Identification of Serum Amyloid A as a Biomarker to Distinguish Prostate Cancer Patients with Bone Lesions

LYLY LE,1 KIM CHI,1 SCOTT TYLDESLEY,1 STEPHANE FLIBOTTE,1 DEBORAH L. DIAMOND,2 MICHAEL A. KUZYK,1 and MARIANNE D. SADAR1*

Background: Prostate cancer has a propensity to metastasize to the bone. Currently, there are no curative treatments for this stage of the disease. Sensitive biomarkers that can be monitored in the blood to indicate the presence or development of bone metastases and/or response to therapies are lacking. Surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS) is an affinity-based approach that allows sensitive and high-throughput protein profiling and screening of biological samples.

Methods: We used SELDI-TOF MS for protein profiling of sera from prostate cancer patients (n = 38) with and without bone metastases in our effort to identify individual or multiple serum markers that may be of added benefit to those in current use. Serum was applied to ProteinChip® surfaces (H4 and IMAC) to quickly screen samples and detect peaks predominating in the samples obtained from patients with bone metastases. Unique proteins in the bone metastasis cohort observed by SELDI-TOF MS were identified by two-dimensional gel electrophoresis, in-gel trypsin digestion, and tandem MS. The identities of the proteins were confirmed by ELISA and immunodepletion assays.

Results: The cluster of unique proteins in the sera of patients with bone metastases was identified as isoforms of serum amyloid A. Machine-learning algorithms were also used to identify patients with bone metastases with a sensitivity and specificity of 89.5%.

Conclusions: SELDI-TOF MS protein profiling in combination with other proteomic approaches may provide diagnostic tools with potential clinical applications and serve as tools to aid in the discovery of biomarkers associated with various diseases.

Prostate cancer is the second-leading cause of death from cancer in American males (1). Hormonal therapy, radiotherapy, and chemotherapy for metastatic disease have limited durations of efficacy, and men diagnosed with metastatic disease succumb over a period of several months to a few years (2). The predominant site of metastases for hormone-refractory prostate cancer is bone, occurring in >90% of patients, with <20% of patients having clinically demonstrable soft tissue involvement (3,4). Bone metastases tend to be solely blastic (bone forming) in an individual, but ~20% of cases exhibit a combination of blastic and lytic (bone destroying) lesions. Lytic lesions occur in 15–30% of patients (5). Bone metastases are associated with a poor prognosis and with the development of hormone-refractory disease, for which there is currently no effective treatment (6–9). Once the disease is hormone refractory, the typical survival time is <2 years.

Bone tissue is dynamically maintained by two tightly coupled processes of resorption and formation, which are controlled by osteoclasts and osteoblasts, respectively. Bone remodeling liberates potential biomarkers. Examples of biomarkers for resorption are associated with collagen degradation, such as the carboxy-terminal telopeptides of type I collagen, pyridinoline, and deoxypyridinoline. Markers for bone formation, or osteogenesis, include the carboxy-terminal propeptide of type I collagen, which is cleaved from procollagen molecules during the formation of type I collagen; alkaline phosphatase; and osteocalcin. Bone formation involves three steps: The first is synthesis of collagen, which is associated with
increased proliferation and increased concentrations of carboxy-terminal propeptide of type I collagen. The second phase involves matrix maturation, at which time concentrations of alkaline phosphatase increase. The third phase involves mineralization, which is marked by increased concentrations of osteocalcin. Studies examining the potential usefulness of biochemical markers for the detection of bone metastasis have generally focused on markers for bone turnover in prostate cancer patients and have shown a lack of sensitivity compared with bone scintigraphy scans.

Serum concentrations of prostate-specific antigen (PSA) are useful in combination with a physical examination for the initial diagnosis of prostate cancer, but when used alone, PSA is associated with a high degree of false positives. Increased PSA concentrations in the serum of men with prostate cancer have been suggested to be directly correlated to the tumor volume (10, 11). One of the best applications of serum PSA measurements is for the monitoring of treatment response and relapse in prostate cancer patients. However, serum PSA concentrations cannot predict the presence of bone or soft tissue metastases (12). At the molecular level, PSA gene expression is regulated by the androgen receptor via androgen response elements (13). Immunohistochemical studies confirmed the presence of an androgen receptor in metastatic prostate cancer biopsies (14, 15).

Bone-derived factors such as interleukin-6 (IL-6) (16–20) and insulin-like growth factor-I (21) have been shown to stimulate androgen-responsive reporters or PSA gene expression in the absence of androgen [for a review, see Ref. (22)]. IL-6 is also increased in the sera of patients with metastatic prostate cancer (23–25) and hormone-refractory disease (26), and may be a surrogate marker for androgen-independent disease. Other potential markers of prostate cancer currently being investigated include cadherin (27), human kallikrein 2 (28), and chromogranin A (29).

Despite a growing list of potential markers of prostate cancer, sensitive assays to detect, and ultimately predict susceptibility to, metastases by use of a blood test are still urgently required. Serum concentrations of markers for bone turnover, such as alkaline phosphatase, become increased in response to systemic therapy (30), reducing the clinical application of this biomarker for the monitoring of prostate cancer patients, particularly in the case of those patients potentially receiving intermittent androgen suppression. Detection of bone lesions is generally made by clinical symptoms and bone scans, but the latter may not reflect response to therapy (31). Identification of a biomarker for bone or other metastatic lesions that could be measured concurrently with PSA in the blood of patients would enable more accurate clinical assessment of the stage of the disease and may also be of use in the stratification of patients with respect to anticipated or actual response to therapy where current methods lack reliability. We applied surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS) profiling to sera from prostate cancer patients with and without positive bone scans in our effort to identify individual or multiple serum markers that may be complementary to those in current use as well as to provide further insights into the mechanisms governing the apparent prostate/bone symbiosis. Here we show that rapid SELDI-TOF MS profiling, combined with MS identification of proteins that were isolated by two-dimensional polyacrylamide gel electrophoresis designed to resolve the observed cluster seen with SELDI-TOF MS, correlated with the bone metastases cohort. The cluster of proteins were identified as isoforms of serum amyloid A (SAA). Immunologic quantification validated a correlation of SAA within the bone metastasis group. Machine-learning algorithms were used to identify patients with bone metastases. To our knowledge, this is the first study that compares SELDI-TOF MS spectra, ELISA, and machine-learning algorithms to differentiate metastatic disease.

Materials and Methods

STUDY PARTICIPANTS

Serum samples from each group were collected simultaneously dependent on the patient’s availability in the clinic between July 2001 and May 2002. These volunteer patients had prostate cancer and were being treated at the Vancouver Cancer Centre of the British Columbia Cancer Agency in Vancouver, Canada. A 10-mL sample of blood was collected into a serum separator tube (red-top Vacutainer; BD-Canada), clotted for 30 min at room temperature, and centrifuged (10 min at 1000 g), and the supernatant serum was aliquotted (250 μL) and flash frozen at −80 °C. A total of 38 serum samples were collected from 19 patients considered not to have bone metastases and from 19 patients with confirmed bone metastases as indicated by a positive bone scan. The demographics of these patients are provided in Table 1 with the Gleason scores, initial and/or current PSA values, and serum alkaline phosphatase values. All patients had histologically confirmed adenocarcinoma of the prostate and gave signed informed consent. For patients with PSA >10 μg/L, Gleason score of 7, or T3 disease, a bone scan was performed to confirm no bone metastases. No clinical symptoms were present to suggest bone involvement, and/or serum alkaline phosphatase was within the reference interval. All patients considered to have bone metastases were selected based on a positive bone scan. Exclusion criteria for both groups of patients included (a) any previous malignancy, unless the previous malignancy was diagnosed and definitively treated at least 5 years

Nonstandard abbreviations: PSA, prostate-specific antigen; IL-6, interleukin-6; SELDI-TOF MS, surface-enhanced laser desorption/ionization time-of-flight mass spectrometry; SAA, serum amyloid A; TFA, trifluoroacetic acid; SDS, sodium dodecyl sulfate; MALDI, matrix-assisted laser desorption/ionization; MS/MS, tandem mass spectrometry; SVM, support vector machine; and PMF, peptide mass fingerprinting.
and equilibrated with three 100-

reversed-phase surface. Spots were then washed twice

dry on the spot, thereby forcing an interaction with the

washed four times with 100

(50 mmol/L) for 5 min. The ProteinChip Arrays were then

H4 ProteinChip Arrays were preactivated with 2

ProteinChip Arrays (Ciphergen Biosystems, Inc.).

analyzed on both reversed-phase (hydrophobic surface)

dilation optimal with no statistically significant gain in informa-

tion obtained when binding times are increased up to 2 h

in length (data not shown).

Reversed-phase (acyl chains of 16 methylene groups)

H4 ProteinChip Arrays were preactivated with 2 μL of

acetonitrile (500 mL/L) in a humidity chamber for 5 min

at room temperature. The acetonitrile was removed, and

spots were allowed to briefly air dry before application of

dilated serum samples (2 μL). Samples were allowed to

dry on the spot, thereby forcing an interaction with the

reversed-phase surface. Spots were then washed twice

with 2 μL of acetonitrile (300 mL/L). A saturated solution

of sinapinic acid in acetonitrile (500 mL/L)–trifluoroacetic

acid (prepared as described previously) were added to

each spot and allowed to air dry before mass analysis

using a ProteinChip Reader (Ciphergen Biosystems).

Reversed-phase (acyl chains of 16 methylene groups)

H4 ProteinChip Arrays were preactivated with 2 μL of

acetonitrile (500 mL/L) in a humidity chamber for 5 min

at room temperature. The acetonitrile was removed, and

spots were allowed to briefly air dry before application of

diluted serum samples (2 μL). Samples were allowed to

dry on the spot, thereby forcing an interaction with the

reversed-phase surface. Spots were then washed twice

with 2 μL of acetonitrile (300 mL/L). A saturated solution

of sinapinic acid in acetonitrile (500 mL/L)–trifluoroacetic

acid (prepared as described previously) were added to

each spot and allowed to air dry before mass analysis

using a ProteinChip Reader (Ciphergen Biosystems).

The spots were washed with three 50-μL volumes of

phosphate-buffered saline to remove proteins interacting

diluted serum samples (20

μL) were denatured with 30 μL of 8 mol/L urea, 10 g/L

CHAPS and vortex-mixed for 15 min at 4 °C. To lower the

urea concentration, we added 100 μL of 1 mol/L urea, 1.25 g/L

CHAPS and vortex-mixed the sample. Finally, we added 600 μL of binding buffer (phosphate-buffered saline) to the denatured serum samples and loaded 50 μL

per spot. The sample was allowed to bind to the Protein-

Chip Array for 30 min with shaking at room temperature.

A 30-min binding time has previously been found to be

optimal with no statistically significant gain in informa-

tion obtained when binding times are increased up to 2 h

in length (data not shown).

The spots were washed with three 50-μL volumes of

phosphate-buffered saline to remove proteins interacting

nonspecifically with the IMAC-Cu

surface, followed by

five 50-μL washes with water to remove salts. Arrays

were then air-dried, and two 0.5-μL volumes of sinapinic

acid (prepared as described previously) were added to

each spot and allowed to air dry before mass analysis

using a ProteinChip Reader (Ciphergen Biosystems).

The ProteinChip reader, or linear SELDI-TOF MS, has

delayed extraction (time lag focusing) and a mass analysis

range from <0.1 to >500 kDa. The sensitivity is better

than 700:1 mean signal-to-noise ratio on analysis of 150

fmol of bovine IgG when digitally smoothed with a (0.2

times peak width) sliding window average. Mass accu-

racy with internal calibration is better than ± 200 ppm

(0.02%) at m/z 1000–10 000 and better than ± 500 ppm

(0.05%) at m/z 10 000–300 000. The mass accuracy with

external calibration is better than ± 1000 ppm (0.1%) at m/z

1000–10 000 and better than ± 2000 ppm (0.2%) at m/z

10 000–300 000. Spectra were generated by use, on aver-

age, of 90–105 laser shots using laser intensities of 215

with a sensitivity of 8 (IMAC-Cu

) or laser of 250 with a

sensitivity of 9 (H4). The instrument was externally cali-

brated by use of bovine ubiquitin, cytochrome C, super-

previously and there had never been any evidence of recurrence; and (b) history of significant medical or psychiatric disorder that would impair the ability to obtain informed consent. The study was approved by the research ethics board of the University of British Columbia.

SELDI-TOF MS ANALYSIS OF SERUM PROTEINS

Aliquots of serum samples were thawed on ice, and protein concentrations were measured using the bicineho-

linic acid assay (Pierce). All samples for MS analysis were
diluted to a final concentration of 1 g/L in HEPES (pH 7.4), 30 mmol/L NaCl. Diluted serum samples were
analyzed on both reversed-phase (hydrophobic surface)
H4 and Immobilized Metal Affinity Capture (IMAC) ProteinChip Arrays (Ciphergen Biosystems, Inc.).

Reversed-phase (acyl chains of 16 methylene groups)

H4 ProteinChip Arrays were preactivated with 2 μL of

acetonitrile (500 mL/L) in a humidity chamber for 5 min

at room temperature. The acetonitrile was removed, and

spots were allowed to briefly air dry before application of
diluted serum samples (2 μL). Samples were allowed to
dry on the spot, thereby forcing an interaction with the

reversed-phase surface. Spots were then washed twice

with 2 μL of acetonitrile (300 mL/L). A saturated solution

of sinapinic acid in acetonitrile (500 mL/L)–trifluoroacetic

acid (prepared as described previously) were added to

each spot and allowed to air dry before mass analysis

using a ProteinChip Reader (Ciphergen Biosystems).

The ProteinChip reader, or linear SELDI-TOF MS, has
delayed extraction (time lag focusing) and a mass analysis
range from <0.1 to >500 kDa. The sensitivity is better

than 700:1 mean signal-to-noise ratio on analysis of 150

fmol of bovine IgG when digitally smoothed with a (0.2

times peak width) sliding window average. Mass accu-
racy with internal calibration is better than ± 200 ppm

(0.02%) at m/z 1000–10 000 and better than ± 500 ppm

(0.05%) at m/z 10 000–300 000. The mass accuracy with

external calibration is better than ± 1000 ppm (0.1%) at m/z

1000–10 000 and better than ± 2000 ppm (0.2%) at m/z

10 000–300 000. Spectra were generated by use, on aver-
age, of 90–105 laser shots using laser intensities of 215

with a sensitivity of 8 (IMAC-Cu

) or laser of 250 with a

sensitivity of 9 (H4). The instrument was externally cali-
brated by use of bovine ubiquitin, cytochrome C, super-

Table 1. Demographics, Gleason scores, current and initial PSA concentrations, and alkaline phosphatase concentrations of the prostate cancer patients included in the study.

<table>
<thead>
<tr>
<th>Metastases</th>
<th>Range</th>
<th>Gleason score</th>
<th>PSAC, μg/L</th>
<th>PSAI, μg/L</th>
<th>ALP, U/L</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>52–73</td>
<td>5–10</td>
<td>2–2600</td>
<td>7–170</td>
</tr>
<tr>
<td>Median</td>
<td>61</td>
<td>7</td>
<td>185</td>
<td>93</td>
<td>178</td>
</tr>
<tr>
<td>Mean</td>
<td>62</td>
<td>7</td>
<td>367</td>
<td>283</td>
<td>310</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
<td>608</td>
<td>479</td>
<td>252</td>
</tr>
<tr>
<td>n</td>
<td>19</td>
<td>13</td>
<td>18</td>
<td>15</td>
<td>18</td>
</tr>
<tr>
<td>No metastases</td>
<td>Range</td>
<td>59–78</td>
<td>5–9</td>
<td>0.02–10</td>
<td>4.8–168.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.92</td>
<td>15</td>
<td>84</td>
</tr>
<tr>
<td>Median</td>
<td>71</td>
<td>7</td>
<td>2.02</td>
<td>26.7</td>
<td>86.5</td>
</tr>
<tr>
<td>Mean</td>
<td>70</td>
<td>7</td>
<td>3.09</td>
<td>37.3</td>
<td>32.7</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td></td>
<td>18</td>
<td>19</td>
<td>12</td>
</tr>
<tr>
<td>n</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>12</td>
</tr>
</tbody>
</table>

P

Welch Ttest

Wilcoxon

0.02098

1.02 × 10⁻⁶

0.05751

0.00072

0.00164

0.00161

a PSAC, current PSA; PSAI, initial PSA; ALP, alkaline phosphatase.

b Number of patients.
oxide dismutase, β-lactoglobulin, and horseradish peroxidase (Ciphergen Biosystems) with matrix application identical to that for the serum samples. Before data analysis, spectra were baseline subtracted by use of a modified piecewise convex-hull algorithm, which attempts find the bottom of the spectrum and correct the peak height and area, and normalized by use of total ion current.

**TWO-DIMENSIONAL GEL ELECTROPHORESIS**

Each serum sample (500 μg of protein) was solubilized in 8 mol/L urea, 40 mL/L CHAPS, 40 mmol/L Tris, 2 mL/L Bio-Lyte 3/10 ampholytes (Sequential Extraction Buffer 2; Bio-Rad) with 5 mmol/L tris(2-carboxyethyl)phosphine reducing agent and incubated for 1 h at room temperature. The samples were loaded on 7-cm immobilized pH gradient strips (ReadyStrip IPG strips; linear gradient, pH 3–10; Bio-Rad). Isoelectric focusing of samples (Bio-Rad Protein Cell) was carried out as follows: 250 V for 15 min, followed by a slow voltage ramp over 8000 V-h to 4000 V, with a maximum of 50 μA per strip. Voltage was held at 4000 V, with a maximum of 50 μA per strip, for 20 000 V-h, followed by a 500-V hold. Before the second dimension, the strips were reduced and alkylated for 20 min in 62.5 mmol/L Tris (pH 8.45), 5 mmol/L tris(2-carboxyethyl)phosphine, 25 g/L acrylamide, 200 mL/L glycerol, 20 g/L sodium dodecyl sulfate (SDS), 6 mol/L urea. The second dimension consisted of a 16.5% Separation Gel (4.5 cm), 10% spacer gel (1 cm), and a 4% stacking gel (32). The IPG strips were held in place with 1% agarose in 62.5 mmol/L Tris (pH 8.3), 500 mmol/L NaCl, 2 mL/L methanol. Western blotting was performed on 31 protein bodies were diluted in 50 g/L milk in Tris-buffered saline–Tween. All incubations were performed at room temperature unless otherwise stated. The SAA protein was detected by enhanced chemiluminescence (Amer sham).

**PSA, ALKALINE PHOSPHATASE, IL-6, AND SAA**

PSA was detected with the enzymatic immunoassay (Abbott IMx). Alkaline phosphatase was measured by an absorbance assay using the Dimension® clinical chemistry system (Dade Behring). SAA concentrations in diluted serum samples were measured in triplicate by ELISA (Biosource International), according to the manufacturer’s instructions. The detection limit of the assay is 5 μg/L, and all SAA isoforms were detected. IL-6 concentrations in serum were measured in triplicate by ELISA according to the manufacturer’s instructions (Biosource International). The detection limit of the assay is <2 ng/L. The assay detects both free and receptor-associated IL-6.

**WESTERN BLOT ANALYSIS AND IMMUNODEPLETION ASSAY**

A representative sample from each group was separated by 20% Tris-tricine-SDS–polyacrylamide electrophoresis and transferred to an Immobilon-P membrane (Millipore) in 25 mmol/L Tris, 192 mmol/L glycine, 1 mL/L Tween 20. All proteins bands plus a blank region from the edge of the gel were excised from two-dimensional gels. Individual spots were washed several times for at least 2 h with 500 mL/L methanol, 50 mL/L acetic acid, reduced with 10 mmol/L dithiothreitol, and alkylated with 100 mmol/L iodoacetamide before in-gel trypsin digestion with 10 μL of 20 μg/L trypsin (Boehringer) overnight at 37 °C. Supernatants were then collected and gel pieces extracted with additional ammonium bicarbonate (10 μL of a 25 mmol/L solution) for 15 min, which was pooled with the original supernatant (sample). The samples were zip-tipped (Millipore uC18); washed with 1 mL/L TFA in water; eluted with 2 μL of 500 mL/L acetonitrile, 2,5-dihydroxybenzoic acid, 1 mL/L TFA onto a 100-well stainless steel plate (Applied Biosystems); and analyzed via matrix-assisted laser desorption/ionization (MALDI) on a QStar Pulsar I (Applied Biosystems/MDS SCIEX) equipped with a MALDI I ion source. The QStar was calibrated by use of two tandem MS (MS/MS) fragments of Glu-fibrinogen peptide (Sigma) with [M+H]+ masses of 175.1195 and 1570.6774. After manual inspection of the survey scan to prevent the selection of keratin and trypsin autolysis peaks, the most intense ions were selected for MS/MS fragmentation with an operator-controlled collision energy. All spectra were acquired in positive-ion mode. All product ion scans were acquired from 100 m/z to an upper range that included the peptide mass selected for MS/MS fragmentation. The most intense peptide peaks were selected for fragmentation, and analysis of the resulting data was carried out with the Mascot search engine.

A sample from each patient group was depleted of SAA by immunoprecipitation using the Seize Primary Immunoprecipitation Kit (Pierce Endogen). Each sample was immunodepleted with an 8 molar excess of anti-SAA (Biosource International) coupled to AminoLink Plus Coupling Gel (Pierce Endogen), according to manufacturer instructions. The eluate (depleted serum) and
washes were collected and subsequently analyzed by SELDI-TOF MS using a WCX₂ (weak cation-exchange surface) ProteinChip Array.

SUPPORT VECTOR MACHINE ANALYSIS
Spectra obtained from IMAC-Cu⁺⁺ ProteinChip Arrays were analyzed by mapping the raw (nonfiltered and non-baseline-subtracted) TOF spectra to mass spectra consisting of 16 384 channels with a mass calibration given by \( m/z = aC^2 \), where \( C \) is the channel number and \( a = 0.0001 \) Da. Masses above 26.8 kDa were not considered in the analysis. All mass spectra were then normalized to the same total area. An automated procedure to find and fit the significant peaks in the mass spectra was developed and implemented in the C language. The resulting open-source X Windows application, named Nonprofit (for nonlinear protein fit), should compile and run on any UNIX operating system. It is freely available and can be obtained by contacting one of the authors (sflibotte@bcgsc.ca). In brief, a spectrum representing the average of all samples was first divided in several sections, the width and position of each section being automatically selected by the software, by use of a heuristic approach to obtain the best possible fits in the following step. Each section was then fitted with the appropriate number of gaussian peaks superimposed on a quadratic background, the number of peaks producing the best fit being found with an iterative procedure. The positions, widths, and heights of the peaks were fitted at the same time as the background parameters. The relative peak width was fixed, which means that all peaks in a given section had the same width but that this width was allowed to vary in the fitting procedure. For each section, the result of the fit of the average spectrum was then used as a template to set the initial parameters for fitting all spectra in the data set, the number peaks and the relative positions of the peaks being fixed. This procedure assumed that all peaks to be fitted appeared in the average spectrum, which is a reasonable assumption because it is unlikely that a peak appearing in only a few spectra would be useful in a classification scheme. This procedure corrected for an imperfect gain matching between spectra because the absolute positions of the peaks were free to vary during the fit; only their relative positions were fixed. For each spectrum, the position and area of each gaussian peak were saved for subsequent analysis and data mining. Such a fitting procedure has the advantage of being able to extract meaningful abundance even for proteins associated with mass peaks that were not completely resolved. Examples of fits obtained with the Nonprofit software are shown in Fig. s1 of the Data Supplement that accompanies this article at http://www.clinchem.org/content/vol51/issue4/.

Support vector machines (SVMs) were used to determine whether machine-learning algorithms could differentiate prostate cancer patients with bone lesions from those without bone lesions. Peaks under \( m/z \) 975 were not fitted because many of them showed obvious signs of saturation. The automated procedure applied divided the spectra into 39 sections for fitting, and a total of 270 peaks were fitted for each spectrum, to give >10 000 gaussian peaks. With the help of the fitted peak areas, a classifier was built to distinguish between the spectra associated with the patients who tested positive for bone metastasis and those who did not. SVM was chosen because this learning scheme performed very well as a classifier in association with our peak-fitting procedure on larger SELDI-TOF MS data sets publicly available (data not shown). Furthermore, SVMs are known to perform well in the challenging situation in which the number of samples in the data set is not large compared with the number of attributes. For example, SVMs have been successfully used as a classification scheme in microarray experiments (34). In our case, the attributes were the peak areas. Suppose that \( n \) attributes are being considered to build a classifier; each sample (or instance in the machine-learning terminology) can then be represented by a point in an \( n \)-dimensional space. SVMs find the maximum margin hyperplane, which is the hyperplane separating the two classes of samples in this \( n \)-dimensional space while maximizing the distance between the hyperplane and the closest training point. With our fitting procedure on SELDI-TOF MS data, we found that it is not usually necessary to select the attributes to reduce the dimensionality before building a good classifier with SVM. We therefore do not have to select which type of attribute filter or wrapper would be the most efficient. In practice, we have used the sequential minimal optimization algorithm as implemented in the Weka machine-learning software (35). To maximize the use of this relatively small data set, we have assessed the performance of the classifier with a leave-one-out cross-validation scheme. In other words, 37 of the samples were used to build a classifier and predict the class of the remaining sample, and this procedure was repeated 38 times, leaving out each sample one at a time.

STATISTICAL ANALYSIS
For our purposes, specificity is defined as the ratio of the patients who do not have the protein peak (or above-normal concentrations) and do not have bone metastases to the total number of patients that do not have bone metastases. Sensitivity is defined as the ratio of the patients with bone metastases who had the biomarker (or above-normal concentrations) to the total number of patients with bone metastases. Comparisons between patients with and without bone metastases were performed with the Welch \( T \)-test and the nonparametric Wilcoxon test. The latter test does not assume that the variables follow gaussian distributions. Statistical significance was assumed when \( P \) was \( <0.05 \).

Univariate statistical analysis of SELDI-TOF MS peak masses and relative intensity values was conducted with the nonparametric Mann–Whitney \( U \)-test.
Results

Identification of a unique cluster of proteins in the sera from prostate cancer patients with bone metastases

We analyzed 38 serum samples from prostate cancer patients by SELDI-TOF MS on the same day, using reversed-phase H4 ProteinChip Arrays. The limit of detection of the ProteinChip surfaces has been determined to typically be in the low femtomole range with a linear response over 2–3 orders of magnitude \((36, 37)\). To determine the degree of peak intensity variation observed within our analyses, we analyzed technical replicates \((n = 8)\) of a representative sample from each cohort on H4 ProteinChip arrays (see Table s1 in the online Data Supplement). Mean CVs observed for peaks across the \(m/z\) 2500–150 000 range were <25% (Table 1 in the online Data Supplement), and peaks in the \(m/z\) 10 000–15 000 range of interest exhibited even less variation, with a mean CV of 20%. The H4 data collected for the entire 38 sample set are shown in Fig. 1A as a gel-view depiction of the spectra. This spectrum shows a cluster of proteins at \(\sim m/z\) 11 700 that are predominantly detected in samples from patients with bone metastases (lanes 27–38). This cluster was also detected in two samples obtained from patients not considered to have bone metastases (lanes 18 and 19). Enhanced analysis of this region of the spectra for lane 4 (without bone metastasis) and lane 37 (with bone metastasis) is shown in Fig. 1B. With the exception of lanes 18 and 19, sera from patients without bone metastases failed to have this cluster of peaks at \(m/z\) 11 700.

Statistical analysis of peak intensities of the H4 ProteinChip array spectra revealed that the differences observed of \(\sim 11.7\) kDa in sera of patients with bone metastases were statistically significantly increased \([m/z\) 11 488 \((P = 0.0020)\); \(m/z\) 11 537 \((P = 0.0004)\); \(m/z\) 11 639 \((P = 0.0022)\); \(m/z\) 11 680 \((P = 0.0051)\); Fig. 2]. A plot of the peak intensities of the neighboring statistically insignificant \((P = 0.3812)\) \(m/z\) 13 885 peak is included for reference. The dominant peak of the four peaks is \(\sim 11.7\) kDa. The patients with bone metastases who had this cluster had variable Gleason scores \((5–10)\), current PSA concentrations \((40–2600 \mu g/L)\), and alkaline phosphatase \((152–726 U/L)\). The Gleason score is the sum of the numbers assigned to the two glandular patterns that predominate in a biopsy of the prostate malignancy and represents the microscopic appearance of the tumor \((38)\). Tumors with Gleason scores of 2–4 are considered less aggressive than tumors assigned a Gleason score of 8–10. The two patients not considered to have bone metastases who had this cluster at \(\sim 11.7\) kDa (lanes 18 and 19) had Gleason scores of 7 and 9, respectively.

Identification of the cluster of proteins by two-dimensional gel electrophoresis and MS

Preparative two-dimensional electrophoresis was used to isolate the cluster of proteins observed at 11.7 kDa in the

Fig. 1. SELDI-TOF MS spectra of sera from prostate cancer patients. (A), gel view of SELDI-TOF MS spectra obtained for serum from prostate cancer patients with and without bone metastases with a reversed-phase H4 ProteinChip Array. SELDI spectra were normalized by total ion current normalization in this and all subsequent SELDI-TOF MS spectra presented. Lanes 1–19, patients without bone metastases; lanes 20–38, patients with bone metastases. Gleason score is shown at the right of the spectra. n, not available. Spectra were sorted according to SAA concentrations. (B), enlargement of the \(m/z\) 10 000–15 000 region to highlight the unique cluster of proteins at 11.7 kDa in the spectra for sera from a patient without bone metastasis (top) and a patient with bone metastasis (bottom) obtained with a reversed-phase H4 ProteinChip Array. Spectra in B correspond to lanes 4 and 37 panel A. ¥ indicates a cluster of peaks seen in both spectra.
SELDI-TOF MS spectra. Serum samples chosen for this experiment were from one patient not considered to have bone metastases and from one who did have bone metastasis. The samples correspond to lanes 8 and 37 in Fig. 1. The gels are shown in Fig. s2A of the online Data Supplement and exhibited four differences in the mass region ~11.7 kDa when we compared the serum from the patient with bone metastases with the serum obtained from the patient who did not have bone metastases. These protein spots, plus a blank region at the edge of the gel, were excised and subjected to an in-gel trypsin digestion protocol. The resulting trypsin fragments from each spot were analyzed in a QStar Pulsar I with the MALDI I ion source to identify the unknown proteins, which were identified to be variants of SAA. The amino acid sequence of unprocessed SAA is shown in Fig. s2B of the online Data Supplement, with the variable amino acids boxed. A MALDI-TOF mass spectrum of peptides derived from spot 4 of the two-dimensional gel is shown in Fig. s2C of the online Data Supplement. A Mascot search using the peptide mass fingerprinting (PMF) data indicated that four peptides matched with peptides for SAA, giving a sequence coverage of 40%. Significant confidence (P < 0.05) in the identity was indicated by a Mascot score as high as 80. Further confirmation of protein identification was provided by MS/MS fragmentation analysis, which provided a sequence tag for database searching. Identification of the m/z 2178 ion confirmed its identity as a SAA peptide with a Mascot score of 131 (Fig. s2D in the online Data Supplement). MS/MS analysis of m/z 2178 detected most b and y ions (Fig. s2D in the online Data Supplement). Combined with PMF data, all four spots from the two-dimensional gel yielded very high confidence hits with SAA variants.

**WESTERN BLOT ANALYSIS AND IMMUNODEPLETION ASSAYS**

A representative sample from each group was selected for Western blot analysis based on whether the SELDI spectrum was negative or positive for what we putatively identified as SAA proteins. For this analysis, the samples corresponded to lanes 8 and 37 of Fig. 1 (negative and positive for SAA, respectively). Western blot analysis with an antibody against SAA detected a single band at a molecular mass of ~12 kDa in sample 37, whereas nothing was detected in serum from a patient not considered to have bone metastases (Fig. 3). Only one band could be detected in serum from a patient with confirmed bone metastases by Western blot when using a 20% tricine gel. Although the antibody used recognizes all four isoforms of SAA, these isoforms could not be resolved by Western blot analysis.

To confirm that the peaks detected by SELDI-TOF MS (Fig. 1) were indeed isoforms of SAA, we next analyzed serum from a representative sample of each cohort (lanes 7 and 35 in Fig. 1) both before and after immunodepletion of SAA. Selection of samples was based on the same rationale as described above. The spectrum obtained for serum from the patient with bone metastases had the cluster of peaks at 11.7 kDa, with intensities ranging from ~5 to 13 (Fig. 4). After immunodepletion, the intensities for these peaks were all <1 (compare panels A and B in Fig. 4). The spectra obtained with serum from the patient
without bone metastases did not have the cluster of peaks at 11.7 kDa, either before or after immunodepletion of SAA (panels C and D in Fig. 4). A protein cluster at ~13.8 kDa was only slightly diminished after immunodepletion, indicating the specificity of the assay.

CONFFIRMATION OF INCREASED SAA CONCENTRATIONS IN SERA FROM PATIENTS WITH BONE METASTASES

Taken together, the results of the two-dimensional gel electrophoresis, Western blotting, and immunodepletion experiments described above demonstrate that the 11.7-kDa cluster of peaks in the SELDI-TOF MS mass spectra correspond to SAA. To validate that the serum SAA protein concentrations correlated with peak intensities in the SELDI-TOF MS spectra, we performed an ELISA for SAA in all serum samples and plotted the values vs peak intensity obtained by SELDI-TOF MS for each patient. This plot showed that peak intensity correlated to SAA values (Fig. 5).

CORRELATION OF SAA AND IL-6 IN SERUM OF PROSTATE CANCER PATIENTS

SAA has been reported to be increased by IL-6 (39). IL-6 has also been reported to be increased in patients with recurrent prostate cancer (26). We therefore measured IL-6 in these same serum samples and compared the results to the SAA concentrations detected (Table 2). The results (Fig. 6A) indicated that the two of the patients with bone metastases (lanes 37 and 38) and the highest SAA concentrations also have the highest IL-6 concentrations in this cohort (Fig. 6B). One patient considered not to have bone metastases (Fig. 6, lane 19) also had a very high IL-6 concentration when compared with others in this cohort. This particular patient also had the highest SAA concentrations and was the only patient to have PSA >100 μg/L.
at the time of diagnosis in the cohort with no bone metastases. The probability of metastatic disease increases with increasing PSA concentrations. No specific trend could be readily observed between SAA and IL-6; some patients with very high SAA concentrations did not have high IL-6 concentrations (e.g., Fig. 6, lane 35).

 Serum alkaline phosphatase is a marker for osteoblastic lesions (40); however, in patients with confirmed bone metastases, it was increased to above-normal concentrations (>200 U/L) in only 44% (8 of 18) of the samples. Serum concentrations of alkaline phosphatase were generally higher in patients with positive bone scans (Fig. 6C, lanes 20–38), but there was no obvious trend relating these concentrations to SAA (Fig. 6A), IL-6 (Fig. 6B), or PSA (Fig. 6, D and E; see Tables 1 and 2).

**Table 2. SAA and IL-6 in the sera of prostate cancer patients with and without bone metastases**

<table>
<thead>
<tr>
<th>Metastases</th>
<th>SAA, mg/L</th>
<th>IL-6, µg/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>No metastases</td>
<td>Range</td>
<td>5.3–986</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>227.63</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>304.69</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>19</td>
</tr>
<tr>
<td>No metastases</td>
<td>Range</td>
<td>1.0–317</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>40.36</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>92.68</td>
</tr>
<tr>
<td></td>
<td>n</td>
<td>19</td>
</tr>
<tr>
<td><strong>P</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Welch Test</td>
<td>0.01800</td>
<td>0.433</td>
</tr>
<tr>
<td>Wilcoxon</td>
<td>0.00011</td>
<td>0.212</td>
</tr>
</tbody>
</table>

**Discussion**

Noninvasive tools for the diagnosis, prognosis, and monitoring of cancer are required. Biomarkers that can be measured in the blood of patients provide a convenient, relatively noninvasive approach to diagnosing and monitoring patients. The ideal biomarker should have low false-positive and -negative rates and be sensitive, specific, quantifiable, and capable of detecting tumors at an early stage to allow appropriate therapy and prognostic reference and to predict the likelihood of recurrence (41, 42). Serum PSA is a good biomarker for monitoring progression of hormone-refractory prostate cancer, but it cannot distinguish bone metastases from local recurrence or soft tissue secondary lesions (12, 43). Absolute PSA concentrations in recurrent disease have not consistently been shown to be prognostic for survival or predictive of response. Serum alkaline phosphatase is routinely monitored as a marker for bone turnover, but it lacks sensitivity and is associated with a transient flare in response to systemic therapy such as androgen ablation (44). Application of genomic and proteomic high-throughput approaches holds the promise of identifying new biomarkers.

SVM

Using linear class boundaries and a leave-one-out cross-validation scheme with spectra obtained from IMAC surfaces, SVMs correctly classified 34 of 38 samples, for a classification accuracy of 89.5%. Two samples associated with a positive test for bone metastasis and two samples associated with a negative test were misclassified. This corresponds to a sensitivity and specificity of 89.5%. The analysis pipeline described was applied to data obtained from other ProteinChip surfaces, but the best classification accuracy was obtained with the IMAC data. The simultaneous use of all data from various ProteinChip surfaces did not improve classification accuracy. No obvious unique peaks for bone metastases samples could be manually detected with IMAC-Cu⁺⁺ ProteinChip Arrays.

**SVM**

Using linear class boundaries and a leave-one-out cross-validation scheme with spectra obtained from IMAC surfaces, SVMs correctly classified 34 of 38 samples, for a classification accuracy of 89.5%. Two samples associated with a positive test for bone metastasis and two samples associated with a negative test were misclassified. This corresponds to a sensitivity and specificity of 89.5%. The analysis pipeline described was applied to data obtained from other ProteinChip surfaces, but the best classification accuracy was obtained with the IMAC data. The simultaneous use of all data from various ProteinChip surfaces did not improve classification accuracy. No obvious unique peaks for bone metastases samples could be manually detected with IMAC-Cu⁺⁺ ProteinChip Arrays.

**Discussion**

Noninvasive tools for the diagnosis, prognosis, and monitoring of cancer are required. Biomarkers that can be measured in the blood of patients provide a convenient, relatively noninvasive approach to diagnosing and monitoring patients. The ideal biomarker should have low false-positive and -negative rates and be sensitive, specific, quantifiable, and capable of detecting tumors at an early stage to allow appropriate therapy and prognostic reference and to predict the likelihood of recurrence (41, 42). Serum PSA is a good biomarker for monitoring progression of hormone-refractory prostate cancer, but it cannot distinguish bone metastases from local recurrence or soft tissue secondary lesions (12, 43). Absolute PSA concentrations in recurrent disease have not consistently been shown to be prognostic for survival or predictive of response. Serum alkaline phosphatase is routinely monitored as a marker for bone turnover, but it lacks sensitivity and is associated with a transient flare in response to systemic therapy such as androgen ablation (44). Application of genomic and proteomic high-throughput approaches holds the promise of identifying new biomarkers.

SVM

Using linear class boundaries and a leave-one-out cross-validation scheme with spectra obtained from IMAC surfaces, SVMs correctly classified 34 of 38 samples, for a classification accuracy of 89.5%. Two samples associated with a positive test for bone metastasis and two samples associated with a negative test were misclassified. This corresponds to a sensitivity and specificity of 89.5%. The analysis pipeline described was applied to data obtained from other ProteinChip surfaces, but the best classification accuracy was obtained with the IMAC data. The simultaneous use of all data from various ProteinChip surfaces did not improve classification accuracy. No obvious unique peaks for bone metastases samples could be manually detected with IMAC-Cu⁺⁺ ProteinChip Arrays.
Fig. 6. Concentrations of biomarkers in serum of prostate cancer patients without (lanes 1–19) and with bone metastases (lanes 20–38). (A), SAA concentrations (mg/L) were measured by ELISA. Reference values are <7 mg/L (54) and values above 95.4 (131.4) mg/L have been reported for prostate cancer (55). Columns, mean; error bars, SE (n = 3). (B), serum IL-6 concentrations (ng/L) in prostate cancer patients. Serum IL-6 concentrations have been reported to be 1.91 (0.61) ng/L for healthy individuals, 7 (2.18) ng/L for patients with prostate cancer, and 40–60 ng/L for patients with bone metastases. (C), alkaline phosphatase (ALP; U/L) concentrations in serum. The reference values are <200 U/L. (D), PSA (µg/L) values at the time of sample collection. (E), PSA (µg/L) concentrations at the time of diagnosis. NA, not available. Samples were sorted according to SAA concentrations.
Table 3. Comparison of SELDI-TOF MS, ELISA, and SVM in the detection of bone metastasis in prostate cancer patients. 

<table>
<thead>
<tr>
<th>Analysis</th>
<th>No metastases</th>
<th>Metastases</th>
</tr>
</thead>
<tbody>
<tr>
<td>SELDI-TOF MS</td>
<td>10% (2/19)</td>
<td>58% (11/19)</td>
</tr>
<tr>
<td>ELISA</td>
<td>21% (4/19)</td>
<td>84% (16/19)</td>
</tr>
<tr>
<td>SVM</td>
<td>89.5% (34/38)</td>
<td></td>
</tr>
</tbody>
</table>

* Numbers indicate the percentage of patients in whom SELDI detected SAA peaks or ELISA values for SAA were above the reference interval in each cohort of patients. Values given for SVM indicate the percentage of correct assignments of patients into their respective cohorts.

prostatic tissue. There is an inherent lack of specificity for SAA; it is generally increased in patients with a broad spectrum of primary and disseminated diseases (54, 55). SAA has received considerable attention for being a potential marker for neoplastic activity, however, because of its inverse correlation with patient survival in spite of the lack of elucidation of its role in metastatic processes [for a review see Ref. (51)]. SAA is induced by IL-6 (39), and although IL-6 is reported to be increased in patients with metastatic prostate cancer, we were unable to confirm this finding in our samples (see Fig. 6, A and B), thereby negating a possible mechanistic role correlating IL-6 to increased serum concentrations of SAA in prostate cancer patients with bone metastases. In 1984, SAA was proposed for use in the monitoring of active prostate cancer that was not in remission (56). Since that initial publication, SAA has not been reported for monitoring of prostate cancer, probably because of its lack of specificity and the successful emergence of PSA as a biomarker for this disease.

Nevertheless, to our knowledge, the work presented here is the first to identify a biomarker for metastasis based on a combination of SELDI-TOF MS, two-dimensional gel electrophoresis, in-gel trypsin digestion, PMF, and MS/MS followed by validation by ELISA and immunodepletion assay, with the results obtained with the various approaches compared with results obtained with machine-learning algorithms. Results from these comparison are shown in Table 3 and indicate that SVM was the most sensitive approach for differentiating prostate cancer patients with bone metastases, with a sensitivity of 89.5% compared with 84% for ELISA and 58% for SELDI-TOF MS alone. Application of SVM involves the analysis of ~270 peaks representing peptides or proteins per spectrum. Thus, SVM analysis uses multiple proteins/peptides rather than just the single SAA cluster of isoforms, as are used with manual inspection of SELDI-TOF MS spectra or ELISA. These data support that measurement of multiple proteins or a panel of proteins can enhance the detection rate [for a review, see Ref. (57)]. Thus, SELDI-TOF MS protein profiling may provide a diagnostic tool with potential clinical applications as well as aid in the discovery of biomarkers associated with various diseases.

This work was supported by Health Canada (M.D.S) and Genome British Columbia (M.D.S.). We thank H.A. Hare, H. Adomat, and A. Garcia for technical assistance and the University of Victoria/Genome BC Proteomics Centre (Dr. R. Olafson) for MS analyses.

References


32. Schagger H, von Jagow G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem 1987;166:368–79.


