G. Tan, H.Y. Tang, D.W. Speicher, in preparation.
[33] P.G. Righetti, A. Castagna, P. Antonioli, E. Boschetti, Electrophoresis (2005) 297.
[34] V. Thulasiraman, S. Lin, L. Gheorghiu, J. Lathrop, L. Lomas, D. Hammond, E. Boschetti, Electrophoresis (2005) 3561.
[35] L. Guerrier, V. Thulasiraman, A. Castagna, F. Fortis, S. Lin, L. Lomas, P.G. Righetti, E. Boschetti, J. Chromatogr. B (2006) 34.
[36] H. Zhang, X. Li, D.B. Martin, R. Aebersold, Nat. Biotechnol. (2003) 660.
[37] H. Zhang, E.C. Yi, X. Li, P. Mallick, K.S. Kelly-Spratt, C.D. Masselon, D.G. Camp II, R.D. Smith, C.J. Kemp, R. Aebersold, Mol. Cell. Proteom. (2005) 144.
[38] C.S. Spahr, S.A. Susin, E.J. Bures, J.H. Robinson, M.T. Davis, M.D. McGinley, G. Kroemer, S.D. Patterson, Electrophoresis (2000) 1635.
[39] K. Gevaert, B. Ghesquiere, A. Staes, L. Martens, J. Van Damme, G.R. Thomas, J. Vanderkerckhove, Proteomics (2004) 897.
[40] T. Liu, W.-J. Qian, W-N.U. Chen, J.M. Jacobs, R.J. Moore, D.J. Anderson, M.A. Gritsenko, M.E. Monroe, B.D. Thrall, D.G. Camp II, R.D. Smith, Proteomics (2005) 1263.
[41] U.S. Patent No.:6,638,408. Title: METHOD AND DEVICE FOR SEPARATION OF CHARGED MOLECULES BY SOLUTION ISOELECTRIC FOCUSING, David W. Speicher and Xun Zuo. Issue Date: October 28, 2003.
[42] X. Zuo, D.W. Speicher, Proteomics (2002) 58.
[43] H.Y. Tang, N. Ali-Khan, L.A. Echan, N. Levenkova, J.J. Rux, D.W. Speicher, Proteomics 5 (2005) 3329.
[44] B.B. Haab, B.H. Geierstanger, G. Michailidis, F. Vitzthum, S. Forrester, R. Okon, P. Saviranta, A. Brinker, M. Sorette, L. Perlee, S. Suresh, G. Drwal, J.N. Adkins, G.S. Omenn, Proteomics 5 (2005) 3278.
[45] H. Wang, S.G. Clouthier, V. Galchev, D.E. Misek, U. Duffner, C.-K. Min, R. Zhao, J. Tra, G.S. Omenn, J.L.M. Ferrara, S.M. Hanash, Mol. Cell. Proteom. 4 (2005) 618.
[46] D.E. Misek, R. Kuick, H. Wang, V. Galchev, B. Deng, R. Zhao, J. Tra, M.R. Pisano, R. Amunugama, D. Allen, A.K. Walker, J.R. Strahler, P. Andrews, G.S. Omenn, S.M. Hanash, Proteomics 5 (2005) 3343.
[47] T. Liu, W.-J. Qian, M.A. Gritsenko, D.G. Camp II, M.E. Monroe, R.J. Moore, R.D. Smith, J. Prot. Res. (2005) 2070.
[48] J.S.D. Zimmer, M.E. Monroe, W.-J. Qian, R.D. Smith, Mass Spectrom. Rev. (2006) 450.
[49] T. Liu, W-J. Qian, M.A. Gritsenko, W. Xiao, L.L. Moldawer, A. Kaushal, M.E. Monroe, S.M. Varnum, R.J. Moore, S.O. Purvine, R.V. Maier, R.W. Davis, R.G. Tompkins, D.G. Camp II, R.D. Smith, and the Inflammation and the Host Response to Injury Large Scale Collaborative Research Program, Mol. Cell. Proteom. 5 (10) (2006) 1899.
[50] G. Wunderlich, K. Zophel, L. Crook, S. Smith, B.R. Smith, W.G. Franke, Thyroid (2001) 819.
[51] F.J. Esteva, C.D. Cheli, H. Fritsche, M. Fornier, D. Slamon, R.P. Thiel, D. Luftner, F. Ghani, Breast Cancer Res. (2005) R436.
[52] ARUP Laboratories. ARUP User's Guide. Available at: http://www.aruplab.com/testing/user_guide.jsp. Accessed May 19, 2006.
[53] H. Alfthan, C. Haglund, J. Dabek, U.-H. Stenman, Clin. Chem. (1992) 1981.
[54] Swiss Institute of Bioinformatics, Expert Protein Analysis System (ExPASy), SWISS-2DPAGE Human Plasma. Available at: http://us.expasy.org/cgi-bin/map2/def?PLASMA_HUMAN. Accessed: 5/24/2006.