Substrate optimization and clinical validation of reporter peptides for MS-based protease profiling in serum specimens: A new approach for diagnosis of malignant disease

DIEGO YEPES[1], ANETTE JACOB[2], MARC DAUBER[2], VICTOR COSTINA[1], RALF HOFHEINZ[3], MICHAEL NEUMAIER[1] and PETER FINDEISEN[1]

[1]Institute for Clinical Chemistry, Medical Faculty Mannheim of the University of Heidelberg, University Hospital Mannheim, Theodor-Kutzer-Ufer 1-3, 68167 Mannheim; [2]Functional Genome Analysis, German Cancer Research Center, INF 580, 69120 Heidelberg; [3]III Medical Clinic, Medical Faculty Mannheim of the University of Heidelberg, University Hospital Mannheim, Theodor-Kutzer-Ufer 1-3, D-68167 Mannheim, Germany

Abstract. The progression of many solid tumors is characterized by the release of tumor-associated proteases, such as cancer procoagulant, MMP2 and MMP7. Consequently, the detection of tumor-specific proteolytic activity in serum specimens has recently been proposed as a new diagnostic tool in oncology. However, tumor-associated proteases are highly diluted in serum specimens and it is challenging to identify substrates that are specifically cleaved. In this study, we describe the systematic optimization of a synthetic peptide substrate using a positional scanning synthetic combinatorial library (PS-SCL) approach. The initial reporter peptide (RP) comprises of the cleavage site, WKPYDAAD, that is part of the coagulation factor X, the natural substrate of the tumor-associated cysteine protease cancer procoagulant (EC 3.4.22.26). Specifically, the amino acid substitution of aspartic acid (D) in position P1' against asparagine (N) improved the processing of respective RPs in serum specimens from patients with colorectal tumors compared to healthy controls. Proteolytic fragments of RPs accumulated during prolonged incubation with serum specimens and were quantified with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Finally, the optimized RP with the cleaved motif WKPYNAAD was combined with the RPs, VPLSLTMG and IPVSLRSG, that were cleaved by the tumor-associated proteases, MMP2 and MMP7, respectively. The diagnostic accuracy of MS-based protease profiling was evaluated for this triplex RP mix in a cohort of 50 serum specimens equally divided into colorectal cancer patients and healthy control individuals. Multiparametric analysis showed an AUC value of 0.90 for the receiver operating characteristic curve and was superior to the classification accuracy of the single markers. Our results demonstrate that RPs for MS-based protease profiling can systematically be optimized with a PS-SCL. Furthermore, the combination of different RPs can additionally increase the classification accuracy of functional protease profiling, and this in turn could lead to improved diagnosis, monitoring and prognosis of malignant disease.

Introduction

Proteases play an important role in different biological processes including cell differentiation, tissue remodelling, haemostasis, immunity, angiogenesis, apoptosis and malignant disease (1). Specifically for solid tumors, proteases are well-known factors for promoting local progression and distant metastasis, and there is increasing evidence that proteases also have key functions in the early stages of malignant disease (2). The tumor-associated proteases are either secreted directly by the tumor or they originate from surrounding connective tissue and infiltrating leucocytes as a result of tumor-stroma interaction (3).

Some tumor-associated proteases, such as matrix metallo-proteinases (MMPs), cathepsines or cancer procoagulant (CP) are released into the bloodstream and can be used for diagnostic (4-7) and prognostic purposes (8). However, due to their low concentration, the detection of tumor-associated proteolytic activity in serum specimens is very challenging. Furthermore, serum has high proteolytic activity that is related to diverse

---

exopeptidases (9) as well as different endoproteases (e.g., from the coagulation cascade and the complement system) (10). Nevertheless, tumor-associated protease activity in serum specimens of cancer patients can be monitored by using specific synthetic substrates that are selectively cleaved by the protease of interest (4-7). With the use of appropriate synthetic reporter peptides (RPs) for the spiking of serum specimens, the reaction conditions that comprise of substrate concentration, incubation time and buffer composition, can be optimized and standardized accordingly. Furthermore, the proteolytic fragments accumulate to a level where they become readily detectable with mass spectrometry (6). This approach is similar to established diagnostic assays measuring the proteolytic activity of distinct enzymes, e.g., coagulation factors (11). With RP spiking, the pre-analytical problems relating to sample collection, storage, and handling, can be alleviated (12).

The success of mass spectrometry-based functional endoprotease profiling with RP spiking has two essential prerequisites (6). Firstly, the RPs must be protected from unwanted proteolytic processing by genuine serum proteases. Any cleavage of RPs by endoproteases is followed by further proteolytic degradation through exoproteases (13). Secondly, and most importantly, suitable RPs with appropriate amino acid sequences have to be identified, that can be cleaved by tumor-associated proteases with the utmost specificity and selectivity (14).

In this study, we present an approach for the optimization of MS-based RP profiling by modulating the cleavage site of a given RP with a positional scanning synthetic combinatorial library (PS-SCL) (15). As a model, we have chosen the cleavage sequence, WKPYDAAD, known to be processed by the tumor-associated protease CP (10). The positions P2, P1 and P'1 (16) were systematically varied and 24 different peptides were synthesized. Serum specimens from colorectal cancer (CRC) patients and healthy control individuals were used to analyze the peptide fragments generated by proteolytic activity using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS). The exchange of the amino acid, aspartic acid (D), in position P1' against asparagine (N) resulted in improved classification accuracy of the serum specimens from non-malignant controls and tumor patients. To alleviate interindividual differences between protease profiles (17) the optimized RP WKPYNAAD was combined with the RPs, VPLSLTMG and IPVSLRSG, that are cleaved by the tumor-associated proteases, MMP2 and MMP7, respectively (18). In a proof-of-principle experiment with sera from CRC patients and healthy controls, the diagnostic performance could further be enhanced by multiplex assay.

Using systematic permutations of the model substrate, we investigated the specificity of protease activities in CRC patients and controls. Furthermore, the effects of combining the candidate model peptide with cleavage substrates for MMP2 and MMP7 on classification were assessed in a multiplex setting. Our results demonstrate that the optimization of RPs for MS-based functional protease profiling is crucial prior to its use in diagnostic applications. Currently, although time consuming and laborious, the growing capabilities of high throughput synthesis of peptide libraries in microarray formats (19) allow for increased optimization and multiplicity.

This could lead to the development of clinically useful tests of proteomics profiling in oncology.

Materials and methods

PS-SCL and RP synthesis. The PS-SCL was synthesized in the functional genome analysis laboratory of the German Cancer Research Centre (Heidelberg, Germany). A 24-PS-SCL was developed in order to optimize the cleavage motif, WKPYDAAD, (Table I) that is part of the coagulation factor, X heavy chain (P00742). This sequence is cleaved after the amino acid tyrosine (Y) by the tumor-associated protease CP (EC 3.4.22.26) (10) and based on the work of Mielicki (20), we decided to vary the positions P2, P1 and P1' of the cleavage motif (16). All peptides were dissolved in DMSO as 4 mM stock solution and aliquots were kept at -20˚C until further use.

Cell culture. The human CRC cell line, HT-29, was purchased from the American Type Culture Collection (ATCC). Cells were grown in RPMI medium that was supplemented with 10% FBS, penicillin (100 units/ml) and streptomycin (100 μg/ml) in a 24-well cell culture plate. The cells were grown at 37˚C and 5% CO_2 and harvested prior to complete confluence. The supernatant was removed from the cell culture plate and immediately centrifuged at 5000 rpm for 5 min. The cleared supernatant was stored at -80˚C until further use.

Serum samples. Whole blood specimens were acquired from patients with metastatic colorectal tumors and healthy controls at the University Hospital Mannheim. Blood collection was performed after we had obtained institutional review board approval and written informed consent from all patients. After a 30-min clotting time at room temperature, the specimens were centrifuged at 20˚C for 10 min at 3000 x g. The serum was removed and aliquots were stored at -80˚C until further use. Any handling and processing of serum specimens from tumor patients and healthy controls was performed in a strictly blinded manner. Quality control (QC) samples (pooled serum specimens from tumor patients and controls) were randomly integrated into the study and measured alongside the test samples.

PS-SCL screening with serum specimens. Each substrate was diluted to 100 μM in 2X digestion buffer (200 mM Tris-HCl pH 7.4, 20 mM $CaCl_2$). Pooled serum specimens from tumor patients and healthy controls were diluted 1:3 with PBS pH 7.4 (PAA Laboratories). Subsequently, 50 μl from 100 μM/ml substrate solutions were mixed with 50 μl 1:3 diluted serum sample and incubated at 37˚C for 2 h. Proteolytic fragments of RPs were extracted from serum specimens as previously described (6). Briefly, streptavidin sepharose high-performance slurry (GE Healthcare) was washed twice and resuspended in PBS pH 7.4 to yield 50% slurry. We then added 50 μl of streptavidin slurry to each sample and incubated the mixture for 15 min at room temperature. After binding, the streptavidin sepharose was centrifuged at 1100 x g for 5 min. A 20 μl aliquot of the supernatant was further purified using the IMAC SL nickel magnetic beads (Bruker Daltonics) prior to MALDI-TOF-MS analysis as described previously (6).

Table I. PS-SCL for the substrate, WKPYDAAD.

| | Cancer procoagulant PS-SCL | | | | | | | |
	P4	P3	P2	P1	P1'	P2'	P3'	P4'
Group 1								
RP1	W	K	P	Y	D	A	A	D
RP2	W	K	P	Y	**E**	A	A	D
RP3	W	K	P	Y	**Q**	A	A	D
RP4	W	K	P	Y	**N**	A	A	D
RP5	W	K	P	**W**	D	A	A	D
RP6	W	K	P	**F**	D	A	A	D
RP7	W	K	P	**W**	**E**	A	A	D
RP8	W	K	P	**F**	**E**	A	A	D
RP9	W	K	P	**W**	**Q**	A	A	D
RP10	W	K	P	**F**	**Q**	A	A	D
RP11	W	K	P	**W**	**N**	A	A	D
RP12	W	K	P	**F**	**N**	A	A	D
Group 2								
RP13	W	K	**V**	Y	D	A	A	D
RP14	W	K	**V**	Y	**E**	A	A	D
RP15	W	K	**V**	Y	**Q**	A	A	D
RP16	W	K	**V**	Y	**N**	A	A	D
RP17	W	K	**V**	**W**	D	A	A	D
RP18	W	K	**V**	**F**	D	A	A	D
RP19	W	K	**V**	**W**	**E**	A	A	D
RP20	W	K	**V**	**F**	**E**	A	A	D
RP21	W	K	**V**	**W**	**Q**	A	A	D
RP22	W	K	**V**	**F**	**Q**	A	A	D
RP23	W	K	**V**	**W**	**N**	A	A	D
RP24	W	K	**V**	**F**	**N**	A	A	D

Amino acids that differ from the original sequence of RP1 are indicated in bold letters. The reporter peptides (RPs) from group 1 (RP1-RP12) had proline in the position P2 and showed different cleavage patterns when compared to the RPs from group 2 with valine (V) at position P2 (see Fig. 1). The non-prime site (P4-P1) and prime site positions (P1'-P4') are labelled according to the nomenclature described by Schechter and Berger (16).

MALDI-TOF measurements. We diluted 1 μl of each sample 1:10 in matrix solution, and 1 μl of the resulting mixture was spotted onto a MTP 384 polished steel target (Bruker Daltonics). MALDI-TOF-MS measurements were performed using an Autoflex II (Bruker Daltonics) operating in the positive mode. The matrix solution was prepared by dissolving 0.3 g/l of α-cyano-4-hydroxycinnamic acid solution (Bruker Daltonics) in ethanol/acetone (2:1, v/v), and external mass calibration was then performed. The MS spectra for peaks in the range of 1-3 kDa were generated by summarizing 1200 laser shots (60 shots; 20 different spot positions; hexagonal cyclic movement) and a maximal laser power of 30%. For each spectrum a peak-picking algorithm (centroid, signal-to-noise ratio >10) was applied by using FlexAnalysis software (Bruker Daltonics).

Kinetics of proteolytic processing for selected RPs. The RPs 1 and 4 from the 24-PS-SCL (Table I) were diluted to 100 μM in 2X digestion buffer (200 mM Tris-HCl pH 7.4, 20 mM $CaCl_2$). The pooled serum specimens from tumor patients (QCT) (n=5) and healthy controls (QCH) (n=5) were diluted 1:3 with PBS and the resulting 50 μl were placed in 96-well microplates (Greiner). Subsequently, 50 μl from 100 μM substrate solutions were added to the pooled serum specimens and incubated at 37°C for 0, 0.5, 1, 2, 4, 6, 8 and 16 h. The proteolytic cleavage of RPs was interrupted by freezing the spiked specimens at -20°C prior to further processing. After incubation the resulting digests were diluted 1:10 with 0.1% TFA prior to MALDI-TOF measurements (21).

Reproducibility of RP spiking. To monitor the reproducibility of RP spiking, two distinct quality control samples were generated, comprising of serum specimens from five colorectal tumor patients (QCT) and five healthy control individuals (QCH), respectively. Both samples were aliquotted and stored at -80°C until further use.

Each QCT and QCH sample was spiked with the RPs, RP1 and RP4, as described above. After incubation for 16 h at 37°C the specimens were frozen at -20°C prior to MALDI-TOF-MS analysis. The proteolytic processing of the RPs resulted in the accumulation of the anchor peptide, D-Ahx-Ahx-HHHHHHt [mass to charge ratio (m/z) 1283] (Table II), and the respective peak intensities were used for quantification. The distribution of the m/z 1283 peak was visualized with whisker box plots, and standard deviation (SD) and coefficient of variation (CV) was calculated with Origin software, version 8.1.

RP validation and multiplex protease profiling in serum specimens. For validation, the optimized RP (RP4) with the cleavage site, WKPYNAAD, was compared to the original substrate (RP1) WKPYDAAD in a set of 25 serum specimens from CRC patients, 25 serum samples from patients with inflammation and high C-reactive protein but no malignant disease and 25 specimens from healthy control individuals. The RPs were diluted with 2X digestion buffer at a final concentration of 50 μM and incubated for 16 h with the serum specimens.

For multiplex protease profiling, the selected RPs from the PS-SC library with the optimized cleavage site, WKPYNAAD (RP4), and the original substrate, WKPYDAAD (RP1), together with two other RPs (RPA and RPB) with known cleavage sites for MMP2 and MMP7, respectively (18), were chosen for this 'proof-of-principle' experiment (Table II). The set of four RPs was incubated with 25 serum specimens from CRC patients and 25 serum specimens from healthy controls. The RPs were added to the serum specimens as described above and incubated at 37°C for 16 h. The proteolytic cleavage of RPs was interrupted by freezing at -20°C prior to further processing. The spiked specimens were diluted 1:10 with 0.1% TFA prior to MALDI-TOF-MS analysis (21).

Statistical and cluster analyses. The Wilcoxon-Mann-Whitney test was performed with the statistical analysis program Origin version 8.1. The receiver operating characteristic (ROC) calculations were performed with the Weka 3 (22)

Table II. Sequences of reporter peptides and proteolytic fragments.

Name of reporter peptide	Linker	Amino acid sequence [non-prime site (P4-P1) - prime site (P1'-P4')]	Linker	C-terminal affinity tag	m/z (M+H)+	Nomenclature in figures
		-P4-P3-P2-P1-P1'-P2'-P3'-P4'-				
RP1	Abu-Ahx-	-W -K -P -Y -D -A -A -D -	-Ahx-Ahx	-HHHHHH-t	2539.9	
		P -Y -D -A -A -D -	-Ahx-Ahx	-HHHHHH-t	1800.9	RP1-P2
		Y -D -A -A -D -	-Ahx-Ahx	-HHHHHH-t	1703.8	RP1-P1
		D -A -A -D -	-Ahx-Ahx	-HHHHHH-t	1540.6	RP1-P1'
		A -A -D -	-Ahx-Ahx	-HHHHHH-t	1425.5	RP1-P2'
		A -D -	-Ahx-Ahx	-HHHHHH-t	1354.4	RP1-P3'
		D -	**-Ahx-Ahx**	**-HHHHHH-t**	**1283.4**	***RP1-P4'**
			Ahx-Ahx	-HHHHHH-t	1168.3	RP1-AP
RP4	Abu-Ahx-	-W -K -P -Y -N -A -A -D -	-Ahx-Ahx	-HHHHHH-t	2538.9	
		P -Y -N -A -A -D -	-Ahx-Ahx	-HHHHHH-t	1799.9	RP2-P2
		Y -N -A -A -D -	-Ahx-Ahx	-HHHHHH-t	1702.8	RP2-P1
		N -A -A -D -	-Ahx-Ahx	-HHHHHH-t	1539.6	RP2-P1'
		A -A -D -	-Ahx-Ahx	-HHHHHH-t	1425.5	RP2-P2'
		A -D -	-Ahx-Ahx	-HHHHHH-t	1354.5	RP2-P3'
		D -	**-Ahx-Ahx**	**-HHHHHH-t**	**1283.4**	***RP2-P4'**
			Ahx-Ahx	-HHHHHH-t	1168.3	RP2-AP
RPA	Abu-Ahx-	-V -P -L -S -L -T -M -G -	-Ahx-Ahx	-HHHHHH-t	2391.9	
		L -T -M -G -	-Ahx-Ahx	-HHHHHH-t	1570.8	RPA-P1'
		T -M -G -	-Ahx-Ahx	-HHHHHH-t	1457.6	RPA-P2'
		M -G -	-Ahx-Ahx	-HHHHHH-t	1356.5	RPA-P3'
		G -	**-Ahx-Ahx**	**-HHHHHH-t**	**1225.3**	**RPA-P4'**
			Ahx-Ahx	-HHHHHH-t	1168.3	*RPA-AP
RPB	Abu-Ahx-	-I -P -V -S -L -R -S -G -	-Ahx-Ahx	-HHHHHH-t	2402.7	
		L -R -S -G -	-Ahx-Ahx	-HHHHHH-t	1617.7	RPB-P1'
		R -S -G -	-Ahx-Ahx	-HHHHHH-t	1504.6	RPB-P2'
		S -G -	-Ahx-Ahx	-HHHHHH-t	1312.4	RPB-P3'
		G -	**-Ahx-Ahx**	**-HHHHHH-t**	**1225.3**	**RPB-P4'**
			Ahx-Ahx	-HHHHHH-t	1168.3	*RPB-AP*

The predicted cleavage sites of reporter peptide (RP)1 and RP4 for the cancer procoagulant, RPA for MMP2 and RPB for MMP7 are underlined (41). The systematic ladder-like degradation of RPs is shown. The dominating fragments of MS spectra after incubation of RPs for 16 h 37˚C in serum specimens are indicated with bold letters. (see Fig. 6). *Anchor-peptide with highest signal intensity.

software suite. For unsupervised hierarchical clustering, the software programs, Cluster and TreeView, were used (23).

Results

Optimization of RPs for CP with PS-SCL screening. The RP with the cleavage site, WKPYDAAD, was processed by the tumor-associated protease CP (EC 3.4.22.26) and has previously been described in the endoprotease profiling of serum specimens from CRC patients (6). The protease-substrate interaction can further be optimized by the use of synthetic peptide libraries (15). In order to incorporate diversity, amino acids of the original sequence were permuted in

positions P2, P1 and P1' as described in Materials and methods (Table I).

The screening of the PS-SCL showed relatively homogeneous results for the HT-29 cell culture supernatant that has a high procoagulant activity (24) (Fig. 1). Most RPs from group 1 with proline at position P2 (Table I) were cleaved between position P1 and P1' at the predicted site for the cysteine protease CP (10). Specifically, the RPs, RP1, RP4 and RP11, were cleaved in culture supernatant of HT-29 cells (Fig. 1A). In contrast, the exchange of the amino acid proline (P) versus valine (V) at position P2 of the non-prime site (group 2, Table I) resulted in a shift of the cleavage site, and the RPs were cut between P2 (V) and P1 (Fig. 1B). This suggests
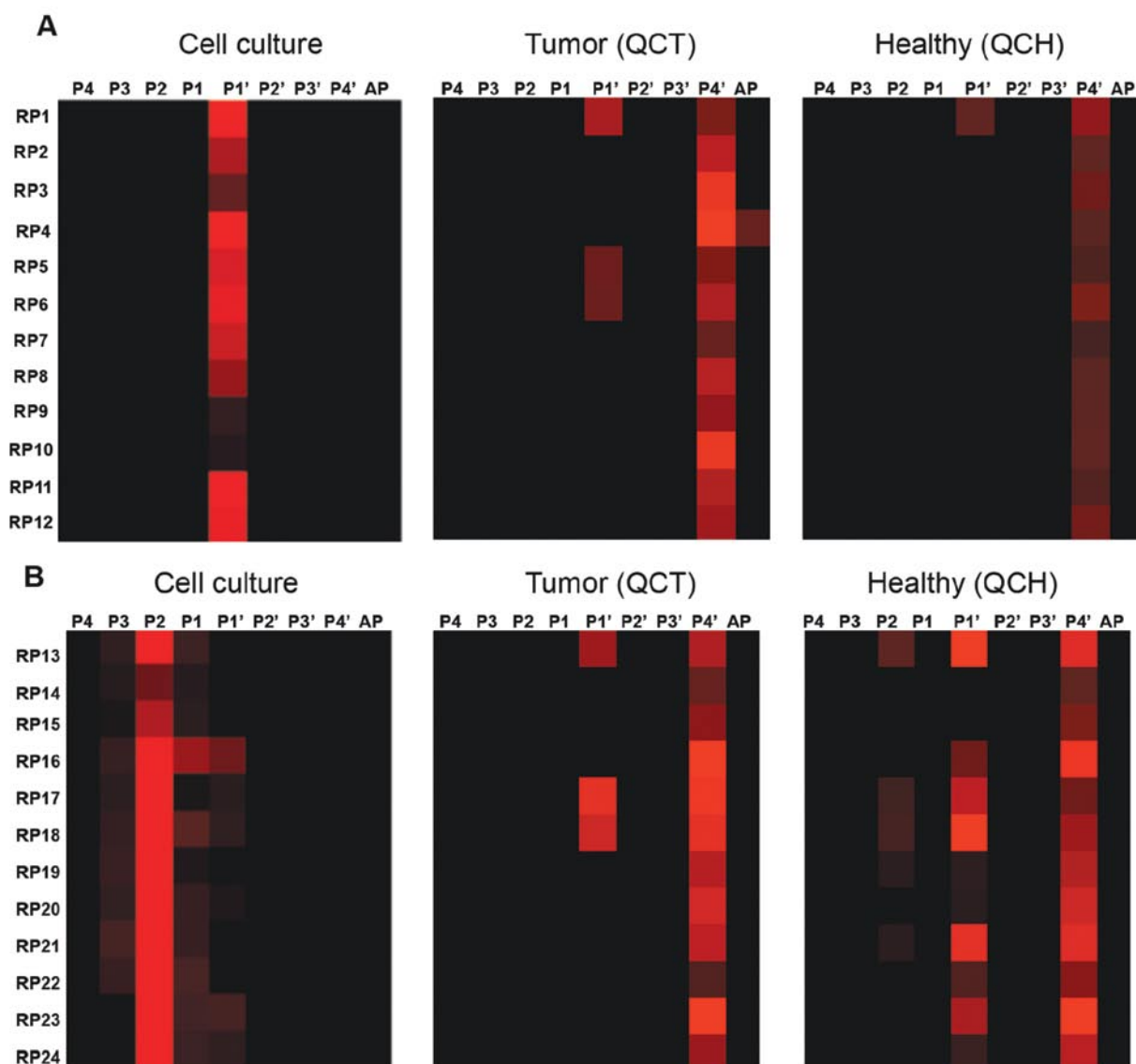
Figure 1. Processing of 24 RPs from the PS-SCL in different specimens. RPs were incubated with cell culture supernatant from HT29 cells, and serum from CRC patients (QCT) and healthy controls (QCH) for 2 h at 37°C prior to MALDI-TOF-MS. The signal intensities were visualised with TreeView software (23) and red colour indicates high signal intensity of the respective proteolytic fragments. (A) RPs RP1-12 (group 1, see Table I). (B) RPs RP13-24 (group 2, see Table I). AP, anchor peptide (see Table II).

that the endoprotease processing the RPs from group 2 is different from CP.

However, the exoprotease content of cell culture supernatant from CRC cells is limited (24,25), leaving the once-cleaved fragments stable. In contrast, serum comprises of a multitude of different proteases (9) and RPs that are first cleaved by endoproteases are further degraded by a multitude of exo-proteases resulting in a 'ladder-like' degradation (6). The prolonged incubation of RPs results in an accumulation of protease-resistant C-terminal fragments that are designated as anchor peptides (6). Accordingly, the fragment, D-Ahx-Ahx-HHHHHH-t (m/z 1283), is the predominant signal that resulted from the cleavage of RPs from group 1 after 2 h (Fig. 1A) and up to 16 h of incubation (Fig. 6). Specifically, RP4 was cleaved in the pooled serum specimen of tumor patients and the signal intensity of the anchor peptide (m/z 1283) originated from RP4 was higher when compared to RP1 or RP11 (Fig. 1A). Most importantly, the proteolytic

processing of RP4 in the pooled serum from healthy controls was weak and clearly different to the strong signal of the tumor specimen (Fig. 1A).

In contrast, the RPs from group 2 did not show different patterns of proteolytic processing when pooled serum specimens of tumor patients and healthy controls were compared (Fig. 1B). Furthermore, the cleavage of these RPs did not only result in the accumulation of anchor peptides, but also longer fragments wer seen, making the interpretation of results more difficult.

In summary, there is one promising RP (RP4) that showed increased signal intensities in pooled serum specimens from tumor patients when compared to healthy controls and therefore, this was selected for further validation studies.

Kinetics of proteolytic processing for RPs, RP1 and RP4, in serum specimens. The RP, RP4, as the most promising candidate from the PS-SCL screening and the original substrate,
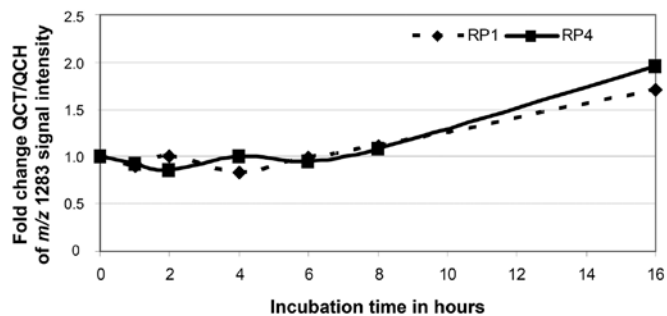
Figure 2. Determination of optimal incubation time for RPs. The RPs, RP1 and RP4, were incubated in pooled serum specimens from tumor patients (QCT) and healthy controls (QCH) for different times as indicated in the graph prior to MALDI-TOF-MS. The signal intensity of the anchor peptide (m/z 1283.4) (see Table II) was extracted from each spectrum and the mean values were calculated for the triplicate measurements. For each RP the ratio of m/z 1283.4 signal intensity in the QCT specimens was divided with the corresponding values from the QCH sample for any incubation time. A fold change of 1 indicates no different processing of the RPs in the QCT and QCH specimens. A fold change of >1 indicates an enhanced cleavage of RPs in the QCT specimens when compared to the QCH samples.

Figure 3. Reproducibility of RP spiking. We incubated 100 μmol/l of the full-length RPs, RP1 and RP4, with pooled serum specimens from tumor patients (QCT) and healthy controls (QCH) for 16 h at 37°C prior to MALDI-TOF-MS. Analyses were serially performed ten times for each combination of RP and type of specimen. The distribution of signal intensity of the peak (m/z 1283.4) for repetitive analyses is shown. The line inside each box represents the median; the open square inside each box represents the mean; the limits of each box represent the 25th and 75th percentiles, and the whiskers represent the minimum and maximum values.

RP1, were chosen for further protease profiling experiments. The RPs were incubated in serum specimens for different times ranging from 0 to 16 h in order to determine the optimal incubation time. Two pooled serum specimens, one from CRC patients (QCT) and the other from healthy controls (QCH) were spiked with RP1 and RP4 and MALDI-TOF-MS analyses were performed in triplicate for each time-point. Signal intensities of the peptide fragment, D-Ahx-Ahx-HHHHHHt (m/z 1283), were extracted from the MS spectra. The ratio of the mean signal intensity of the m/z 1283 peak in tumor and control samples (QCT/QCH) was calculated for each incubation time-point (Fig. 2). In the QCT and QCH specimens, no differences were observed in the proteolytic processing of RP1 or RP4 for up to 8 h of incubation. In contrast, after 16 h of incubation there was a 2-fold stronger proteolytic processing of RP1 and RP4 in the pooled serum specimen from tumor patients (QCT). Accordingly, incubation times of 16 h were chosen for all further experiments.

Although the QCT/QCH-ratios for signal intensities of the m/z 1283 peak resulting from the cleavage of RP1 and RP4 were almost identical, we noted that RP4 processing gave consistently higher signals when compared to RP1. After 16 h of incubation of RP4, the mean signal intensities of the proteolytic product (m/z 1283) were 6.983 arbitrary units (a.u.) and 3.567 a.u. for the QCT and QCH specimens, respectively. In contrast, the incubation of the same serum specimens with RP1 resulted in mean signal intensities of 2.357 a.u. for QCT specimens and 1.383 a.u. for QCH specimens (Fig. 3). In general, higher signal intensities are favourable for the good reproducibility of MALDI-TOF-MS measurements as demonstrated below.

Reproducibility of RP spiking. To test the reproducibility of RP spiking, we analyzed two serum samples (QCT, QCH) ten times with independent sample preparation, MALDI-TOF-MS measurement and data analysis for every specimen. The RPs,
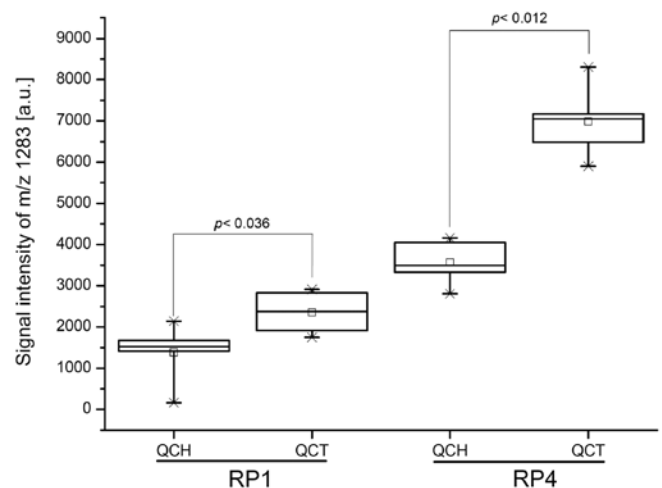
RP1 and RP4, were added to the serum specimens and incubated for 16 h as described above.

The CV for reproducibility of RP1 spiking was 53.3% (mean 1.383 a.u., SD 787) for the QCH sample and 22.2% (mean 2.357 a.u., SD 523) for the QCT sample. In contrast, the reproducibility of RP spiking with RP4 showed higher signal intensities with better reproducibility. The CV for reproducibility of RP4 spiking was 15.5% (mean 3.567 a.u., SD 553) for the QCH sample and 12.8% (mean 6.983 a.u., SD 895) for the QCT sample. The QCT samples could clearly be separated from QCH samples by the use of RP4 as well as RP1 spiking (Fig. 3).

Clinical validation of optimized RP4 in a sample cohort. The RPs, RP1 and RP4, were further validated with serum specimens from 25 CRC patients, 25 patients without malignancy but high CRP levels and 25 healthy controls. The spiked serum specimens were incubated for 16 h at 37°C prior to MALDI-TOF-MS analysis as described in Materials and methods.

For both RPs (RP1 and RP4) the signal intensities of the corresponding anchor peptides (m/z 1283) were significantly higher in the tumor specimens when compared to the healthy controls (Fig. 4). For RP1, the mean value of signal intensity for the m/z 1283 peak was 3.486 a.u. in the specimens from healthy controls and 2.59-fold higher (9.041 a.u.) in the specimens from CRC patients. Accordingly for RP4, the mean value was 7.974 a.u. in the specimens from healthy controls and 2.78-fold higher (22.187 a.u.) in the specimens from CRC patients. However, only RP4 but not RP1 could additionally discriminate serum specimens from tumor patients and patients without malignant disease but high CRP levels. The mean value for the signal intensity of the anchor peptide (m/z 1283) originated from RP4 in serum specimens of CRC
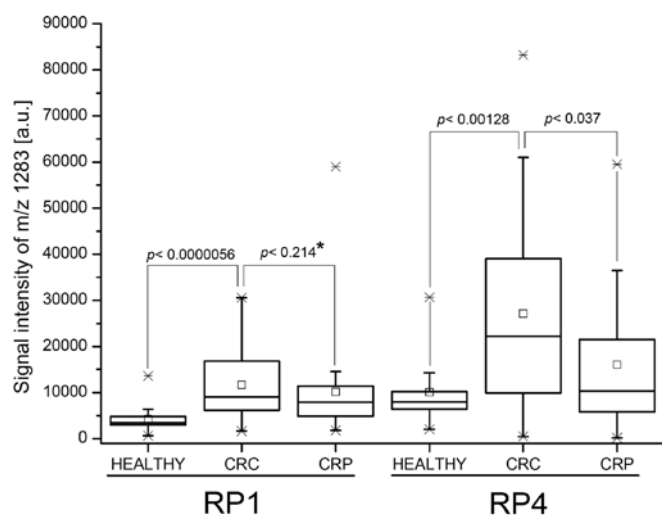
Figure 4. Validation of RP4 in a cohort of clinical material. The RPs, RP1 and RP4, were evaluated with serum specimens from 25 healthy controls, 25 CRC patients and 25 patients with inflammatory disease and high serum CRP levels but without malignant disease. After 16 h of incubation at 37˚C the samples were diluted 1:10 with 0.1% TFA and measured by MALDI-TOF-MS. Statistical analyses were performed with the Mann-Whitney test and statistical significance was assumed for p-values of <0.05. For description of the whisker box plot graph refer to legend to Fig. 3. Asterisks indicate outliers.
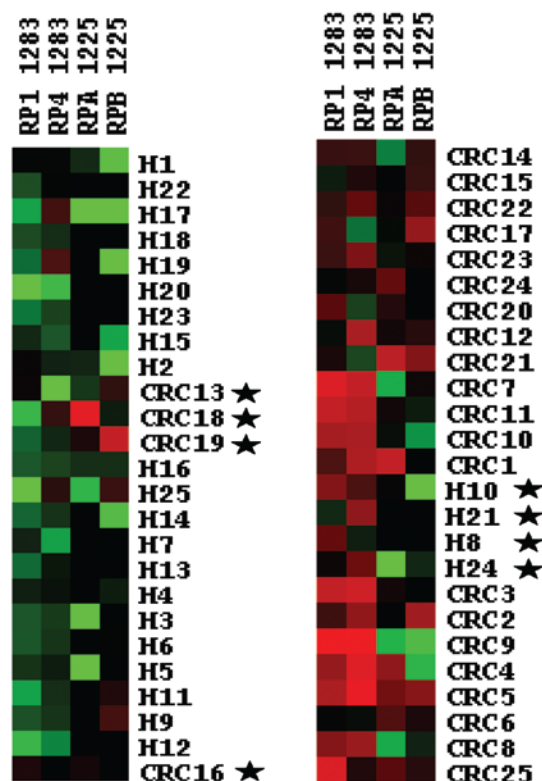

Figure 5. Multiplex analysis of serum with four different RPs. The RPs, RP1, RP4, RPA and RPB, were spiked in serum specimens of 25 colorectal tumor patients and 25 serum specimens from healthy control individuals. After 16 h of incubation at 37˚C the MALDI-TOF-MS analyses were performed as described in Materials and methods. The signal intensities of anchor peptides (m/z 1283 and m/z 1225) originating from the cleavage of RP1/RP4 and RPA/RPB, respectively were extracted from the MS spectra. These data were used to cluster all samples in two different groups with the help of the k-means algorithm in the cluster software (23). Signal intensities are color coded and green/red indicates low/high signal intensities. The left panel comprises mainly of specimens from healthy controls (21/25) and the right panel comprises mainly of samples from CRC patients (21/25). Columns indicate anchor peptides resulting from the proteolytic cleavage of respective reporter peptides. Rows indicate specimens from healthy controls (H) and colorectal tumor patients (CRC). The asterisks mark misclassified specimens.

patients was 22.187 a.u. whereas in serum specimens from patients with non-malignant disease but high CRP levels, the mean value was only 10.382 a.u. and this difference was statistically significant with a p-value of <0.05. In contrast, the mean value for the signal intensity of the anchor peptide (m/z 1283) originated from RP1 in serum specimens of CRC patients was only 9.041 a.u. and statistically not different (p-value of 0.214) to the mean value of 7.925 a.u. in serum specimens from patients with non-malignant disease but high CRP levels (Fig. 4).

Functional protease profiling with different RPs (multiplex analysis). In order to test the hypothesis that the combination of different RPs can improve the diagnostic classification of tumor samples (26), two additional RPs were added to the protease profiling experiments. Besides the RPs, RP1 and RP4, that are cleaved by the cysteine protease 'cancer pro-coagulant', we selected two RPs designated as RPA and RPB (Table II) that are known to be cleaved by the tumor-associated proteases, MMP2 and MMP7, respectively (18). RPA and RPB were also added to the serum specimens of 25 CRC patients and 25 healthy controls, incubated for 16 h and analysed with MALDI-TOF-MS. The signal intensities of the anchor peptides (Table II) were extracted from the MS spectra for ROC curve calculations and multiparametric analyses with Cluster and TreeView software (23). The area under the ROC-curves (AUC) for the single markers ranged from 0.55 (RPB) to 0.86 (RPA). However, the highest AUC value of 0.90 was observed for the combination of all 4 RPs (RP1, RP4, RPA and RPB).

The unsupervised hierarchical clustering resulted in diag-nostic sensitivity and a specificity of 84%, meaning that 4 out of the 25 specimens from the colorectal tumor patients were classified as false-negative and 4 out of the 25 specimens from the healthy controls were classified as false-positive (Fig. 5).

Discussion

The dysregulation of protease activity plays an important role in the initiation and progression of malignant disease (1,2). Tumor-associated proteases that are secreted into the blood-stream are candidates for laboratory diagnostic use (6,7,13). Adding RPs to serum specimens allows for the monitoring of tumor-related proteolytic activity (5-7). Most importantly, the RP spiking can easily be standardized and thereby alleviates major limitations of current MS-based clinical proteomic profiling approaches (4,6,12). However, blood serum has a high intrinsic proteolytic activity (9), and successful protease profiling analyses should therefore depend on the selection of adequate RPs, exclusively cleaved by tumor-associated proteases and not by intrinsic serum proteolytic activities.

In a previous study, we developed a RP with the cleavage motif, WKPYDAAD, for the MS-based detection of the tumor-associated cysteine protease 'cancer procoagulant' (6). The aim of our present study was to further optimize the
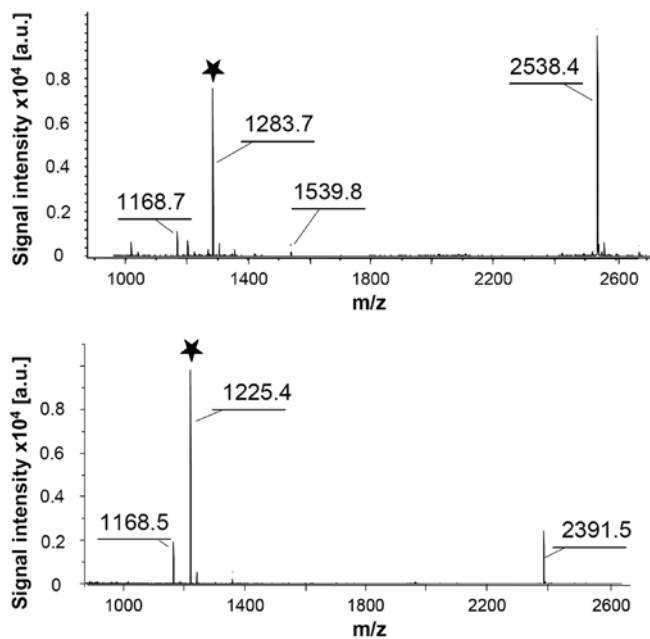
Figure 6. Exemplary MALDI-TOF-MS profiles that illustrate the proteolytical degradation of RPs in serum. The RPs, RP4 and RPA, were incubated for 16 h at 37°C in pooled serum from CRC patients (QCT) prior to MALDI-TOF-MS. Upper graph: m/z 2538.4, Bio-Abu-Ahx-WKPYNAAD-Ahx-Ahx-HHHHHH-t. m/z 1283.7, D-Ahx-Ahx-HHHHHH-t. m/z 1539.8, NDAAD-Ahx-Ahx-HHHHHH-t. m/z 1168.7, Ahx-Ahx-HHHHHH-t. Lower graph: m/z 2391.5, Bio-Abu-Ahx-VPLSLTMG-Ahx-Ahx-HHHHHH-t. m/z 1225.7, G-Ahx-Ahx-HHHHHH-t. m/z 1168.7, Ahx-Ahx-HHHHHH-t (see Table II). Asterisks indicate proteolytic fragment with highest signal intensity.

diagnostic sensitivity and specificity of RP spiking by systematically varying the amino acid sequence of the cleavage motif, WKPYDAAD, in order to improve the classification accuracy of tumor samples. For this purpose, a PS-SCL was constructed. Over the years, a number of studies have demonstrated the power of PS-SCL to effectively optimize protease-substrate interactions, when the proteases are available in highly purified forms (15,27,28). However, the use of PS-SCL has not been reported for the screening of protease activity in clinical specimens so far, in which the enzymatic activities are far more complex. For example, several synthetic peptidyl substrates for the tumor-associated protease CP have been described (29). However, these substrates are also cleaved by other high abundance proteases of serum specimens and are therefore of limited use for diagnostic purpose for detecting CP activity in clinical specimens.

The *in vitro* stability of RPs in human serum specimens from non-malignant controls is a most critical issue and often decreases the diagnostic specificity of functional protease profiling (14). Furthermore, the peptidyl substrates should have a high affinity for tumor-associated proteases to warrant high diagnostic sensitivities. Some of the amino acid substitutions of the cleavage motif, WKPYDAAD, are associated with dramatic effects concerning the processing of the RP in serum specimens from tumor patients and healthy controls. The exchange of proline (P) against valine (V) at position P2 led to a complete loss of diagnostic specificity and the RPs were cleaved in serum specimens from tumor patients and

healthy controls in a nearly identical manner. Proline (P) is a cyclic and hydrophobic amino acid that is disrupting protein conformation whereas valine (V) with its aliphatic side chain is highly hydrophobic (30). It can be concluded that proline at position P2 is essential for the cleavage related to the activity of CP (20). In contrast, valine at position P2 induces the processing of RPs by obviously genuine serum proteases.

When amino acids with similar physicochemical properties were substituted at positions P1 (aromatic amino acids Y, W, and F) and P1' (acidic/polar amino acids D, E, Q, N), the proteolysis rate of RPs, that affects diagnostic sensitivity, could further be improved and the sequence motif, WKPYNAAD (RP4), turned out to be more effective than the original sequence WKPYDAAD (RP1). We were not able to perform enzyme kinetic measurements, as analyses were performed with clinical specimens that comprised of a multitude of different proteases. However, in a cohort of 25 specimens from tumor patients, the proteolytic processing of the RP, RP4, with a median signal intensity of 22.187 a.u. was approximately 2-fold higher than the corresponding value of 9.041 a.u. for the RP, RP1 (Fig. 4). These data indicate that the exchange of aspartic acid (D) against asparagine (N) at the position P1' of the sequence WKPYDAAD induces a faster cleavage by the tumor-associated protease CP. More importantly, the proteolytic processing of the RP, RP4, in serum specimens from tumor patients was also clearly above the level of the median signal intensity of 10.382 a.u. observed in serum specimens from patients with non-malignant but inflammatory disease. In contrast, the proteolytic processing of the RP, RP1, was not statistically different in specimens from CRC patients and patients with inflammatory disease (Fig. 4). The proteolytic activity of serum specimens is not only altered in malignant disease, but also during inflammatory processes (31,32) and it is obvious that specific tumor-associated proteolytic activity could be detected with the improved RP, RP4, but not with RP1.

However, few outliers of the healthy control individuals showed high signal activity for the cleaved RPs, RP1 and RP4, hypothesized to be surrogate markers for CP activity. This, in part, could be due to the fact that CP is not exclusively present in serum specimens from tumor patients, but can also be found in serum specimens from healthy control individuals (33). It could also be possible that the RP can also be cleaved by other endoproteases, thereby resulting in the decreased diagnostic specificity of the test. Furthermore, a recent imaging study showed incidental findings in a substantial number of participants, which resulted in further treatment such as surgery, for a previously undiagnosed tumor (34). This highlights the problem that in a 'healthy' control population, there could be a few cases of undetected malignancy, thereby overestimating diagnostic unspecificity. In contrast, some cancer patients had low signal intensities for the cleaved RPs, RP1 and RP4, indicating no increased activities of CP (Fig. 5), resulting in low diagnostic sensitivities in these cases.

It is likely that tumor heterogeneity and the progression of malignant disease result in different protease patterns (17,35). The use of one single RP is unlikely to be sufficient to ensure high diagnostic accuracy, thus calling for combinations of different RPs in a multiplex fashion. For example, MMP2 and MMP7 are known to play important roles in the progression of CRC (2). Our approach to combine RP1 and RP4, RPA

(MMP2) and RPB (MMP7) resulted in improved AUC for ROC curves, indicating that multiplex analysis has superior diagnostic efficacies (36).

While our approach shows the methodological feasibility of designing protease profiling assays, it requires validation in studies within a clinical setting to comply with recent recommendations (37), before inclusion in diagnostic procedures. Further studies are required in order to focus on the systematic identification of RP sequences that are most efficiently and specifically cleaved by tumor-associated proteases. Furthermore, the characterization and identification of proteases involved in the processing of RPs is also essential for improving sample collection and storage for the preservation of tumor-associated protease activity.

Finally, MALDI-TOF-MS has two serious shortcomings as regards functional protease profiling for diagnostic use. Firstly, the combination of different RPs for multiplex analysis has to be evaluated carefully, as the unwanted phenomenon of ion suppression has to be taken into account (38). Secondly, signal intensities of MALDI-TOF-MS analyses have unsatisfactory interday reproducibility that is hardly acceptable for routine analyses (39) and we therefore suggest, that protease profiling should be evaluated with LC/MS/MS technology (40) which would allow for the exact quantification of multiple analytes within a single run.

Acknowledgements

References

1. Lopez-Otin C and Bond JS: Proteases: multifunctional enzymes in life and disease. J Biol Chem 283: 30433-30437, 2008.
2. Egeblad M and Werb Z: New functions for the matrix metalloproteinases in cancer progression. Nat Rev Cancer 2: 161-174, 2002.
3. Gocheva V, Wang HW, Gadea BB, et al: IL-4 induces cathepsin protease activity in tumor-associated macrophages to promote cancer growth and invasion. Genes Dev 24: 241-255, 2010.
4. Findeisen P, Peccerella T, Post S, Wenz F and Neumaier M: Spiking of serum specimens with exogenous reporter peptides for mass spectrometry based protease profiling as diagnostic tool. Rapid Commun Mass Spectrom 22: 1223-1229, 2008.
5. Villanueva J, Nazarian A, Lawlor K and Tempst P: Monitoring peptidase activities in complex proteomes by MALDI-TOF mass spectrometry. Nat Protoc 4: 1167-1183, 2009.
6. Peccerella T, Lukan N, Hofheinz R, et al: Endoprotease profiling with double-tagged peptide substrates: a new diagnostic approach in oncology. Clin Chem 56: 272-280, 2010.
7. Dekker LJ, Burgers PC, Charif H, et al: Differential expression of protease activity in serum samples of prostate carcinoma patients with metastases. Proteomics 10: 2348-2358, 2010.
8. Somiari SB, Somiari RI, Heckman CM, et al: Circulating MMP2 and MMP9 in breast cancer - potential role in classification of patients into low risk, high risk, benign disease and breast cancer categories. Int J Cancer 119: 1403-1411, 2006.
9. Wildes D and Wells JA: Sampling the N-terminal proteome of human blood. Proc Natl Acad Sci USA 107: 4561-4566, 2010.
10. Rawlings ND, Morton FR, Kok CY, Kong J and Barrett AJ: MEROPS: the peptidase database. Nucleic Acids Res 36: D320-D325, 2008.
11. Witt I: Test systems with synthetic peptide substrates in haemostaseology. Eur J Clin Chem Clin Biochem 29: 355-374, 1991.
12. Findeisen P, Post S, Wenz F and Neumaier M: Addition of exogenous reporter peptides to serum samples before mass spectrometry-based protease profiling provides advantages over profiling of endogenous peptides. Clin Chem 53: 1864-1866, 2007.
13. Villanueva J, Nazarian A, Lawlor K, Yi SS, Robbins RJ and Tempst P: A sequence-specific exopeptidase activity test (SSEAT) for 'functional' biomarker discovery. Mol Cell Proteomics 7: 509-518, 2008.
14. Tempst P: Response to 'Letter to the Editor' by Diamandis EP, Kulasingam V and Sardana G. J Clin Invest eletter: http://www.jci.org/eletters/view/26022, 2006.
15. Schneider EL and Craik CS: Positional scanning synthetic combinatorial libraries for substrate profiling. Methods Mol Biol 539: 59-78, 2009.
16. Schechter I and Berger A: On the size of the active site in proteases. I. Papain. Biochem Biophys Res Commun 27: 157-162, 1967.
17. Lin TS, Chiou SH, Wang LS, et al: Expression spectra of matrix metalloproteinases in metastatic non-small cell lung cancer. Oncol Rep 12: 717-723, 2004.
18. Xia Z, Xing Y, So MK, Koh AL, Sinclair R and Rao J: Multiplex detection of protease activity with quantum dot nanosensors prepared by intein-mediated specific bioconjugation. Anal Chem 80: 8649-8655, 2008.
19. Beyer M, Nesterov A, Block I, et al: Combinatorial synthesis of peptide arrays onto a microchip. Science 318: 1888, 2007.
20. Mielicki WP: Biochemistry of cancer procoagulant. Haemostasis 31 (Suppl 1): 8-10, 2001.
21. Matharoo-Ball B, Ratcliffe L, Lancashire L, et al: Diagnostic biomarkers differentiating metastatic melanoma patients from healthy controls identified by an integrated MALDI-TOF mass spectrometry/bioinformatic approach. Proteomics Clin Appl 1: 605-620, 2007.
22. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P and Witten IH: The WEKA Data Mining Software: An Update; SIGKDD Explorations. Vol. 1, Issue 1, 2009.
23. Eisen MB, Spellman PT, Brown PO and Botstein D: Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95: 14863-14868, 1998.
24. Amirkhosravi A, Biggerstaff JP, Warnes G, Francis DA and Francis JL: Determination of tumor cell procoagulant activity by Sonoclot analysis in whole blood. Thromb Res 84: 323-332, 1996.
25. Volmer MW, Stuhler K, Zapatka M, et al: Differential proteome analysis of conditioned media to detect Smad4 regulated secreted biomarkers in colon cancer. Proteomics 5: 2587-2601, 2005.
26. Kozlov IA, Melnyk PC, Hachmann JP, et al: A high-complexity, multiplexed solution-phase assay for profiling protease activity on microarrays. Comb Chem High Throughput Screen 11: 24-35, 2008.
27. Backes BJ, Harris JL, Leonetti F, Craik CS and Ellman JA: Synthesis of positional-scanning libraries of fluorogenic peptide substrates to define the extended substrate specificity of plasmin and thrombin. Nat Biotechnol 18: 187-193, 2000.
28. Furlong ST, Mauger RC, Strimpler AM, Liu YP, Morris FX and Edwards PD: Synthesis and physical characterization of a P1 arginine combinatorial library, and its application to the determination of the substrate specificity of serine peptidases. Bioorg Med Chem 10: 3637-3647, 2002.
29. Mielicki WP, Mielicka E and Gordon SG: Cancer procoagulant activity studies using synthetic peptidyl substrates. Thromb Res 87: 251-256, 1997.
30. Betts MJ and Russell RB (eds): Amino acid properties and consequences of subsitutions. In: Bioinformatics for Geneticists. Wiley, New Jersey, pp289-316, 2003.
31. Sharony R, Yu PJ, Park J, Galloway AC, Mignatti P and Pintucci G: Protein targets of inflammatory serine proteases and cardiovascular disease. J Inflamm 7: 1-17, 2010.
32. Consolo M, Amoroso A, Spandidos DA and Mazzarino MC: Matrix metalloproteinases and their inhibitors as markers of inflammation and fibrosis in chronic liver disease (Review). Int J Mol Med 24: 143-152, 2009.
33. Gordon SG and Cross BA: An enzyme-linked immunosorbent assay for cancer procoagulant and its potential as a new tumor marker. Cancer Res 50: 6229-6234, 1990.

34. Orme NM, Fletcher JG, Siddiki HA, *et al*: Incidental findings in imaging research: evaluating incidence, benefit, and burden. Arch Intern Med 170: 1525-1532, 2010.
35. Bok RA, Hansell EJ, Nguyen TP, Greenberg NM, McKerrow JH and Shuman MA: Patterns of protease production during prostate cancer progression: proteomic evidence for cascades in a transgenic model. Prostate Cancer Prostatic Dis 6: 272-280, 2003.
36. Faraggi D and Kramar A: Methodological issues associated with tumor marker development. Biostatistical aspects. Urol Oncol 5: 211-213, 2000.
37. Mischak H, Apweiler R, Banks R, *et al*: Clinical proteomics: A need to define the field and to begin to set adequate standards. Proteomics Clin Appl 1: 148-156, 2007.
38. Schlosser G, Pocsfalvi G, Huszar E, Malorni A and Hudecz F: MALDI-TOF mass spectrometry of a combinatorial peptide library: effect of matrix composition on signal suppression. J Mass Spectrom 40: 1590-1594, 2005.
39. Szajli E, Feher T and Medzihradszky KF: Investigating the quantitative nature of MALDI-TOF MS. Mol Cell Proteomics 7: 2410-2418, 2008.
40. Kuzyk MA, Smith D, Yang J, *et al*: Multiple reaction monitoring-based, multiplexed, absolute quantitation of 45 proteins in human plasma. Mol Cell Proteomics 8: 1860-1877, 2009.
41. Crooks GE, Hon G, Chandonia JM and Brenner SE: WebLogo: a sequence logo generator. Genome Res 14: 1188-1190, 2004.