View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Queensland University of Technology ePrints Archive

Clinical Chemistry 51:10
1933–1945 (2005)

Oak Ridge Conference

Analysis of Albumin-Associated Peptides and Proteins from Ovarian Cancer Patients

Mark S. Lowenthal,[1] Arpita I. Mehta,[3] Kristina Frogale,[1] Russell W. Bandle,[1] Robyn P. Araujo,[1] Brian L. Hood,[5] Timothy D. Veenstra,[5] Thomas P. Conrads,[5] Paul Goldsmith,[1] David Fishman,[4] Emanuel F. Petricoin III,[2*] and Lance A. Liotta[2*]

Background: **Albumin binds low–molecular-weight molecules, including proteins and peptides, which then acquire its longer half-life, thereby protecting the bound species from kidney clearance. We developed an experimental method to isolate albumin in its native state and to then identify [mass spectrometry (MS) sequencing] the corresponding bound low–molecular-weight molecules. We used this method to analyze pooled sera from a human disease study set (high-risk persons without cancer, n= 40; stage I ovarian cancer, n = 30; stage III ovarian cancer, n = 40) to demonstrate the feasibility of this approach as a discovery method.**

Methods: **Albumin was isolated by solid-phase affinity capture under native binding and washing conditions. Captured albumin-associated proteins and peptides were separated by gel electrophoresis and subjected to iterative MS sequencing by microcapillary reversed-phase tandem MS. Selected albumin-bound protein fragments were confirmed in human sera by Western blotting and immunocompetition.**

Results: **In total, 1208 individual protein sequences were predicted from all 3 pools. The predicted sequences were largely fragments derived from proteins with diverse biological functions. More than one third of these fragments were identified by multiple peptide sequences, and more than one half of the identified species were in vivo cleavage products of parent proteins. An estimated 700 serum peptides or proteins were predicted that had not been reported in previous serum databases. Several proteolytic fragments of larger molecules that may be cancer-related were confirmed immunologically in blood by Western blotting and peptide immunocompetition. BRCA2, a 390-kDa low-abundance nuclear protein linked to cancer susceptibility, was represented in sera as a series of specific fragments bound to albumin.**

Conclusion: **Carrier-protein harvesting provides a rich source of candidate peptides and proteins with potential diverse tissue and cellular origins that may reflect important disease-related information.**

© 2005 American Association for Clinical Chemistry

The circulatory proteome holds great promise as a reservoir of information useful for disease detection and therapeutic monitoring *(1)*. Despite this potential, comprehensive characterization of the circulatory proteome is difficult because of the wide dynamic range of protein concentrations that exist between larger molecules such as albumin (g/L) and the sought after biomarkers (below ng/L) *(2)*. Here we describe a method for amplifying the yield of low-abundance, low–molecular-weight proteins and peptide fragments in serum as a means of providing a new window into the information content contained within the circulation. This method takes advantage of the tendency for low-abundance molecules to associate with high-abundance proteins such as albumin and thereby acquire the carrier's longevity in the serum *(3)*. This report describes the application of this method to identify candidate proteins and peptides in serum pools from diseased and control ovarian cancer patients.

Current protocols for investigation of the serum proteome often recommend prefractionation by native depletion of high-abundance species (e.g., albumin, im-

[1] Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD.
[2] Center for Applied Proteomics and Molecular Medicine, George Mason University, Manassas, VA.
[3] Tufts University School of Medicine, Howard Hughes Medical Institute, Boston MA.
[4] National Ovarian Cancer Early Detection Program, New York University, New York, NY.
[5] National Cancer Institute Biomedical Proteomics Program, SAIC, NCI, Frederick, MD.
*Address correspondence to Dr. Petricoin or Dr. Liotta at: Center for Applied Proteomics and Molecular Medicine, George Mason University, Manassas, VA 20110. E-mail lliotta@gmu.edu or epetrico@gmu.edu.
Received April 25, 2005; accepted August 1, 2005.
Previously published online at DOI: 10.1373/clinchem.2005.052944

munoglobulins, lipoproteins) *(4–8)*. This approach can inadvertently remove a high percentage of candidate low-abundance proteins and peptides because most low–molecular-weight molecules in serum form complexes with high-abundance proteins, which protect the bound molecules from clearance and allow them to remain in circulation. Albumin is the most abundant protein in plasma and serum, present at 50 g/L, and has a half-life of 19 days in humans *(9–11)*. Because the kidney generally and efficiently filters out molecules <60 kDa, smaller proteins and peptides will be protected from clearance by association with albumin *(9, 10)*. Plasma protein binding can be an effective means of extending the pharmacokinetic properties of otherwise short-lived molecules *(9, 11–13)*. Dennis et al. *(9)* demonstrated that a 58-fold increase in peptide longevity could be attained through albumin binding and association. Albumin binding of biologically important proteins and peptides is well documented. For example, the amino-terminal peptide of HIV-1, gp41, and the 14-kDa fragment of streptococcal protein G are known to specifically bind with human serum albumin *(14, 15)*. Vitamin A transport and homeostasis is controlled by a specific protein–protein interaction between the 21-kDa retinol-binding protein and transthyretin, which was shown to reduce glomerular filtration of the low–molecular-mass retinol-binding protein *(16)*.

It has recently been shown that most low–molecular-weight molecules that have been visualized by mass spectrometry (MS)[6] profiling methods in the past exist under native conditions in a complexed state with high–molecular-weight proteins *(17, 18)*. Direct analysis of the low–molecular-weight portion of size-fractionated native serum revealed few unbound molecular species. These previous studies *(3, 17, 18)* demonstrated that most of the low–molecular-weight species were found in the high–molecular-weight fraction because they were all complexed to other proteins. These same investigators went on to denature the protein complexes and then sequenced the size-fractionated low–molecular-weight repository. The selectivity of binding of low–molecular-weight constituents for different classes of carrier proteins was explored further *(17, 18)*, revealing that subsets of bound low–molecular-weight peptides and sequenced protein fragments associated with albumin were different from those bound to immunoglobulin, apolipoprotein, and transferrin *(18)*. These previous studies established the scientific rationale for the present investigation into the carrier-protein–bound information archive. To evaluate the possibility that this repository of bound molecules contains potential disease-related information, we report the results of an investigation comparing the albumin-associated peptides in serum study sets derived from patients with pathologically documented ovarian cancer vs healthy but high-risk patients who had been disease-free for 5 years after serum collection.

Two key properties of carrier proteins enable them to function as efficient harvesters of lower-abundance putative biomarkers in the circulation: overwhelming relative abundance and long half-lives compared with their cargo. Considering the binding reaction between a biomarker (B) and its carrier protein (C):

$$B + C \underset{k_r}{\overset{k_f}{\leftrightharpoons}} BC$$

it is reasonable to suppose that the reaction dynamics are governed by mass action kinetics because a carrier protein would not usually bind more than one molecule of the biomarker of interest if the carrier existed in high relative abundance.

Thus, the rate of formation of the biomarker–carrier-protein complex (BC) is proportional to the product of the biomarker concentration [B] and the carrier-protein concentration [C], with a constant of proportionality k_f. The dissociation of the biomarker–carrier-protein complex into free biomarker and carrier protein, on the other hand, is proportional to the concentration of the bound biomarker [BC], with a constant of proportionality k_r. The forward and reverse reaction rates are given by:

$$R_f = k_f\,[B][C]$$

and

$$R_r = k_r\,[BC]$$

respectively. The vast excess of carrier protein means that only a tiny proportion of its total concentration will be consumed in the reaction with the biomarker, so that the free carrier-protein concentration [C] may be considered a constant. The free carrier-protein concentration may then be absorbed into the forward rate constant to yield an amplified effective forward rate constant, k_f^*, giving rise to the forward reaction rate:

$$R_f = k_f^*\,[B]$$

where

$$k_f^* = k_f\,[C]$$

The reaction arrives at an equilibrium when $R_f = R_r$, or:

$$k_f^*\,[B] = k_r\,[BC]$$

Importantly, because the effective forward rate constant k_f^* has been amplified by the (very large) carrier-protein concentration, it is expected to be many orders of magnitude larger than the reverse rate constant k_r, so that the reaction between the biomarker and its carrier protein is strongly biased in the forward direction. The equilibrium

[6] Nonstandard abbreviations: MS, mass spectrometry; μLC-MS/MS microcapillary reversed-phase tandem MS; and PVDF, polyvinylidene difluoride.

ratio between the bound and free forms of the biomarker is given by:

$$\rho_{eq} = \frac{[BC]}{[B]} = \frac{k_f^*}{k_r}$$

Thus, the putative biomarker is likely to exist almost entirely in complexed association with a carrier protein, even when its innate binding affinity, characterized by the "natural" forward rate constant, k_f, is low *(3)*. Indeed, with an albumin concentration in the blood of 6×10^{-4} mol/L *(19)*, molecules with low-affinity binding constants, k_f/k_r, of 15×10^4 L/mol [reported for flavonoid interaction with albumin *(20)*] and 9.34×10^6 L/mol [reported for interactions between protoporphyrins and albumin *(21)*] correspond to 98.9% and 99.98% probability, respectively, that the molecule of interest will be complexed with albumin.

Moreover, the relatively long half-life of the carrier protein [~19 days for albumin compared with a few hours or less for an unbound molecule *(9, 22–25)*] facilitates the prolonged presence of smaller peptides and proteins in the bloodstream. Mehta et al. *(3)* showed that if a protein is secreted into the bloodstream at a constant rate p, and is subsequently eliminated through the kidneys at a rate proportional to its concentration (with a constant of proportionality of e), its steady-state concentration in the circulation is given by the ratio p/e. In the absence of carrier-protein binding, the steady-state concentration of biomarker in the circulation would be p_b/e_b, where p_b and e_b are the production and elimination rates of the biomarker, respectively. When bound to a carrier protein, however, the biomarker acquires the carrier protein's longer half-life, so that the steady-state concentration becomes p_b/e_c, where e_c is the relatively small elimination rate of the carrier protein. Because the elimination rate of a protein is inversely proportional to its half-life, amplification of the biomarker's concentration in the bloodstream as a result of carrier-protein binding is $e_b/e_c = t_{(1/2)c}/t_{(1/2)b}$, where $t_{(1/2)c}$ and $t_{(1/2)b}$ are the half-lives of the carrier protein and biomarker, respectively. Thus, the longer the half-life of the carrier protein, the greater is its ability to amplify the concentration of bound species in the bloodstream. In this way, the relative abundance of large carrier proteins and their longer half-lives work together to increase the total concentration of complexed molecules in the bloodstream. This report describes the application of the carrier-protein sequestration principle to the isolation and mass spectrometric identification of candidate low–molecular-weight proteins and peptides that exist in the serum of ovarian cancer patient disease study sets.

Materials and Methods

CLINICAL SERUM SAMPLES

Serum samples were collected under full patient consent and Institutional Review Board approval. Serum was collected before physical evaluation, diagnosis, and treatment and stored at −80 °C. The ovarian study set consisted of 40 samples from unaffected, high-risk patient, 30 samples from patients with stage I ovarian cancer, and 40 samples from patients with stage III ovarian cancer. The gynecologic oncology clinic where the samples from the cases were collected is located in a building that is separate but contiguously linked to the building in which the National Ovarian Cancer Early Detection Program at Northwestern University is located, where the high-risk control samples were collected. A special attribute of this sample set is that the same personnel were involved in the blood collection, handling, and storage of all biospecimens. In addition, all blood specimens were processed in an identical manner under the same methodology. Healthy control sera were collected from unaffected women determined to be at increased risk for ovarian cancer. These women were enrolled in the National Ovarian Cancer Early Detection Program and had no evidence of any cancer for 5 years as determined by twice yearly 3-dimensional color Doppler ultrasound and extensive clinical evaluation by a board-certified gynecologic oncologist.

Increased risk was determined by classic genetic pedigree analysis and by the presence or absence of factors such as BRCA mutation status. The determination of increased risk was based on commercially available computer-generated risk algorithm programs such as BRCA Pro. All patients in the high-risk clinic were seen by board-certified genetic counselors and geneticists who defined the risk. Similarly, the serum specimens from women with ovarian cancer were procured in a gynecologic oncology clinic from symptomatic women who were later surgically staged and found to have epithelial ovarian carcinoma. Each sample was accompanied by a verified pathology diagnosis. Briefly, specimens were collected in red-top Vacutainer Tubes and allowed to clot for 1 h on ice, followed by centrifugation at 4 °C for 10 min at 1210*g*. The serum supernatant was divided in aliquots and stored at −80 °C until needed. Samples were selected for our analysis by a random process categorized by the pathology diagnosis as cancer or benign.

CAPTURE OF NATIVE SERUM ALBUMIN AND ELUTION OF COMPLEXED PROTEIN SPECIES

A schematic representation of the basic experimental technique is illustrated in Fig. 1. Native, diluted serum is introduced into an affinity column so that the carrier protein (albumin) is captured along with any bound molecules. The bound subproteome consisting of the carrier proteins and their peptide "cargo" is eluted, dissociated, and separated by 1-dimensional gel electrophoresis. Each entire gel lane is cut out, finely subdivided into molecular mass regions, subjected to in-gel trypsin digestion, and prepared for electrospray mass spectrometric analysis.

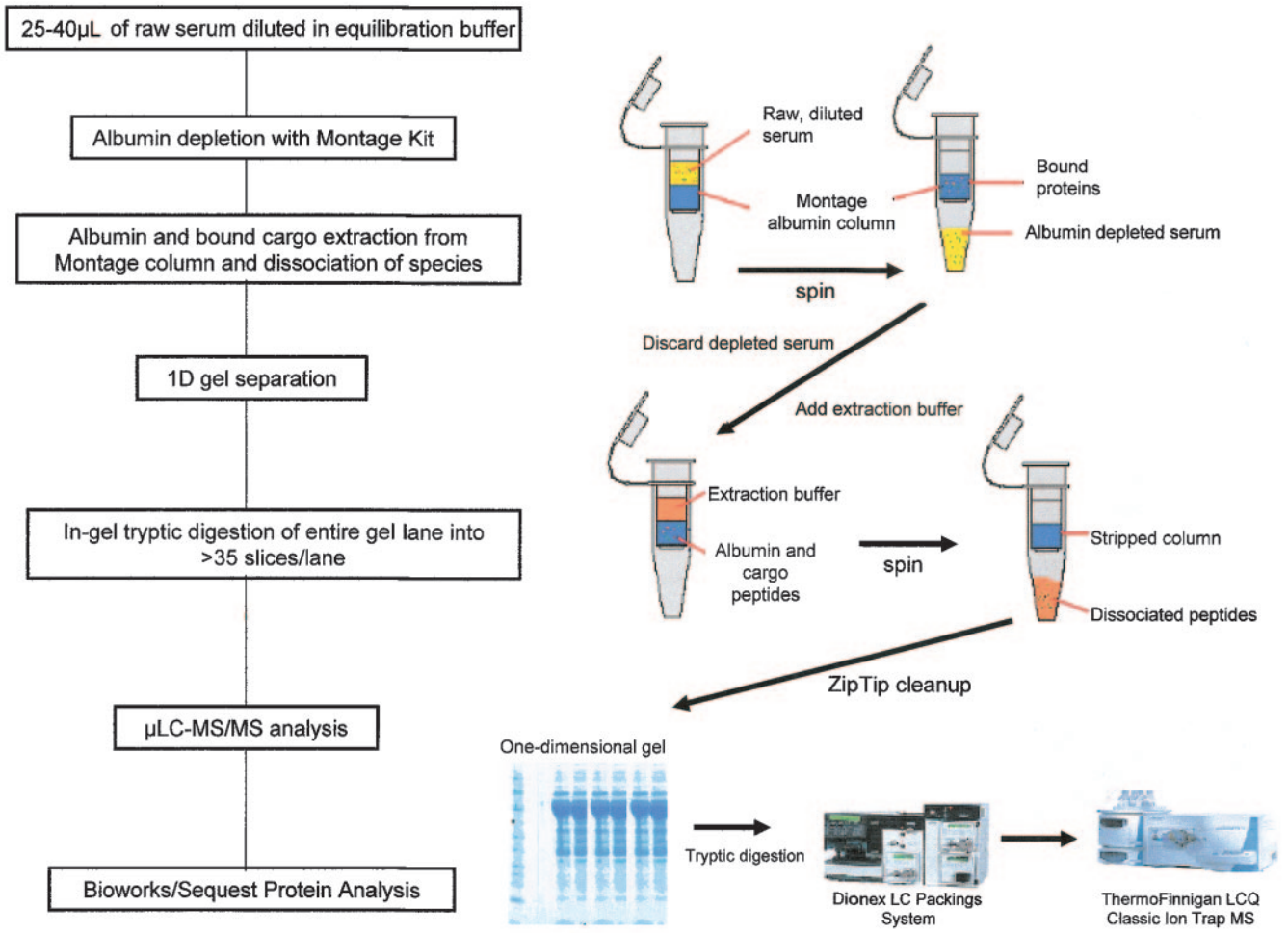

Fig. 1. Schematic representation of the experimental technique.

Raw serum is introduced into a Millipore albumin depletion column. Albumin complexed with other proteins and peptides is retained. After gentle washing under native conditions, the protein and peptide cargo are dissociated with organic solvent. Eluted proteins are fractionated by 1-dimensional (*1D*) gel electrophoresis followed by in-gel trypsin digestion and subjected to μLC-MS/MS analysis.

PURIFICATION OF ALBUMIN AND BOUND PEPTIDE

Typically, 25 μL of human stage-specific (pooled) cancer serum (~3.1 mg of protein) was diluted to 200 μL with Equilibration Buffer (Millipore) and run through a (Montage) albumin-specific affinity column twice. The bound protein was washed thoroughly with two 200-μL volumes of proprietary wash buffer (provided by the manufacturer). These fractions were combined and labeled as a "flow-through" fraction. The bound proteins were eluted from the column by equilibrating with acetonitrile–H_2O–trifluoroacetic acid (70:30:0.2 by volume) for 30 min, followed by a slow spin-through of the elution mixture, repeated once. The eluate (retentate fraction) was lyophilized to <10 μL in a HetoVac roto (CT 110) and reconstituted in an H_2O–acetonitrile–formic acid (95:5:0.1 by volume) buffer. Samples were desalted with a ZipTip cleanup and reconstituted in a 1:1 mixture of water and sodium dodecyl sulfate sample buffer (20 μL total volume).

1-DIMENSIONAL PROTEIN GEL SEPARATION AND DIGESTION

The flow-through and retentate fractions were kept on ice in 20 μL of sample buffer from 25 μL of original serum, and then were heated for 5 min at 95 °C and loaded on 1-dimensional precast gels to separate albumin from the proteins/peptides/fragments of interest. The proteins and fragments were visualized with a Gel Code Blue Stain Reagent (Pierce) according to the manufacturer's protocols. The entire lane was excised from the gel and finely sliced into very small molecular-weight regions (~35 slices/lane). Gel bands were reduced, alkylated, and digested with porcine modified trypsin according to a

standard protocol *(26)*, and peptides were concentrated and prepped for mass spectrometric analysis.

MICROCAPILLARY REVERSED-PHASE TANDEM MS

Samples were lyophilized to near dryness and reconstituted in 6.3 μL of Buffer A (H_2O–acetonitrile–formic acid; 95:5:0.5 by volume) for MS analysis. Microcapillary reversed-phase tandem MS (μLC-MS/MS) analysis was performed with a Dionex LC Packings liquid chromatography system coupled on-line to a ThermoFinnigan LCQ Classic ion trap mass spectrometer with a modified nanospray source. Reversed-phase separations were performed on an in-house, slurry-packed capillary column. The C_{18} silica-bonded column was a 10-cm long (75-μm i.d.) fused-silica column packed with 5-μm beads (pore size, 300 Å; Vydac). A PepMap C_{18} cartridge (5-mm; Dionex) acted as a desalting column. Sample was injected in microliter pick-up mode and washed with Buffer A for 5 min before elution with a linear gradient with buffer B (acetonitrile–H_2O–formic acid; 95:5:0.1 by volume) up to 85% over 95 min at a flow rate of 200 nL/min. Full MS scans were followed by 4 MS/MS scans of the most abundant peptide ions (in a data-dependent mode), and collision-induced dissociation was performed at a collision energy of 38% with the ion spray voltage set to 1.80 kV, capillary voltage set to 22.80 V and temperature set to 180 °C.

DATA ANALYSIS AND REPETITIVE SEQUENCING

Data analysis was performed by searching MS/MS spectra against the European Bioinformatics Institute of the nonredundant proteome set of Swiss-Prot, TrEMBL, and Ensembl entries through the Sequest Bioworks Browser (ThermoFinnigan), with a static modification of +57 Da on cysteine residues and a dynamic modification for oxidation of methionine of +15.9994 Da. Peptides were considered legitimate hits after the correlation scores (refer below) were filtered and the MS/MS data were manually inspected. The criteria used to filter data in this report are at least as stringent as most literature citations *(17, 18, 27–30)*:

Charge	X_{corr}	ΔCn	Rsp
+1	>1.9	>0.1	=1
+2	>2.2	>0.1	=1
+3	>3.5	>0.1	=1

Accepted peptide hits were required to have an X_{corr} ranking = 1 relative to all other peptides in the database. The albumin extraction, gel electrophoresis, protein digestion/extraction, and μLC-MS/MS analysis were repeated in 5 subsequent trials, each time yielding diminishing returns of new identifications for low-abundance peptide hits. Repetitive sequencing of peptides in multiple trials was an additional means to validate the reproducibility of our experimental procedure, both within and between cancer stages.

VALIDATION BY SERUM WESTERN BLOTTING

The primary antibody that recognized BRCA2 was synthesized in house. Rabbits were immunized with a peptide corresponding to an exact antigenic region of BRCA2, and the resulting polyclonal anti-BRCA2 antibody was affinity-purified (see below). The specificity of the antibody was verified against the full-length (390 kDa) BRCA2 protein extracted from HeLa cell nuclear extract. Subsequent preincubation of the primary antibody with an immunizing synthetic peptide overlapping the antigenic region of interest successfully competed away the representative band of native BRCA2 at 390 kDa. After verification of the specificities of the antibody and competition peptide, this experimental procedure was applied to pooled ovarian cancer and control serum samples.

Prepared serum samples were heated for 5 min at 95 °C in sample buffer containing 20 mL/L β-mercaptoethanol, followed by centrifugation at 10 000*g* for 1 min to remove insoluble material. Samples were then subjected to 1-dimensional electrophoresis and electroblotting at 30 V for 2 h on ice. Membranes were incubated overnight at 4 °C in 50 g/L nonfat dry milk, 75 g/L glycine, and 1 mL/L Tween 20 in water to block unoccupied protein binding sites.

The blocked membranes were rinsed twice with wash buffer [10 mmol/L Tris (pH 7.5), 150 mmol/L NaCl, 1 g/L bovine serum albumin, 1 mL/L Tween 20] and then incubated with 1 mg/L primary antibody in wash buffer containing 50 g/L nonfat dry milk, with rocking, for 2 h at room temperature. For peptide blocking/competition assays, 10 μg of primary antibody was incubated with 100 μg of the corresponding immunization peptide in 400 μL of wash buffer for 1 h at room temperature with end-over-end mixing. The peptide-treated antibody solution was diluted to 10 mL (1 mg/L final antibody concentration) in wash buffer containing 50 g/L nonfat dry milk before incubation with polyvinylidene difluoride (PVDF) membrane.

The membranes were washed 5 times (3 min each) in 50 mL of wash buffer and subsequently incubated in 10 mL of horseradish peroxidase–conjugated goat anti-rabbit IgG (1:50 000 in wash buffer) for 1 h at room temperature. After the PVDF membranes were washed thoroughly, signals were developed by enhanced chemiluminescence.

PEPTIDE-SPECIFIC ANTIBODIES

A peptide representing amino acid residues 980–993 (DKIPEKNNDYMNKW) of the BRCA2 sequence was synthesized (Anaspec) and conjugated to keyhole limpet hemocyanin for immunization as described previously *(31)*. The resulting antisera were affinity-purified over columns of peptides conjugated to Affigel 15 (Bio-Rad) and concentrated in stirred cells with YM-30 membranes (Millipore). The concentrates were subjected to gel-filtration chromatography on 2.6 × 60 cm Superdex 200 columns (GE Healthcare) in phosphate-buffered saline,

and the monomeric IgG fractions were pooled and concentrated. The protein concentrations were determined by the Bradford assay (Bio-Rad).

Results

The ovarian cancer study set was divided according to disease category; serum samples within each disease category were pooled into sets. A total of 110 samples were classified based on pathology into high-risk (n = 40), stage I (n = 30), and stage III (n = 40) ovarian cancer pools, and 5 separate aliquots per disease stage were iteratively sequenced by the experimental procedure in Fig. 1. A total of 1208 unique proteins were predicted in all 3 pools, 446 of these from multiple peptide sequences. An iterative sequencing approach examines the repetitive yield and variability between runs and between stage classifications. The aggregate yield of low-abundance protein identifications is expected to increase with repeated iterations of the experimental method. The correlation between the number of sequencing iterations performed and the total number of peptide sequences, and corresponding protein identifications obtained are described in Table 1. Overall, the number of unique proteins identified by multiple peptide sequences increased at a diminishing rate relative to the number of iterations performed. The rates of single and multiple peptide hits accumulated vs the number of experimental iterations are shown in Fig. 2. The ability to identify new peptides with multiple hits began to diminish by the third iteration (Table 1 and Fig. 2); however, the total number of new peptide identifications and single hit identifications continued to increase even after 5 iterations. Previous work by Liu et al. *(32)* revealed that saturation of new peptide identification occurs at around the 10th iteration. The results presented here support the conclusion that greater coverage of lower-abundance proteins can be achieved by increasing the number of experiments performed on a given sample.

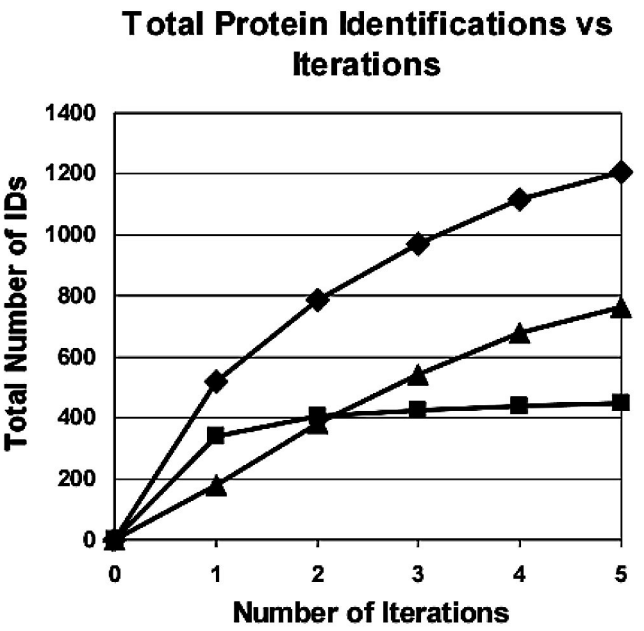

Fig. 2. Number of protein identifications and peptide sequences found relative to the number of iterations of the experimental procedure.

The recurring yield of peptide identifications (*IDs*) is plotted vs the number of independent experimental sample iterations. Each successive iteration represents a separate, new aliquot of the serum pool. Multipeptide identifications reached a plateau earlier in the iteration cycle. ♦, total identifications; ■, multiple peptide identifications; ▲, single peptide identifications.

Table 1. Representation of the total number of peptide hits and protein identifications acquired after subsequent iterations of the experimental method for ovarian cancer stage-specific and normal sera.

Ovarian cancer pool	No. of patient samples in pool	No. of iterations	No. of unique protein IDs[b]	No. of peptides sequenced	No. of multipeptide protein IDs	No. of single peptide protein IDs
High risk	40	5	359	935	147	212
		4	334	908	144	190
		3	282	838	140	142
		2	235	720	133	102
		1	164	565	106	58
Stage I	30	5	453	1027	149	304
		4	425	998	148	277
		3	386	921	142	244
		2	298	732	132	166
		1	168	594	111	57
Stage III-IV	40	5	396	896	150	246
		4	358	848	148	210
		3	301	761	144	157
		2	253	694	140	113
		1	185	560	123	62

[a] The number of patient samples pooled for each group is indicated. We performed 5 iterations for each category and report the total number of peptides sequenced correlating to the total number of unique proteins identified. The protein identifications are classified as single or multiple peptide hits.

[b] ID, identification. Identified proteins are listed in the online Data Supplement.

We have identified >700 different tryptic peptides derived from proteins not previously reported to exist in the blood in published databases from sera from women with various stages of ovarian cancer; many of these proteins are of low abundance (see Tables 1a–1c of the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol51/issue10/). More than 100 proteins falling into putative functional categories (Fig. 3) previously known to be related to cancer were identified by single or multiple peptide hits. In this study, unique single peptide hits were often discovered more than once from the same disease category (Table 1b of the online Data Supplement). That is, the same single peptide hit was generated more than once from different aliquots of the same disease category.

Low-abundance proteins or peptides derived from tissues entering the serum compartment can become complexed with high abundance proteins *(3, 9–17, 32)*. At least one half of all the proteins identified, which were bound to albumin and thus sequenced by the present method, must exist as peptide fragments of whole proteins. This is supported by two points of evidence: (*a*) the peptides were sequenced from a molecular-weight region of a gel that did not correspond to the predicted mass of the intact protein; and (*b*) passive diffusion of protein species through the vascular walls is hindered above apparent molecular masses of 60 kDa *(25)*. Thus, large tissue proteins can be represented in the blood circulation only if they are actively secreted or if they are in vivo cleavage products of the parent protein. BRCA2, which was sequenced by LC-MS/MS and represented by 4 distinct peptides (example BRCA2 spectra from 2 iteratively found peptides are presented in Fig. 4) and subsequently validated by competition Western blotting (Fig. 5C), is not predicted to be in the blood circulation as an intact protein. In fact, the parent protein with a known molecular mass of 390 kDa was not found by serum Western blotting (Fig. 5D). Shown in Fig. 5C are immunocompetition results for 2 cleavage products of BRCA2 (~12 and 25 kDa) that were identified by Western blot analysis of pooled ovarian cancer sera. A rabbit was immunized with a synthetic peptide representing amino acid residues 980–993 (DKIPEKNNDYMNKW) of the BRCA2 protein to produce a polyclonal antibody that ultimately detected the BRCA2 fragments in serum. Competition with the cognate peptide completely extinguished the 12- and 25-kDa bands. The estimated masses of the immuno-fragments (12 and 25 kDa) represent <9% by mass of the total parent protein. Furthermore, 3 other additional and distinct peptides were sequenced (residues 1942–1959, 2390–2401, and 2447–2459) in a separate region of the parent protein. The close proximity of the peptides containing amino acids 2390–2401 and 2447–2459 suggests that they likely originate from a single serum fragment of BRCA2 (Fig. 5, A and B).

Candidate proteins bound to albumin (Table 1; see also

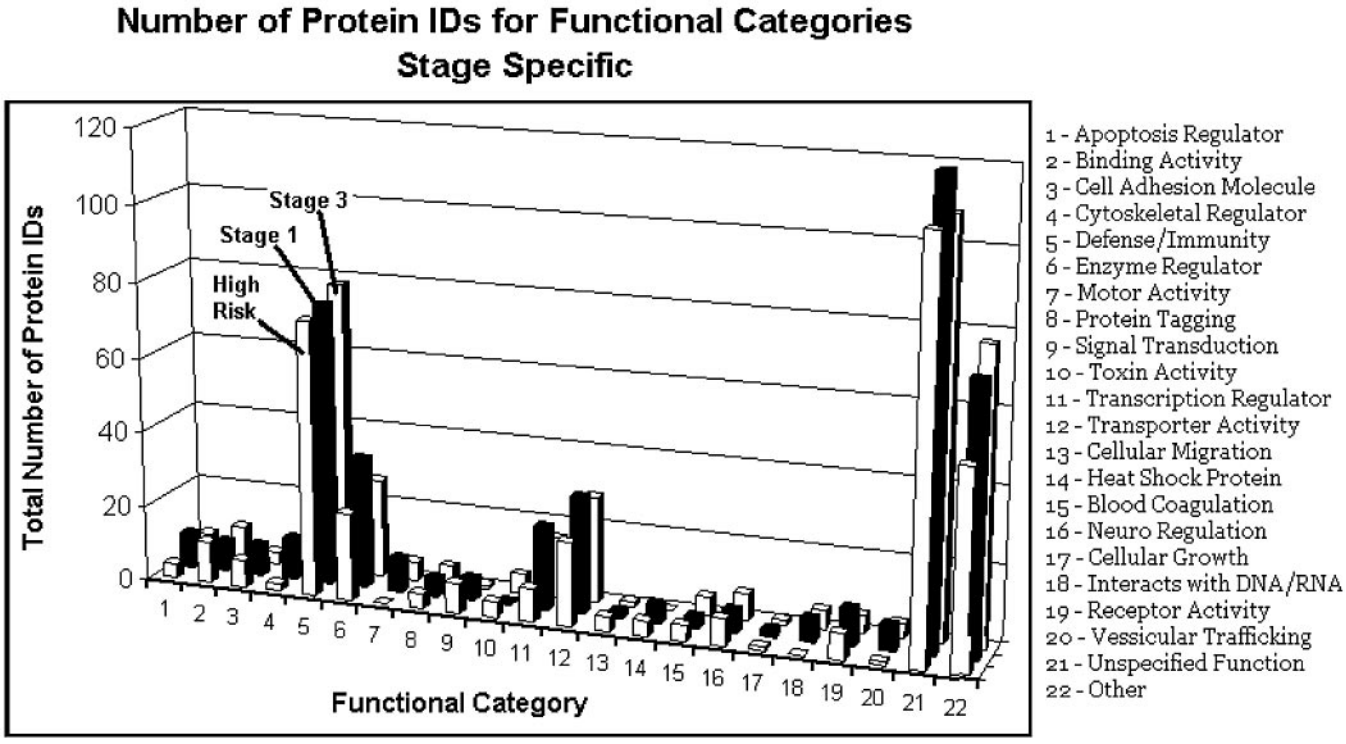

Fig. 3. Functional categories of the predicted albumin-associated proteins identified for each of the stage-specific ovarian cancer and high-risk control serum pools.

Serum pools are designated as high risk, stage I (*stage 1*), and stage III-IV (*stage 3*). *IDs*, identifications.

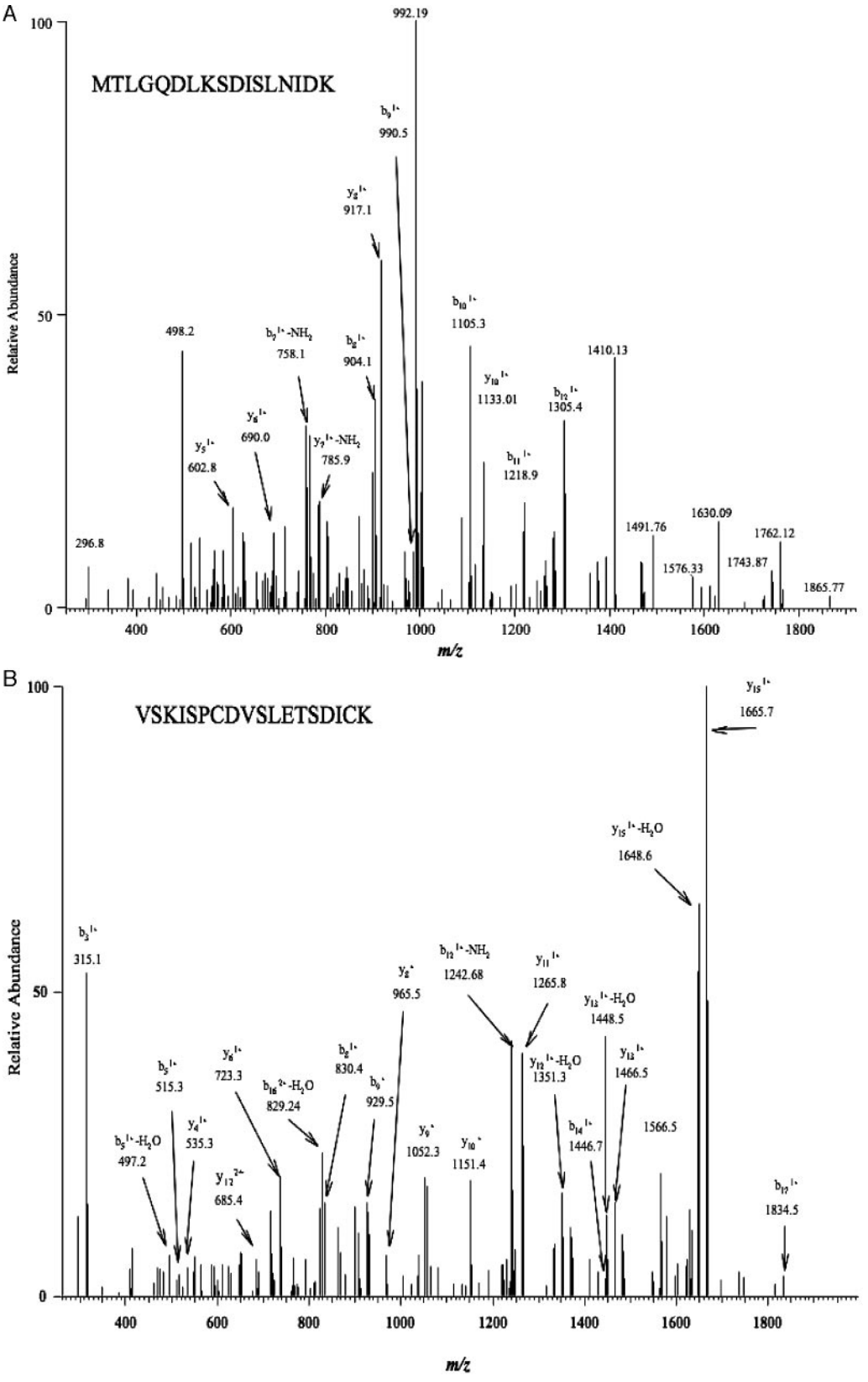

Fig. 4. Representative MS/MS spectra of BRCA2 peptides.

(*A*), result for the BRCA2 peptide identified as MTLGQDLKSDISLNIDK. (*B*), result for the BRCA2 peptide identified as VSKISPCDVSLETSDICK.

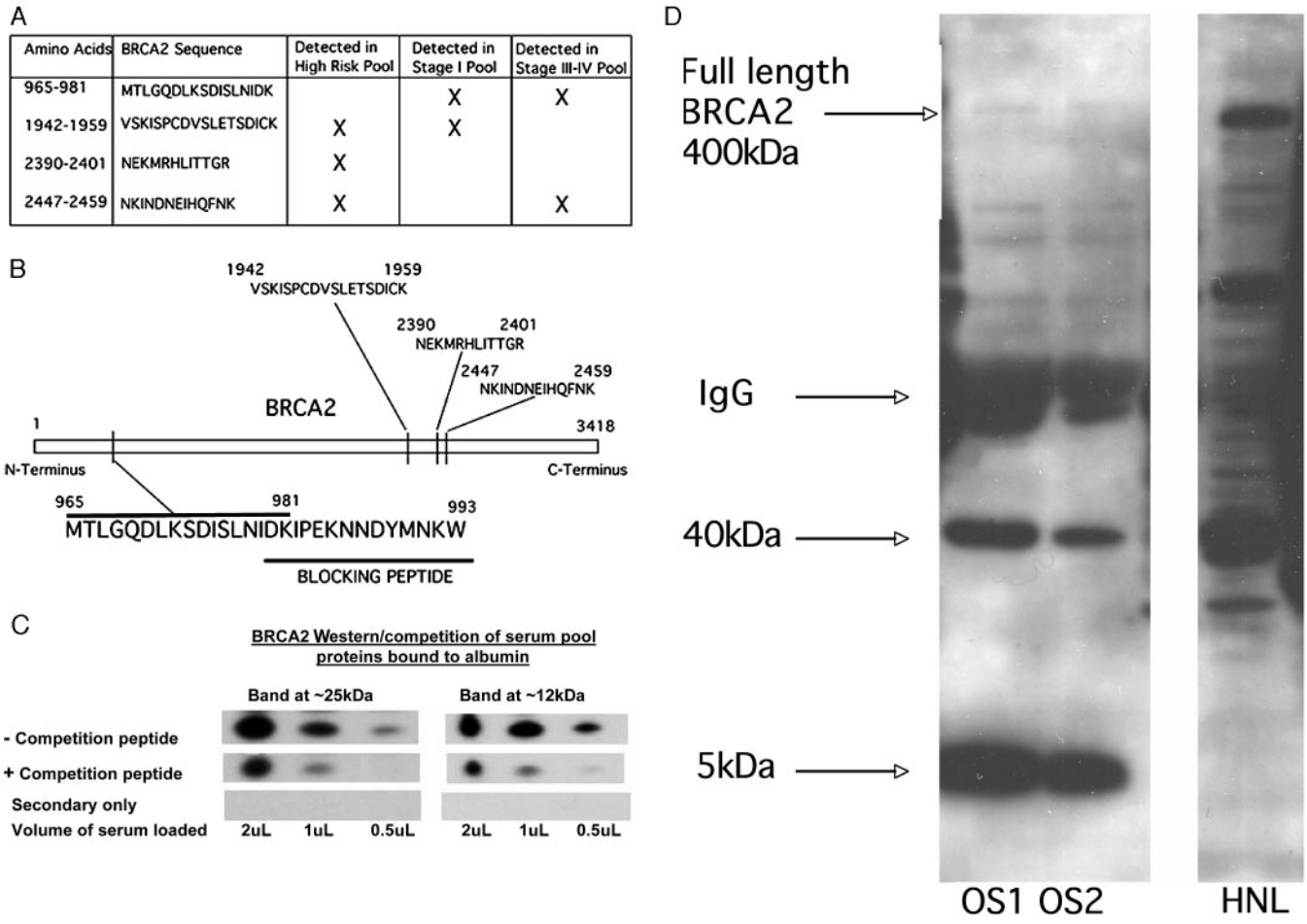

Fig. 5. Confirmation of the presence of BRCA2 peptide in serum.

(*A*), locations of the 4 peptide sequences of BRCA2 found by MS and the serum pool population origins. (*B*), synthetic peptide corresponding to the amino acid sequence overlapping with the sequence found by MS was used as a blocking agent in immunocompetition Western blotting. All sequenced peptides are shown in a spatial orientation within native BRCA2. (*C*), peptide identification by MS was confirmed by competition Western blot analysis using the peptide shown in *B*. Western blots of pooled ovarian cancer are displayed in a dilution curve. *Top row*, standard Western blot using a polyclonal anti-rabbit primary antibody synthesized in house corresponding to the antigenic region of BRCA2 found by MS. *Middle row*, the primary antibody is first incubated with a competition peptide specific for the identical amino acid region of BRCA2 before Western blotting. *Bottom row*, negative control/background of the goat anti-rabbit secondary. The BRCA2 peptides highlighted are evident at ~12 and 25 kDa. The indication of peptide fragments in Western blots of serum is expected because the molecular mass of the parent BRCA2 protein is 390 kDa, an apparent mass much larger than what is known to be able to passively enter the blood stream. (*D*), Western blot using polyclonal anti-BRCA 2 antibodies showing 2 example ovarian cancer sera (*OS1* and *OS2*) compared with total nuclear lysate of HeLa cells (*HNL*). Absence of the full-length BRCA2 protein in the sera is shown.

Tables 1a–1c of the online Data Supplement) are apparently diverse in origin and span a wide variety of predicted physiologic functions (Fig. 3). The breadth of biological and functional classes ranges from defense and immunity to transcription regulation, apoptosis, and cell adhesion. More than one half of the predicted proteins that have been sequenced have no known function reported in the literature. The relative abundances of the different functional classes of those proteins found in ovarian cancer samples are shown in Fig. 3. The diversity of categories represented is characteristic of a typical serum proteome *(23, 26)*. The ovarian cancer serum proteins that have been identified are listed in Tables 1a–1c of the online Data Supplement, and are shown classified as either multiple-peptide or single-peptide identifications. A sequence for CA125, a known ovarian cancer marker *(33, 34)*, was predicted to exist in the stage III ovarian cancer pool, but it was not found in either the stage I or the high-risk serum pools.

Discussion

The data presented here indicate that high-abundance serum carrier proteins such as albumin may act to sequester low–molecular-weight peptide fragments in the blood. Such sequestered peptides may provide a potentially rich source of candidate disease-associated biomarkers for subsequent clinical validation and provide a new opportunity to expand the knowledge base for the molecular composition of the circulation. Although the present work emphasizes the use of albumin, nonalbumin carrier proteins, such as IgG or fibronectin, are associated with a distinct subproteome within human sera *(23)*. The

method of albumin capture, elution of bound peptide and protein fragments, fractionation, digestion, and MS sequencing was repeated on 5 independent aliquots of each disease category serum pool. Because shown in Table 1, this iterative-sequencing approach provided a means to assess sequence discovery reproducibility within the same serum pool sample. Iterative sample sequencing indicated that after 5 rounds of iterations, the yield of novel multiple-peptide predictions not identified in previous iterations appeared to increase more slowly. This would be expected if the peptide sequences are being randomly resampled from the same large population of limited original candidates.

In theory, a similar list of protein fragments could be generated by denaturing and size-fractionating of sera because albumin fractionation and isolation would occur first. This is, in fact, exactly what previous investigators did in their analyses *(17, 18)*. However, because many current biomarker efforts begin by depletion of native albumin (as well as depletion of other high-abundance proteins such as immunoglobulins), the findings of this study begin to more fully describe the molecular information that is being lost by this method of sample preparation. Moreover, direct analysis of the albumin-bound material allows for clearer understanding of the nature and existence of the carrier-protein–bound low–molecular-weight information archive while minimizing issues of small amounts of free/unbound material contaminating the analysis.

The total list of identified albumin-bound entities are shown in Tables 1a–1c of the online Data Supplement (http://www.clinchem.org/content/vol51/issue10/). This list reveals a rich and previously undescribed information archive of putative analytes, many of which had not been described as being represented in the circulation. These molecules are mostly fragments of larger proteins; thus, their validation will necessitate the development of new immunodetection reagents such as anti-peptide antibodies.

It is well known that highly abundant proteins are identified by a larger number of peptides than are lower-abundance proteins. The probability of selecting and identifying a peptide from a low-abundance, and potentially more interesting, protein is therefore much less. The inherent complexity and large dynamic range of protein concentrations of global proteome samples represents a considerable barrier to obtaining multiple peptide identifications for each protein. The large number of peptides present in such mixtures greatly exceeds the capacity of current data-dependent tandem mass spectrometers, even when multidimensional fractionation is used *(4, 35, 36)*. In the present method we have endeavored to reduce this barrier by isolating the major high-abundance protein albumin and then examining the identity of its bound species.

Predicted sequences according to the criteria outlined in the *Materials and Methods* section fell into 2 categories: (*a*) multiple peptide hits, and (*b*) single peptide hits. Multiple peptide hits according to disease pool category are presented in Table 1a of the online Data Supplement by protein identity. Single peptide hits are listed in Table 1b and 1c of the online Data Supplement, along with the corresponding specific predicted sequences. Low-abundance proteins in blood are statistically likely to be discovered by only a single peptide sequence as a function of the dynamic range of proteins in serum; consequently, their probability of detection is low. Furthermore, many of the predicted serum species are peptide fragments of larger parent proteins. This can be concluded because a high percentage of the predicted proteins in Tables 1a–1c of the online Data Supplement have a full-length mass larger than albumin. Peptide fragments have reduced numbers of trypsin domains compared with the intact parent protein. Thus, a single peptide prediction, if correct, can correspond to a low-abundance peptide or a protein fragment with a reduced number of trypsin domains compared with the parent molecule. Because we conducted 5 independent isolation and sequence determinations for each disease category, single hits that were discovered more than once could be tabulated (Table 1b of the online Data Supplement). Single peptide hits discovered more than once may have a higher probability of being valid serum biomarkers than do single-instance single peptide hits.

The relevance of single peptide hits has been addressed recently *(37)*. These investigators pointed out that the elimination of single peptide hits from experimental results is scientifically unwarranted and detrimental to the field in general *(37)*. In fact, when so-called "one-hit wonders" identified by isotope-coded affinity tag proteomics were later assessed for validation by immunoassay detection, >90% were found to retain the differential expression state determined by the initial MS-based assay. Moreover, in the present study, many of the same single-hit tryptic fragments were iteratively and reproducibly found in the same disease category. Finally, for small, low-abundance fragments (with a low number of trypsin cleavage sites), it is statistically likely that only a single hit would be obtained in any given experimental cycle.

Although we identified 4 separate fragments for BRCA2, which increased our confidence in the finding, the fact that a majority of our protein identifications were based on one-hit MS analysis does not diminish their potential for potential validation. It is well known that highly abundant proteins are identified by a larger number of peptides than are lower-abundance proteins. The probability of selecting and identifying a peptide from a low abundance, and potentially more interesting, protein is therefore much less. The inherent complexity and large dynamic range of protein concentrations of global proteome samples represents a considerable barrier to obtaining multiple-peptide identifications for each protein. The large number of peptides present in such mixtures greatly exceeds the capacity of current data-dependent tandem mass spectrometers, even when multidimensional frac-

tionation is used. Moreover, global proteomic studies have consistently shown a low false-positive rate (i.e., <5%) for peptide identification; therefore; for every incorrect one-hit wonder that is thrown out, ~19 correct identifications are also excluded. Excluding one-hit wonders from possible validation would decimate quantitative proteomics studies that use isotope-coded affinity tag reagents and phosphoproteomic experiments, which typically produce only single phosphopeptide identifications. Many investigations have performed orthogonal validation experiments (e.g., Western analysis) that have confirmed that these qualitative and quantitative proteomic results forthcoming from one-hit wonders are indeed valid identifications *(27, 38–42)*.

Ultimately all predicted peptide sequences, whether single or multiple, must remain designated as "candidate" components of the human serum until they are confirmed immunologically. Many of the predicted sequences were derived from molecular-weight fractions smaller than the predicted size of the protein. We therefore concluded that the sequenced species existed as a fragment of a larger parent molecule. Thus, the gel migration location of the band could not be used to identify the protein. Western blotting with peptide competition was performed on the selected fragments identified by MS as belonging to BRCA2 to verify the identity of selected mass spectrometric identifications (Fig. 5C).

BRCA2 is a well-studied tumor suppressor protein related to the p53 pathway that is directly implicated in familial breast cancer and ovarian cancer *(43–46)*. Although the role of BRCA2 in breast and ovarian cancer predisposition is poorly understood, it is known that mutations of the *BRCA2* gene are responsible for one third of hereditary breast cancer cases *(45)*. The data in Fig. 5 suggest that at least 2 fragments of the BRCA2 parent protein exist in blood. A single predicted trypsin-cleaved peptide at amino acids 965–981 was identified in both stage I and stage III ovarian cancer serum pools (Fig. 4A and Fig. 5A). This peptide is represented in competition Western blots by 2 distinct molecular-weight bands (Fig. 5C). It is likely that a fragment containing the peptide at amino acids 965–981 would be cleaved at various residues in vivo and could therefore be represented at multiple molecular weights. This peptide was not predicted in pooled high-risk ovarian cancer serum by MS analysis. On the basis of evidence from the Western blots using a BRCA2 polyclonal antibody and the fact that BRCA2 is too large (390 kDa) to enter the blood circulation in its native form, we conclude that BRCA2 can be represented in the serum as one or more fragments and that at least 2 amino-terminal peptide fragments of BRCA2 <25 kDa can be validated by peptide competition for antibodies that recognize an amino acid sequence adjacent to, and overlapping with, the predicted peptide sequence. Additionally, 2 separate BRCA2 fragments encompassing no less than the amino acid region from 2390–2459 and 1942–1959 (Fig. 4B) are predicted to exist in the serum from mass spectral evidence of multiple peptide sequences identified in this region of BRCA2.

The current methodology applies carrier-protein sequestering to a disease category pool of combined sera. The methodology described here appears to be a successful means to identify candidate proteins and peptides in the sera of patients with known disease states. On the basis of the predicted physiologic functions of the parental protein containing the peptide sequence, these proteins and peptides can potentially be derived from tumor cells, from the host microenvironment, or from interactions between these two tissue compartments. Each pool contained multiple serum specimens procured before pathology-based diagnosis. The rationale for using a pool of multiple serum samples is based on the conservative assumption that even within the same histopathologic diagnosis, cancer is a heterogeneous disease. Thus, a biologically relevant molecule may be expressed only within a subfraction of the cancer population. Pooling samples is statistically the best means to identify a list of candidate peptides and proteins that exist within subsets of the pooled population.

The first disadvantage of using a pool of sera to discover markers is that molecules present in a subset of the combined samples are diluted within the pool. Although an individual sample may have a high concentration of the putative protein or peptide, this sample is diluted within the entire pool. Consequently, this averaging effect will lower the concentration of any individual molecule in the pool before sequence analysis. Lastly, a candidate protein that is identified in one disease pool (e.g., high risk) but not found in a second disease category (e.g., stage I cancer) may still exist in the second category at a concentration below a threshold for statistical sampling probability. Currently we are performing competition Western blots on a large set of serum samples from individual patients diagnosed as either high risk or with stage I, III, or IV ovarian cancer to determine the extent by which pooling samples limits the effectiveness of MS sequencing detection. Despite these drawbacks, as shown in Table 1 and in Tables 1a–1c of the online Data Supplement, carrier-protein sequestration appears to yield a list of predicted sequences derived from proteins predicted in the literature to have diverse physiologic functions.

The predicted sequences presented in Tables 1a–1c of the online Data Supplement demonstrate a rich diversity of peptides and protein fragments in the sera. Although the application of this method yielded different sets of predicted sequences within the control and diseased category pools, these proteins and peptides cannot be considered as candidate diagnostic markers. Once high-throughput means are developed to quantitatively measure each protein and peptide and accurately distinguish specific biomarker fragments from their parent molecules, proper clinical trial studies can examine the diagnostic sensitivity and specificity of a selected panel of molecules.

In conclusion, proteins and peptides associated with serum carrier proteins such as albumin may constitute a rich source of new additions to the human serum proteome. The sequences identified by this means are predicted to be derived from proteins that have been shown in the literature to serve a wide variety of physiologic functions. We can hypothesize that these complexed proteins and peptides may have originated from a range of different tissues and cellular compartments. A large proportion of the predicted sequences represent fragments of larger molecules. A subset of these candidates may eventually be found suitable for full clinical diagnostic testing in adequately powered objective study sets.

References

1. Ardekani AM, Liotta LA, Petricoin EF 3rd. Clinical potential of proteomics in the diagnosis of ovarian cancer. Expert Rev Mol Diagn 2002;2:312–20.
2. Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 2002; 1:845–67.
3. Mehta AI, Ross S, Lowenthal MS, Fusaro V, Fishman DA, Petricoin EF 3rd, et al. Biomarker amplification by serum carrier protein binding. Dis Markers 2003;19:1–10.
4. Adkins JN, Varnum SM, Auberry KJ, Moore RJ, Angell NH, Smith RD, et al. Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry. Mol Cell Proteomics 2002;1:947–55.
5. Pieper R, Su Q, Gatlin CL, Huang ST, Anderson NL, Steiner S. Multi-component immunoaffinity subtraction chromatography: an innovative step towards a comprehensive survey of the human plasma proteome. Proteomics 2003;3:422–32.
6. Rothemund DL, Locke VL, Liew A, Thomas TM, Wasinger V, Rylatt, DB. Depletion of the highly abundant protein albumin from human plasma using the Gradiflow. Proteomics 2003;3:279–87.
7. Wang YY, Cheng P, Chan DW. A simple affinity spin tube filter method for removing high-abundant common proteins or enriching low-abundant biomarkers for serum proteomic analysis. Proteomics 2003;3:243–8.
8. Pieper R, Gatlin CL, Makusky AJ, Russo PS, Schatz CR, Miller SS, et al. The human serum proteome: display of nearly 3700 chromatographically separated protein spots on two-dimensional electrophoresis gels and identification of 325 distinct proteins. Proteomics 2003;3:1345–64.
9. Dennis MS, Zhang M, Meng YG, Kadkhodayan M, Kirchhofer D, Combs D, et al. Albumin binding as a general strategy for improving the pharmacokinetics of proteins. J Biol Chem 2002; 277:35035–43.
10. Lee VHL. Peptide and protein drug delivery. New York: Marcel Dekker, 1990:1–56.
11. Makrides SC, Nygren PA, Andrews B, Ford PJ, Evans KS, Hayman EG, et al. Extended in vivo half-life of human soluble complement receptor type 1 fused to a serum albumin-binding receptor. J Pharmacol Exp Ther 1996;277:534–42.
12. Kurtzhals P, Havelund S, Jonassen I, Kiehr B, Larsen UD, Ribel U, et al. Albumin binding of insulins acylated with fatty acids: characterization of the ligand-protein interaction and correlation between binding affinity and timing of the insulin effect in vivo. Biochem J 1995;15:725–31.
13. Markussen J, Havelund S, Kurtzhals P, Andersen AS, Halstrom J, Hasselager E, et al. Soluble, fatty acid acylated insulins bind to albumin and show protracted action in pigs. Diabetologia 1996; 39:281–8.
14. Gordon LM, Curtain CC, McCloyn V, Kirkpatrick A, Mobley PW, et al. The amino-terminal peptide of HIV-1 gp41 interacts with human serum albumin. AIDS Res Hum Retroviruses 1993;9:1145–56.
15. Sjobring U, Falkenberg C, Nielsen E, Akerstrom B, Bjorck L. Isolation and characterization of a 14-kDa albumin-binding fragment of streptococcal protein G. J Immunol 1988;140:1595–9.
16. Raghu P, Sivakumar B. Interactions amongst plasma retinol-binding protein, transthyretin, and their ligands: implications in vitamin A homeostasis and transthyretin amyloidosis. Biochim Biophys Acta 2004;1703:1–9.
17. Tirumalai RS, Chan KC, Prieto DA, Issaq HJ, Conrads TP, Veenstra TD. Characterization of the low molecular weight human serum proteome. Mol Cell Proteomics 2003;2:1096–103.
18. Zhou M, Lucas DA, Chan KC, Issaq HJ, Petricoin EF 3rd, Liotta LA, et al. An investigation into the human serum "interactome". Electrophoresis 2004;25:1289–98.
19. Corot C, Violas X, Robert P, Gagneur G, Port M. Comparison of different types of blood pool agents (P792, MS325, USPIO) in a rabbit MR angiography-like protocol. Invest Radiol 2003;38: 311–9.
20. Dufour C, Dangles O. Flavonoid-serum albumin complexation: determination of binding constants and binding sites by fluorescence spectroscopy. Biochim Biophys Acta 1721:164–73, 2005.
21. Ding Y, Lin B, Huie CW. Binding studies of porphyrins to human serum albumin using affinity capillary electrophoresis. Electrophoresis 2001;22:2210–6.
22. Klasco RK, ed. United States pharmacopoeia DI, Vol. 1. Drug information for the health care professional. United States Pharmacopeial Convention, Inc. Greenwood Village, CO: Thomson Micromedex, 2000.
23. Peters T Jr. Serum albumin. In: Putnam FW, ed. The plasma proteins: structure, function, and genetic control, Vol. 1, 2nd ed. New York: Academic Press, 1975:133–81.
24. Rainey TG, Read CA. Pharmacology of colloids and crystalloids, the pharmacologic approach to the critically ill patient, Vol. 1, 3rd ed. Baltimore: Williams & Wilkins, 1994:272–90.
25. Wagner BK, D'Amelio LF. Pharmacologic and clinical considerations in selecting crystalloid, colloidal, and oxygen-carrying resuscitation fluids, part 1. Clin Pharm 1993;12:335–46.
26. Kinter M, Sherman NE. Protein sequencing and identification using tandem mass spectrometry. New York: Wiley-Interscience, Inc., 2000:301pp + xvi.
27. Peng J, Elias JE, Thoreen CC, Licklider LJ, Gygi SP. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J Proteome Res 2003;2:43–50.
28. Qian WJ, Liu T, Monroe ME, Strittmatter EF, Jacobs JM, Kangas LJ, et al. Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome. J Proteome Res 2005;4:53–62.
29. Wilmarth PA, Riviere MA, Rustvold DL, Lauten JD, Madden TE, David LL. Two-dimensional liquid chromatography study of the human whole saliva proteome. J Proteome Res 2004;3:1017–23.
30. Xiang R, Shi Y, Dillon DA, Negin B, Horvath C, Wilkins JA. 2D LC/MS analysis of membrane proteins from breast cancer cell lines MCF7 and BT474. J Proteome Res 2004;3:1278–83.
31. Goldsmith P, Backlund PS Jr, Rossiter K, Carter A, Milligan G, Unson CG, et al. Purification of heterotrimeric GTP-binding proteins from brain: identification of a novel form of Go. Biochemistry 1988;27:7085–90.
32. Liu H, Sadygov RG, Yates JR. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem 2004;76:4193–201.
33. Moss EL, Hollingworth J, Reynolds TM. The role of CA125 in clinical practice. J Clin Pathol 2005;58:308–12.

34. Topley N, Michael D, Bowen T. CA125: Holy Grail or a poisoned chalice. Nephron Clin Pract 2005;100:c52–c54.

35. Washburn MP, Wolters D, Yates JR 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol 2001;19:242–7.

36. Yu LR, Conrads TP, Uo T, Morrison RS, Chan K, Lucas DA, et al. Global analysis of the cortical neuron proteome. Mol Cell Proteomics 2004;3:896–907.

37. Veenstra TD, Conrads TP, Issaq HJ. What to do with "one-hit wonders"? Electrophoresis 2004;25:1278–9.

38. Johnson MD, Yu L-R, Conrads TP, Kinoshito Y, Uo T, Lee S-W, et al. Proteome analysis of DNA damage-induced neuronal death using high throughput mass spectrometry. J Biol Chem 2004;279: 26685–97.

39. Conrads KA, Yi M, Simpson KA, Stevens R, Lucas DA, Veenstra TD, et al. A combined proteome and microarray investigation of inorganic phosphate-induced pre-osteoblast cells. Mol Cell Proteomics 2005;Jun 14 [Epub ahead of print].

40. Ballif BA, Villén J, Beausoleil SA, Schwartz D, Gygi SP. Phosphoproteomic analysis of the developing mouse brain. Mol Cell Proteomics 2004;3:1093–101.

41. Hardwidge PR, Rodriguez-Escudero I, Goode D, Donohoe S, Eng J, Goodlett DR, et al. Proteomic analysis of the intestinal epithelial cell response to enteropathogenic *Escherichia coli*. J Biol Chem 2004;279:20127–36.

42. Anderson L. Candidate-based proteomics in the search for biomarkers of cardiovascular disease, J Physiol 2005;563:23–60.

43. Jonkers J, Meuwissen R, van der Gulden H, Petersen H, van der Valk M, Berns A. Synergistic tumor suppressor activity of BRCA2 and p53 in a conditional mouse model for breast cancer. Nat Genet 2001;29:418–25.

44. Esashi F, Christ N, Gannon J, Liu Y, Hunt T, Jasin M, et al. CDK-dependent phosphorylation of BRCA2 as a regulatory mechanism for recombinational repair. Nature 2005;434:598–604.

45. Ford D, Easton DF, Stratton M, Narod S, Goldgar D, Devilee P, et al. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The breast cancer linkage consortium. Am J Hum Genet 1998;62:676–89.

46. Hussain S, Witt E, Huber P, Medhurst A, Ashworth A, Mathew C. Direct interaction of the Fanconi anaemia protein FANCG with BRCA2/FANCD1. Hum Mol Genet 2003;12:2503–10.