A Sequence-specific Exopeptidase Activity Test (SSEAT) for “Functional” Biomarker Discovery*\[S\]

Josep Villanueva\[‡\]¶§, Arpi Nazarian\[‡\], Kevin Lawlor\[‡\], San San Yi\[‡\], Richard J. Robbins¶¶, and Paul Tempst\[‡\]¶§

One form of functional proteomics entails profiling of genuine activities, as opposed to surrogates of activity or active “states,” in a complex biological matrix: for example, tracking enzyme-catalyzed changes, in real time, ranging from simple modifications to complex anabolic or catabolic reactions. Here we present a test to compare defined exoprotease activities within individual proteomes of two or more groups of biological samples. It tracks degradation of artificial substrates, under strictly controlled conditions, using semiautomated MALDI-TOF mass spectrometric analysis of the resulting patterns. Each fragment is quantitated by comparison with double labeled, non-degradable internal standards (all-D-amino acid peptides) spiked into the samples at the same time as the substrates to reflect adsorptive and processing-related losses. The full array of metabolites is then quantitated (coefficients of variation of 6.3–14.3% over five replicates) and subjected to multivariate statistical analysis. Using this approach, we tested serum samples of 48 metastatic thyroid cancer patients and 48 healthy controls, with selected peptide substrates taken from earlier standard peptidomics screens (i.e. the “discovery” phase), and obtained class predictions with 94% sensitivity and 90% specificity without prior feature selection (24 features). The test all but eliminates reproducibility problems related to sample collection, storage, and handling as well as to possible variability in endogenous peptide precursor levels because of hemostatic alterations in cancer patients. Molecular & Cellular Proteomics 7: 509–518, 2008.

In the current vernacular, the term proteomics most often stands for cataloguing large sets of proteins that may or may not include measurements of relative abundance and, on a rare occasion, absolute concentrations. One might describe it as the protein equivalent of microarray-based mRNA profiling. The favorite approach in most proteomics endeavors is iden-tity-based shotgun analysis that involves digesting (e.g. with trypsin) complex protein mixtures into peptides followed by a mass spectrometric readout, utilizing diverse platforms and varying degrees of technological sophistication (1, 2). In some variations on this theme, modified proteins or protein metabolites are the targets of analysis, giving rise to such specialties as phosphoproteomics and peptidomics, again focusing on identity, (relative) quantitation, the sites of modification and processing, or all of the above (3, 4). Even so, they still represent steady state chemical readouts of a particular type of macromolecules in a cell, tissue, or biological fluid. Modifications and some forms of processing influence protein activity, for instance in the case of many enzymes that are essential to the function of molecular machines, pathways, cells, and organisms (5). In fact, post-translational modifications and peptides are themselves the product of specific enzymatic activities, and careful qualitative and quantitative measurements may therefore yield some insights in the activity at work. However, steady state measurements can only provide some quantitative estimates on enzyme activity when contributing factors such as specific activity and half-life of the product or metabolite are known. Conversely determining the precise concentration of an enzyme, by classical proteomics for example, does not necessarily divulge what its activity might be in any particular location or situation as that is also dependent on activation, inhibition, processing, and folding among other features (6, 7). Taken together, quantita-tion of enzymatic activities in a complex biological matrix, either individually or in aggregate form, is clearly in the realm of proteomics but is regrettably a largely unpracticed subspecialty at the moment.

Some notable efforts made toward activity-based proteomics profiling to date involve the use of chemical probes that selectively label, on a whole-proteome background, certain classes of active enzymes but not their inactive forms (8–11). Further to this idea, fluorophore- or affinity tag-derivatized versions have been shown to make excellent tools for the visualization of activity signatures in, and the retrieval and identification of the actual enzymes from, entire proteomes. It could be argued, however, that active states and the availability and access to enzyme active sites are still one notch away from bona fide enzymatic activity, i.e. bringing about

From the \[‡\]Protein Center and \[§\]Molecular Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10065 and \[¶\]Department of Medicine, The Methodist Hospital, Houston, Texas 77030

Received, August 21, 2007, and in revised form, October 23, 2007

Published, MCP Papers in Press, November 6, 2007, DOI 10.1074/ mcp.M700397-MCP200
Comparative Exoprotease Activity Test

The abbreviations used are: FPA, fibrinopeptide A; C3f, complement 3f; Fmoc, 9-fluorenylmethoxy-carbonyl; Boc, t-butoxycarbonyl; DEGR, degradation product; REF, reference; SVM, support vector machine; LOOCV, leave-one-out cross-validation; SSEAT, sequence-specific exoprotease activity test; CLUS, clusterin.

Chemical change in a target, ranging from simple modifications to anabolic or catabolic reactions. In the literal sense, activity-based or functional proteomics would therefore entail profiling based on genuine activity readouts.

Classical post-translational modification profiling could be considered a form of functional proteomics albeit with the limitations stated above. Ideally these or other type (mixed) activities should be measured de novo in a multiplexed manner and with controlled amounts of substrate(s). Under the right conditions, such assays would allow amplification of the output signals, thus potentially visualizing low abundance enzymes (i.e. proteins) on a virtually transparent whole-proteome background, something akin to antibody-based detection but without any prior knowledge of analyte identity. Oncopeptidomics, the analysis of the blood serum peptidome of cancer patients most often for biomarker discovery purposes (12–14), is also a special brand of activity-based proteomics as it basically provides a peptide metabolomic readout of protein synthesis, secretion, processing, and proteolytic (i.e. enzymatic) degradation. This could be particularly relevant to cancer as increased and/or altered expression and secretion of various proteases has been associated with tumor growth, invasion, and metastasis (15, 16). Following comparative peptidomics screens, in combination with multivariate statistical analysis, it has been postulated that serum peptide patterns, as surrogates for proteolytic activities, reflect these important physiological changes in cancer patients and may therefore contain diagnostic information (17). However, no independent verification by other laboratories has been provided so far.

An inherent problem of current oncopeptidomics is that the bulk of the serum peptidome as measured using high throughput, low resolution MS is the result of exoprotease activities targeted at some ex vivo coagulation and complement activation by-products, such as fibrinopeptide A (FPA) and complement 3f (C3f) among others. Whereas these peptides appear to be ideal substrates, their formation depends on a process that is sometimes deregulated in cancer (a disorder known as malignant coagulopathy (18–20)), and their degradation is highly dependent and indeed easily affected by clinical sample collection and handling, thus necessitating great care (21). A promising source of unique, functional cancer biomarkers may therefore go untapped unless a more reliable activity test for the postulated but as yet unidentified exopeptidases can be designed. From a classical proteomics point of view, some of these enzymes may also be really low abundance, especially during early stages of cancer, and therefore “invisible” in traditional MS-based discovery schemes. However, given enough substrate, time, and optimal assay conditions, catalytic product may accumulate to the level that it becomes readily detectable using any type of mass spectrometer.

Here we present a method to monitor controlled, de novo peptide breakdown in large numbers of biological samples using MALDI-TOF mass spectrometry, with relative quantitation of the metabolites via comparison with spiked internal standards, followed by multivariate statistical analysis of the resulting minipeptidome. We show that the test is uniquely suited to probe the altered balance of exopeptidases and/or their modulators in the blood of cancer patients, and we have used that information for class prediction purposes. This novel method may lead to a fundamentally new approach to biomarker development.

MATERIALS AND METHODS

Serum Samples—Human reference serum (catalog number S-7023, lot 034K8997) was purchased from Sigma. Blood samples from volunteer subjects with no known malignancies and from patients diagnosed with metastatic thyroid carcinoma were collected following informed consent, under a protocol approved by the Memorial Sloan-Kettering Cancer Center Institutional Review and Privacy Board, as documented previously (22). The 48 patient (supplemental Table S1A) serum samples used in the current study were taken from a previously collected and documented (e.g. patient age, gender, and pathologic diagnosis) set of 60 samples (22); 48 age- and gender-matched healthy controls (supplemental Table S1B) were then selected from an existing reference set also described previously (22). All 96 samples had been collected at the same location (Memorial Sloan-Kettering Cancer Center) and according to the exact same standard operating procedure (22). As before, each serum sample had been frozen and thawed twice before it was subjected to solid-phase peptide extraction and mass spectrometric analysis. Patient and control serum “pools” (2.4 ml each) were obtained by mixing 50-μl aliquots from each of the 48 individual cancer patient samples and, respectively, the 48 healthy control samples. Each pool was divided in 120 × 20-μl aliquots and stored at −80 °C until further use. Multiple blood samples were obtained, all within a single 10-min period, from a healthy volunteer, “NL106” (supplemental Table S1), for analytical studies and assay optimization purposes.

Synthetic Peptides—All peptides were synthesized, purified (>98% purity), and quality/purity-controlled at the Memorial Sloan-Kettering Cancer Center microchemistry and proteomics core facility as described previously (23, 24). 9-Fluorenylmethoxy-carbonyl-protected (“Fmoc”) l-configuration amino acids (“l-amino acids”; including [13C6]Leu, [13C6,15N]Leu, and [13C5,15N]Val isotope-labeled forms), Fmoc-o-amino acids, Fmoc-l-amino acid resins, Fmoc-o-Arg(Pbf)-Wang resin, Fmoc-l-[13C6,15N]Leu hydroxymethyl polystryrene 4-methylbenzhydrylamine resin, and H-D-Lys(Boc)-2-chlorotrityl resin were obtained from Anaspec (San Jose, CA); Fmoc-o-Val-Wang resin was from Novabiochem; and Fmoc-l-[13C6,15N]Ile and Fmoc-l-[ring-13C6]Phe were from Cambridge Isotope Laboratories (Andover, MA). Sequential coupling of amino acids was carried out using an automated instrument, model 433A (Applied Biosystems, Foster City, CA), except for the isotope-labeled residues that were coupled “off line” by reaction of 0.2 mmol with 100 mg of the resin-anchored, growing peptide chain for 15 h at room temperature while shaking (24). Reagents and solvents were from Applied Biosystems. Peptide stocks of degradable substrates (Table I) and the corresponding reference peptide ladders (supplemental Table S2) were reconstituted...
in 30% acetonitrile, 0.1% formic acid and stored at −80 °C. Quantitation was done by amino acid analysis at the Keck Biotechnology core facility (Yale University, New Haven, CT), and stocks were adjusted to 100 pmol/μl for the substrates and to a final concentration of 10 pmol/μl for each peptide contained in each of the three reference mixtures (seven peptides in the “FPA-derived ladder” mixture, 10 in the “C3f-ladder”, and seven in the “CLUS2-ladder”).

**Exoprotease Activity Assays**—Individual or pooled serum samples (frozen 20-μl aliquots for one-time use; see above) or Sigma reference serum (20 μl) was thawed on ice and transferred to the wells of 0.2-ml polypropylene “Template III PCR” half-skirted, 96-well microtiter plates (USA Scientific, Ocala, FL). Substrate and reference peptide stock solutions were separately diluted into water (10-μl final volume each) to yield the selected concentrations and were then added to the sera. In the case where reference peptides were omitted, water was added to adjust the final volume to 40 μl. The 96-well plate was then covered with aluminum foil and incubated at room temperature for the relevant period of time (25). The amounts of added peptides per reaction and the incubation times were as follows: 40 pmol of degradable C3f plus 2 pmol of each reference peptide incubated for 2 h; 100 pmol of degradable CLUS2 plus 2 pmol of each reference peptide incubated for 2 h. Peptide amounts and incubation times were frequently modified for investigative or optimization purposes as indicated in the text.

Upon completion of the reactions off line, the 96-well plate was transferred to a liquid handling station for automated, magnetic particle-assisted peptide extraction and processing before MALDI-TOF mass spectrometric analysis as described previously (25) with the following modifications to spectral acquisition. Spectra were taken in reflector mode geometry under 20 kV (16.45 kV during delayed extraction) of ion accelerating and −1.4 kV multiplier potentials and with gating of mass ions set to m/z = 500. Delayed extraction was maintained for 80 ns to give time lag focusing after each laser shot. Daily robotic and mass spectrometric performance tests were done using “Sigma” human reference serum, and the effective laser energy delivered to the target was adjusted as necessary (21, 25). In those cases when two sets of individual samples (e.g. 48 healthy control and 48 patient samples) were assayed, sample randomization in the 96-well plate, and by extension on the MALDI target plates, was done as described previously (25). Mass spectra signal processing was as described previously (21, 22, 26); details can be found below.

Ratios were calculated of the normalized ion intensities of the degradation product (“DEGR”) peaks over the normalized ion intensities of the corresponding reference (“REF”) peptide peaks for each rung of the ladder (i.e. each of the individual peptides constituting an entire nested set). The spreadsheet resulting from the duplicate analyses was subjected to the same process, and the “DEGR/REF” ratios for each of the rungs of the same peptide ladder from the duplicate experiments were averaged. This process was repeated for each of the three (C3f, FPA(−Ala), and CLUS2) assay substrates.

**Statistical Analysis**—Spreadsheets containing degradation product/reference peptide ratios (see “Signal Processing”) derived from spectra obtained from duplicate assays of 96 individual samples and using three different substrates separately (i.e. six sets of 96 analyses), were imported into the “GeneSpring” program, version 7.3.1 (Agilent, Palo Alto, CA). Different “experiments” were created to represent the ratios, and no normalizations were applied to the experiment because the data had been normalized previously. In the “Experiment’s Interpretation” section, the analysis mode was set to “Ratio” (signal/control), and all measurements were used. No cross-gene error model was used. Support vector machine (SVM) analyses were done using the Class Prediction Tool in GeneSpring. Leave-one-out cross-validation (LOOCV) analyses were done by SVM modeling of each of the different sets of peptides, as well as of the three sets combined, using the following kernels: linear, polynomial (order 2), polynomial (order 3), and radial. In addition to LOOCV analyses using the proper 96 class labels (i.e. true healthy subjects and thyroid cancer patients), random combinations of class labels were tested for SVM modeling. Ten different random combinations of the 96 class labels were generated for each of the three (C3f, FPA(−Ala), and CLUS2) substrate datasets and subjected to LOOCV analyses. Classification rates were averaged from 10 different analyses and compared with the earlier obtained classifications rates when the true combination of class labels had been used.

**RESULTS**

*A Sequence-specific Exoprotease Activity Test (SSEAT)*

We sought to develop a reproducible test of the global exoprotease activity repertoire within individual proteomes of two or more groups of cultured cells, tissue lysates, or biological fluids by accurate quantification of degradation products and multivariate statistical analysis of the resulting qualitative and quantitative patterns. More specifically, we wanted this test to interrogate the aggregate activity of a particular, yet molecularly undefined, subset of all aminopeptidases and carboxypeptidases within a proteome that act in unison to degrade a peptide (or peptides) with defined sequences. So far, the substrates of choice have resulted from prior discovery efforts that involved comparative peptidomics analyses of groups of biological specimen (17, 22), but they could also be

---

**Comparative Exoprotease Activity Test**

**RESULTS**

*A Sequence-specific Exoprotease Activity Test (SSEAT)*

We sought to develop a reproducible test of the global exoprotease activity repertoire within individual proteomes of two or more groups of cultured cells, tissue lysates, or biological fluids by accurate quantification of degradation products and multivariate statistical analysis of the resulting qualitative and quantitative patterns. More specifically, we wanted this test to interrogate the aggregate activity of a particular, yet molecularly undefined, subset of all aminopeptidases and carboxypeptidases within a proteome that act in unison to degrade a peptide (or peptides) with defined sequences. So far, the substrates of choice have resulted from prior discovery efforts that involved comparative peptidomics analyses of groups of biological specimen (17, 22), but they could also be

---

**Comparative Exoprotease Activity Test**

**RESULTS**

*A Sequence-specific Exoprotease Activity Test (SSEAT)*

We sought to develop a reproducible test of the global exoprotease activity repertoire within individual proteomes of two or more groups of cultured cells, tissue lysates, or biological fluids by accurate quantification of degradation products and multivariate statistical analysis of the resulting qualitative and quantitative patterns. More specifically, we wanted this test to interrogate the aggregate activity of a particular, yet molecularly undefined, subset of all aminopeptidases and carboxypeptidases within a proteome that act in unison to degrade a peptide (or peptides) with defined sequences. So far, the substrates of choice have resulted from prior discovery efforts that involved comparative peptidomics analyses of groups of biological specimen (17, 22), but they could also be
selected by presenting various synthetic peptides or peptide libraries to the same groups of samples in search of substrates that most clearly reveal differential degradative activities. Fig. 1 depicts the schematic of a test that satisfies all the above criteria.

A fixed amount of a suitable, degradable substrate is added to an equal volume of each of the samples under study. If any of these samples contain endogenous peptides identical to the ones expected to arise in the assay, as is the case for instance in serum, the substrate is isotopically “tagged” by incorporation of an amino acid containing a stable isotope ($^{13}$C$_6$ or $^{13}$C$_5$ with $^{15}$N). Ideally this labeled residue must be present in each of the resulting degradation products, making positioning in the sequence critical; i.e., it should be at, or near, the opposite end of where the exoproteolytic activity is expected to occur. This information can be inferred from prior peptidomics analysis, in the case of serum samples, or from pilot studies with unlabeled peptide substrates in biological fluids that contain fairly low concentrations of the endogenous peptides, such as plasma. Substrates are incubated with the samples under strictly controlled conditions of time, temperature, and whatever other variables one might decide to modify for investigative or assay purposes. Incubation times are peptide- and sample-dependent and will typically be optimized during pilot experiments using pooled samples from each of the groups to be compared. Following incubation, the samples are subjected to peptide profiling. The newly formed degradation products and pre-existing endogenous degradation products, if any, are measured simultaneously using an existing platform that consists of magnetic particle-assisted, automated processing and MALDI-TOF mass spectrometry (21, 25, 27). Accurate quantitation of each fragment is done by comparison of the ion intensities or areas under the curve of the corresponding $m/z$ peaks with those of double labeled ($2 	imes ^{13}$C$_6$/15N) reference peptides (“internal standards”) spiked into the samples at known absolute concentrations at the same time when the substrate was added. The reference peptides have been synthesized with D-amino acids and are therefore fully resistant to any proteolytic degradation that, like all other enzymatic reactions in nature, is a stereospecific process. By adding the internal standards before the assay and the subsequent automated peptide extraction, the same degree of adsorptive and processing-related losses applies as for the newly generated peptide degradation products and the existing endogenous peptides. After processing of the TOF spectra (26), all peptide pairs are assigned, and ratios are calculated as commonly done in other MS-based, relative, or absolute quantification strategies (28–32). For comparative analysis of large sample sets, the ion intensity of each reference peptide is given an arbitrary value of 1 for that particular pair. Note that the difference between the $m/z$ ($z = 1$) of the ion peak corresponding to the single and double isotopically labeled peptides as shown in Fig. 1 is 8 amu.

**Assay Optimization**—In keeping with our ongoing studies on oncopeptidomics and biomarker patterns, we have selected a series of candidate substrate peptides based on previously observed peptide ladders in serum (17, 33); nearly all contain one or more rungs that are part of the peptide ion signatures for either thyroid cancer and/or bladder, prostate, and breast cancer (17, 22). In a first round of optimization...

---

**Fig. 1.** Strategy for monitoring differential exoprotease activities in biological samples. A degradable, isotopically labeled substrate and its corresponding set of non-degradable, double labeled reference peptides (representing a nested set of truncated sequences) are added to two samples A and B that may, for instance, represent two different clinical states. At a given time point, the two reactions are stopped and analyzed by magnetic particle-based solid-phase extraction and a MALDI-TOF mass spectrometric readout. Spectra are processed, ion intensities are normalized versus the total ion current, matching peptide pairs are assigned, and the ratios of each of the degradation products versus the corresponding spiked reference peptides are calculated. This ratio represents the relative ion intensity of each peptide that is comparable across all samples.
studies, 100 pmol of eight peptides (precursor proteins and locations in each sequence are given in Table I) were individually incubated with 20 μl of pooled control sera for 15 min. Four degraded, either fully or to some extent, but the other peptides did not (Table I) even after prolonged incubation. FPA was totally degraded into amino acids and small fragments undetectable by our standard MALDI-TOF MS analysis after 15 min (Fig. 2), but tests with N-terminally truncated versions indicated that removal of a single Ala slows down the degradation process of this peptide considerably; it takes hours to generate appreciable amounts of breakdown products (Fig. 2). Degradation rates C3f and of two overlapping fragments of the C-terminal part of the clusterin β-chain were somewhere in between those of FPA and FPA(−Ala) (Table I and Fig. 2). Note that because of the high concentrations of the added substrates the endogenous serum peptides are seemingly absent or appear as very low intensity ions in the spectra shown in Fig. 2. Synthetic, isotopically labeled versions of FPA(−Ala), C3f, and CLUS2 (the positioning of the labeled amino acids is indicated in Table I) were then produced for further study as well as non-degradable, doubly labeled versions of the same three peptides and of all the anticipated degradation products that could be detected using our standard protocol in a series of pilot experiments (the sequences and positioning of all labeled amino acids are listed in supplemental Table S2).

**SSEAT Is Reproducible and Tolerant of Serum Preparation Variability—**We have reported previously that some of the common errors in serum peptidomics, such as variation in clotting times and the times that samples are kept at room temperature before freezing as well as the number of freeze/thaw cycles, will lead to extensive variability in spectral ion intensities and could, therefore, also lead to bias (21). This presents a major problem in busy hospital environments where these steps in the procedure are not always controlled to the minute. We thus investigated how the sometimes hard-to-control variables might affect reproducibility of our MS-based, exoprotease-associated peptide degradation assay. Several serum samples were prepared from freshly collected blood, thereby varying clotting times, times “left on the bench,” and the number of freeze/thaw cycles. Note that our standard procedure calls for 1-h clotting time, no extra time at room temperature, and two freeze/thaw cycles. Labeled FPA(−Ala) (200 pmol) was then added to 20-μl aliquots of control serum and incubated at room temperature for 8 h, and the samples were processed and analyzed by MS. Fig. 3 depicts a series of paired m/z peaks (i.e. isotopic envelopes) that are the matching degradation products (sequences indicated) of the unlabeled endogenous and the labeled exogenous (i.e. added) substrates. The color-coded spectra clearly indicate that the ion signals corresponding to “endogenous” peptides (lower m/z; left position in each of the six panels) varied greatly as expected. In contrast, those corresponding to the degradation products of the added substrate (higher m/z; right in each panel) are almost uniform, indicating that none of the clinical chemistry variables under study affected the outcome and reproducibility of the assay, thus satisfying one important criterion of a prospective test. Similar observations were made when using C3f and CLUS2 as substrates (data not shown).

To assess reproducibility of the analytical portion (automated peptide extraction and MS), the exoprotease activity test was carried out five times over 5 consecutive days using pooled control serum, labeled C3f, and 10 associated reference peptides (supplemental Table S2) as described under “Materials and Methods.” Whereas the coefficient of variation (CV) of the normalized ion intensities of the DEGRs and of the matching REF peptides varied, respectively, from 6.6 to 33.8% and 5.6 to 33%, the CV of the ratios (DEGR/REF) was always less than 14.3% (supplemental Table S3).
To explore the use of the exoprotease activity test for functional biomarker discovery, we analyzed serum samples from 48 patients with metastatic carcinomas and from 48 age- and gender-matched healthy controls. To compare the results with those of standard peptidomics-based studies, we selected the patient and control samples from two existing sets that had been collected previously and analyzed under a strict standard operating procedure and from which a 12-peptide ion thyroid cancer signature has been developed that allowed SVM-aided class prediction with 95% sensitivity and 95% specificity (22). Two serum pools, consisting of equal volume aliquots of all patient samples and, respectively, all control samples, were used to investigate and optimize assay conditions to yield maximum differences in degradative patterns between the two groups. The time course results shown in Fig. 4 point to moderately differential rates of C3f degradation and appearance of selected breakdown products under the conditions used (see “Materials and Methods” for more details).
Methods”). The time plots show the corrected quantitative values (DEGR/REF ratios) for the intact substrate (C3f; see Table I) and for the largest and two smallest rungs in the degradation ladder, all measured in triplicate analyses. Initial removal of the C-terminal Arg appears to occur more rapidly, and the accumulation of smaller breakdown products between 90 min and 3 h of incubation time is more prominent in the cancer patient sera. After quantitation of those three fragments, as well as all other fragments (data not shown), an incubation time of 3 h was selected for all further comparative analyses. Similar optimization studies were also carried out using the CLUS2 and FPA(–Ala) substrates (data not shown).

In the next phase, 48 patient and 48 control samples were assayed individually, using each of the three chosen substrates separately, in duplicate. Amounts of substrate and reference peptides and the specific incubation times were as stated under “Materials and Methods.” Spectra were obtained, aligned, and normalized as described under “Materials and Methods” and are displayed using the Mass Spectra Viewer. Each of the overlays contains 96 spectra with normalized intensities: 48 controls (in blue) and 48 thyroid cancer patients (red). The 13C isotopes from the isotopic envelope are shown for each peptide ion peak. Substrate sequences and position of the label are listed in Table I. Abbreviations of each degradation product shown in the panels correspond to the sequences listed in supplemental Table S2 but are singly labeled just like the substrates from which they are derived.

In Fig. 5, ion intensities of identical degradation products differed to varying extents between cancer and controls in a group-specific manner. Although some peaks were more discriminative than others in this regard, not one showed complete segregation between all members of both groups. However, that was before any quantitative correction. After calculating the ratios for each of the analyte/reference pairs (seven for FPA(–Ala), 10 for C3f, and seven for CLUS2), data were averaged for the duplicate analyses, and the resulting spread sheet (supplemental Table S4, A and B) was used for class predictions using LOOCV and an SVM learning algorithm. The results presented in Fig. 6 indicate that without any prior feature selection the quantitative degradation patterns resulting from any of the three individual substrates allowed LOOCV class prediction
predictions with optimal sensitivities ranging between 69 and 90% and specificities between 67 and 88%. However, by combining all features (24 total), 94% sensitivity (binomial confidence interval of 83–99%) and 90% specificity (binomial confidence interval of 77–96%) could be obtained. All assay datasets were also analyzed after randomization of the class labels (cancer and controls). Classification rates were averaged for 10 independent randomizations and resulted in about 50% (±5%) accuracy as expected.

**DISCUSSION**

We describe an analytical method that is applicable to oncopeptidomics research and that may lead to a fundamentally new approach to biomarker development. The activity test presented herein is aimed at detecting and comparing exoprotease activities in entire proteomes of tissue lysates or biological fluids. It provides a one-step readout of not one but supposedly several amino- and carboxypeptidases that act in concert to degrade peptides of specific sequences. This is particularly relevant to cancer as proteases are well established components of tumor progression and invasiveness (15, 16, 34). The “aggregate” activities that we measure at the end point of a chain of events consisting of synthesis, secretion, processing, and activation of each participating enzyme and are also affected by the presence or absence of specific modulators. Together the variations in the above processes, conditions, and effectors produce physiological patterns that reflect the biological state of an organism, including disease, but that may not be readily detectable by standard “omics” screening methods, hence the need for alternative approaches such as the one presented here. An exoprotease activity test obviously requires a suitable substrate, which may be found by targeted peptidomics screens and multivariate statistics (17), the “discovery” phase of functional biomarker development so to speak.

Kinetics of degradation, the resulting patterns, and the precise amounts of every metabolite at any chosen time point are all key features of peptide exoproteolysis. Each is taken into consideration in design, measurements, and data analysis of our test. Assay time is optimized in pilot analyses to generate maximum differential degradative patterns between two groups of samples. A snapshot is then taken by MALDI-TOF MS of isotope-tagged, de novo peptide breakdown products in each of the individual samples followed by relative quantitation by comparison with known amounts of spiked internal standards. These standards (“reference peptides”) are both non-degradable in biological media and distinguishable by MS from all endogenous and substrate degradation products. Instead of simply monitoring the disappearance of substrate or appearance of any particular fragment over time, or using a ratio function of both, we regard the full array of peptide metabolites derived from multiple substrates as a minipeptidome that is then quantitated (with CV values in the range of 6.3–14.3% over five replicate analyses) and can be subjected to multivariate statistical analysis.

Putting this test into practice, we analyzed samples of 48 patients with metastatic thyroid carcinomas and of 48 healthy controls and obtained LOOCV class predictions with 94% sensitivity and 90% specificity without prior feature selection (i.e. a maximum of 24 features, derived from degradation of three different peptides, were used). We expect that these numbers can be improved upon in the future by making a series of additions and adjustments. First, we are developing additional peptide substrates and producing appropriately matching reference peptides for use in these tests. More degradation products will result in a bigger “minipeptidome” and therefore more features from which to select. Multiple substrates may be used in multiplexed activity assays with or
without sequential, time-dependent withdrawal of aliquots for analysis. Second, the test can be tailored, in terms of substrates, assay conditions, and pattern discovery, to many diagnostic or predictive purposes and other uses in disease management.

Another advantage of this test, and perhaps the single major reason we developed it in the first place, is to eliminate reproducibility problems related to sample collection, storage, and handling that have beset many serum peptidomics studies of the past (21, 35, 36). As already stated, the serum peptidome is derived for a large part from a limited number of endogenous peptides, nearly all byproducts of clotting, that have little to no diagnostic value on their own (17). Further degradation of these “founder” peptides is exceedingly dependent on time, temperature, and the number of freeze/thaw cycles, which greatly affects the eventual peptide patterns and thus compromises any diagnostically useful information that might be extracted from such patterns. However, none of these aforementioned factors appear to affect our rigorously controlled, in vitro exoprotease activity test. By dispensing precise amounts of synthetic substrates into each sample, we also avoid any problems related to variability in the endogenous precursor protein and peptide (e.g., FPA) concentrations, for instance in the case of hemostatic alterations in cancer patients (37–40). Further to hemostatic dysregulation, it is telling that no exopeptidases are cited in the chapters dealing with coagulation in two prominent hematology and blood disease texts, only endopeptidases (41, 42). Thus, the likelihood that malignant coagulopathy or other ephiphenomena would affect exoprotease panels and activities in blood is fairly remote we believe, but it cannot be completely ruled out despite the absence of any evidence to the contrary. Examination of the peer-reviewed literature turned up only one exopeptidase known to be activated during clotting, namely thrombin-activable fibrinolysis inhibitor, a lysine-specific carboxypeptidase (43). As none of the substrates used in our study have lysine in that preferred position, aberrant thrombin-activable fibrinolysis inhibitor activity is most likely not an issue.

We expect that future uses and implementation of the exoprotease activity test will go well beyond those provided in the current account. Quantitative assays may be done more accurately and yield absolute amounts instead of comparative values for each of the metabolites using multiplexed multiple reaction monitoring MS (28, 29). The test may also be used to monitor purification of specific exoproteases or combinations thereof from serum, plasma, or any other biological fluid or tissue lysate or from the conditioned media of selected cancer cell lines that secrete similar “activities”. As large portions of our assay are automated, dozens of column fractions can readily be screened within hours; hundreds can be screened within a day. Purification and identification of cancer-relevant exoproteases would allow raising antibodies for ELISA-based screens with the caveat that knowledge of the presence/absence or concentrations alone may not provide the full picture as it will miss out on important functional information as already discussed.

In conclusion, we have described a new approach and related test uniquely suited to probe the altered balance of exoproteases and/or their modulators in the blood of cancer patients and to prospectively use that information for diagnostic or prognostic purposes. This test now offers the option of a targeted, functional proteomic readout that may be either a supplement or a practical alternative for the classical biomarker discovery techniques. It may also extend the analytical range well beyond abundant serum and plasma components by way of amplification of unique signals on a virtually transparent background.

Acknowledgments—We thank Lisa Balistreri for blood collection and Hans Lilja and Martin Fleisher for helpful discussions.

* This work was supported by United States National Institutes of Health Grants U24CA126485 and R33CA111942 and by the Himan-shu Vakil Research Fund. It is part of the NCI, National Institutes of Health Clinical Proteomic Technologies Initiative and Clinical Proteomic Technology Assessment for Cancer (CPTAC) consortium (Broad Institute of the Massachusetts Institute of Technology and Harvard; Memorial Sloan-Kettering Cancer Center; Purdue University; University of California, San Francisco; and Vanderbilt University School of Medicine). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The on-line version of this article (available at http://www.mcponline.org) contains supplemental material.

To whom correspondence should be addressed: Memorial Sloan-Kettering Cancer Center, 1275 York Ave., New York, NY 10065. E-mail: p-tempst@mskcc.org.

REFERENCES

1. Washburn, M. P., Wolters, D., and Yates, J. R., Ill (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol. 19, 242–247

2. Qian, W. J., Jacobs, J. M., Liu, T., Camp, D. G., II, and Smith, R. D. (2006) Advances and challenges in liquid chromatography-mass spectrometry-based proteomics profiling for clinical applications. Mol. Cell. Proteomics 5, 1727–1744

3. Jensen, O. N. (2004) Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. Curr. Opin. Chem. Biol. 8, 33–41

4. Petricoin, E. F., Belluco, C., Araujo, R. P., and Liotta, L. A. (2006) The blood peptidome: a higher dimension of information content for cancer biomarker discovery. Nat. Rev. Cancer 6, 961–967

5. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2000) Molecular Biology of the Cell, 4th Ed., Garland Science, New York

6. Kobe, B., and Kemp, B. E. (1999) Active site-directed protein regulation. Nature 402, 373–376

7. Sidransky, D. (2002) Emerging molecular markers of cancer. Nat. Rev. Cancer 2, 210–219

8. Greenbaum, D., Baruch, A., Hayrapetian, L., Darula, Z., Burlingame, A., Medzihradzsky, K. F., and Bogoy, M. (2002) Chemical approaches for functionally probing the proteome. Mol. Cell. Proteomics 1, 60–68

9. J. Villanueva, A. Nazarian, K. Lawlor, and Paul Tempst, unpublished observations.
Comparative Exoprotease Activity Test

9. Kato, D., Boatright, K. M., Berger, A. B., Nazif, T., Blum, G., Ryan, C., Chehade, K. A., Salvesen, G. S., and Bogoy, M. (2005) Activity-based probes that target diverse cysteine protease families. Nat. Chem. Biol. 1, 33–38

10. Adam, G. C., Sorensen, E. J., and Cravatt, B. F. (2002) Proteomic profiling of mechanically distinct enzyme classes using a common chemotype. Nat. Biotechnol. 20, 805–809

11. Jessani, N., Niessen, S., Wei, B. Q., Nicolau, M., Humphrey, M., J., Han, W., Noh, D. Y., Yates, J. R., III, Jeffrey, S. S., and Cravatt, B. F. (2005) A streamlined platform for high-content functional proteomics of primary human specimens. Nat. Methods 2, 691–697

12. Schrader, M., and Seile, H. (2006) The process chain for peptidomic biomarker discovery. Dis. Markers 22, 27–37

13. Tammien, H., Zucht, H. D., and Budde, P. (2007) Oncoproteidomics—a commentary on opportunities and limitations. Cancer Lett. 249, 80–86

14. Diamandis, E. P. (2007) Oncoproteidomics: a useful approach for cancer diagnosis? Clin. Chem. 53, 1004–1006

15. DeClerck, Y. A., Mercurio, A. M., Stack, M. S., Chapman, H. A., Zutter, M. M., Muschel, R. J., Raz, A., Matsrisan, L. M., Sloane, B. F., Noel, A., Hendrix, M. J., Coussens, L., and Padarathsingh, M. (2004) Proteases, extracellular matrix, and cancer: a workshop of the path B study section. Am. J. Pathol. 164, 1131–1139

16. Puente, X. S., Sanchez, L. M., Overall, C. M., and Lopez-Otin, C. (2003) Human and mouse proteases: a comparative genomic approach. Nat. Rev. Genet. 4, 544–558

17. Villanueva, J., Shaffer, D. R., Philip, J., Chaparro, C. A., Erdjument-Bromage, H., Olshen, A. B., Fleisher, M., Lilja, H., Brogi, E., Boyd, J., Sanchez-Carbayo, M., Holland, E. C., Cordon-Cardo, C., Scher, H. I., and Tempst, P. (2006) Differential exoprotease activities confer tumor-specific peptide signature patterns. J. Clin. Investig. 116, 271–284

18. Rickles, F. R., and Edwards, R. L. (1983) Activation of blood coagulation in cancer: Trousseau’s syndrome revisited. Blood 62, 14–31

19. Bick, R. L. (1992) Coagulation abnormalities in malignancy: a review. Semin. Thromb. Hemost. 18, 353–372

20. Green, K. B., and Silverstein, R. L. (1996) Hypercoagulability in cancer. Hematol. Oncol. Clin. N. Am. 10, 499–530

21. Villanueva, J., Philip, J., Chaparro, C. A., Li, Y., Toledo-Crow, R., DeNoyer, L., Fleisher, M., Robbins, R. J., and Tempst, P. (2005) Correcting common errors in identifying cancer-specific serum peptide signatures. J. Proteome Res. 4, 1060–1072

22. Villanueva, J., Martorella, A. J., Lawlor, K., Philip, J., Fleisher, M., Robbins, R. J., and Tempst, P. (2006) Serum peptide patterns that distinguish metastatic thyroid carcinoma from cancer-free controls are unbiased by gender and age. Mol. Cell. Proteomics 5, 1840–1852

23. Castells, P., Romagnolo, J., Castle, M., Castells-Josson, K., Erdjument-Bromage, H., and Tempst, P. (1994) Biodiversity of apidaecin-type peptide antibiotics. Prospects of manipulating the antibacterial spectrum and combating acquired resistance. J. Biol. Chem. 269, 26107–26115

24. Castle, M., Nazarian, A., Yi, S. S., and Tempst, P. (1999) Lethal effects of apidaecin on Escherichia coli involve sequential molecular interactions with diverse targets. J. Biol. Chem. 274, 32555–32564

25. Villanueva, J., Lawlor, K., Toledo-Crow, R., and Tempst, P. (2006) Automated serum peptide profiling. Nat. Protoc. 1, 880–891

26. Villanueva, J., Philip, J., DeNoyer, L., and Tempst, P. (2007) Data analysis of assorted serum peptide signatures. Nat. Protoc. 2, 588–602

27. Villanueva, J., Philip, J., Enentenberg, D., Chaparro, C. A., Tanwar, M. K., Holland, E. C., and Tempst, P. (2004) Serum peptide profiling by magnetic particle-assisted, automated sample processing and MALDI-TOF mass spectrometry. Anal. Chem. 76, 1560–1570

28. Anderson, L., and Hunter, C. L. (2006) Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. Mol. Cell. Proteomics 5, 573–588

29. Barnidge, D. R., Goodmanson, M. K., Klee, G. G., and Muddiman, D. C. (2004) Absolute quantification of the model biomarker prostate-specific antigen in serum by LC-MS/MS using protein cleavage and isotope dilution mass spectrometry. J. Proteome Res. 3, 644–652

30. Gerber, S. A., Ruah, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc. Natl. Acad. Sci. U. S. A. 100, 6940–6945

31. Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol. Cell. Proteomics 1, 376–386

32. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat. Biotechnol. 17, 994–999

33. Koomen, J. M., Li, D., Xiao, L. C., Liu, T. C., Coombes, K. R., Abbruzzese, J., and Kobayashi, R. (2005) Direct tandem mass spectrometry reveals limitations in protein profiling experiments for plasma biomarker discovery. J. Proteome Res. 4, 972–981

34. Matsrisan, L. M., Sledge, G. W., Jr., and Mohla, S. (2003) Extracellular proteolysis and cancer: meeting summary and future directions. Cancer Res. 63, 6105–6109

35. Check, E. (2004) Proteomics and cancer: running before we can walk? Nature 429, 496–497

36. Diamandis, E. P. (2004) Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations. Mol. Cell. Proteomics 3, 367–378

37. Nand, S., Fisher, S. G., Salgia, R., and Fisher, R. I. (1987) Hemostatic abnormalities in untreated cancer: incidence and correlation with thrombotic and hemorrhagic complications. J. Clin. Oncol. 5, 1998–2003

38. Rickles, F. R., Levine, M., and Edwards, R. L. (1992) Hemostatic alterations in cancer patients. Cancer Metastasis Rev. 11, 237–248

39. Goldsmith, G. H. (2001) Hemostatic changes in patients with malignancy. Int. J. Hematol. 73, 151–156

40. Davis, M. T., and Patterson, S. D. (2007) Does the serum peptidome reveal hemostatic dysregulation? Ernst Schering Res. Found. Workshop 21, 23–44

41. Jandl, J. H. (1996) Blood: Textbook of Hematology, Little, Brown and Co., New York

42. Stamatoyannopoulos, G., Majerus, P. W., Perlmutter, R. M., and Varmus, H. (2001) The Molecular Basis of Blood Diseases, 3rd Ed., W. B. Saunders Co., Philadelphia

43. Marx, P. F. (2004) Thrombin-activatable fibrinolysis inhibitor. Curr. Med. Chem. 11, 2335–2348