A High-Complexity, Multiplexed Solution-Phase Assay for Profiling Protease Activity on Microarrays

Igor A. Kozlov, Peter C. Melnyk, John P. Hachmann, Anu Srinivasan, Melissa Shults, Chanfeng Zhao, Joseph Musmacker, Nicholas Nelson, David L. Barker and Michal Lebl*

Illumina, Inc., 9885 Towne Centre Drive, San Diego, CA 92121-1975, USA

Abstract: We have developed a miniaturized and multiplexed solution assay for the measurement of protease activity in complex samples. This technology can accelerate research in functional proteomics and enable biologists to carry out multiplexed protease inhibitor screens on a large scale. The assay readout is based on Illumina's universal Sentrix® BeadArrays. The peptide sequences that serve as protease substrates are conjugated to oligonucleotide sequences complementary to the oligo tags on randomly assembled and decoded bead arrays. The peptide portion is C-terminally labeled with a biotin residue and contains a sequence of five histidine residues on the amino terminus. The unique oligonucleotide part of each oligonucleotide-peptide conjugate is attached to amino terminus of the peptide sequence. Upon protease cleavage, the biotin residue is cleaved from the oligonucleotide-peptide conjugate. Following the reaction, all biotin-containing species are captured and removed by incubation with streptavidin beads. The cleaved conjugates that remain in solution are captured by hybridization of their oligo sequence to Sentrix BeadArrays and detected using a labeled antibody against pentahistidine tag of the conjugate or by an antibody sandwich assay. We have generated multiple sets of oligonucleotide tagged peptide substrates of varying complexity (100 to 1000 substrates in a mixture) and show that the response of individual substrate is independent of the complexity of the mixture. Our initial results demonstrate the feasibility of assaying proteases in a multiplexed environment with high sensitivity.

INTRODUCTION

The study of proteomics [1] is vital to understanding life processes and will create new opportunities for the design and development of the next generation of pharmaceuticals and diagnostics. Proteases, subjects of "degradomics" [2], cover about 3% of the proteome and play vital roles in regulating biological processes – metabolism, hormonal and immune response, homeostasis, neurochemistry, viral replication, apoptosis, and many others. Therefore, they are important targets for the development of diagnostics or novel drugs. An example of peptide use in diagnostics is the discovery of fingerprint patterns of peptides generated by protease activity in *ex vivo* experiments specifically defining several cancer types in clinical blood samples [3, 4]. As an example of peptide-based therapeutics, we can mention here HIV protease inhibitors [5-7] which have been successfully developed as drugs.

The first step in the design of a protease assay is finding optimal and specific substrates for the particular protease. There are multitudes of assay formats already available – assays based on the chromatographic or electrophoretic evaluation of a reaction mixture, assays based on mass spectrometric analysis, or assays utilizing radioactivity readouts (see e.g. [8-10]). The most efficient and most widely-used assays are based on fluorescent and/or fluorogenic substrates, especially substrates utilizing fluorescent resonance energy transfer (FRET). FRET has been applied to substrate evaluation in a library format by Meldal *et al.* [11] (for review of FRET-based technologies see [12-14]). However, most of the known assays are tedious, expensive, and very difficult to multiplex. Multiplexing is the key to a high-throughput, reliable and economical assay.

It is always desirable to carry out assays in a format as close to the physiological system as possible. Furthermore, performing the assays in very high complexity and minimal volume will minimize cost and maximize throughput. Unfortunately, these two requirements are sometimes incompatible. Microarray-based technologies are considered the most promising technologies for the study of proteomics (see e.g. [15-38]), because they are economical and high-throughput. However, data from solid-phase assays often tend to be markedly different from data generated in solution-phase assays. We have merged these two requirements to create a potentially optimal solution. The biological reaction is performed in solution, and readout of results is made on solid-phase using microarrays. To be able to utilize the ultra-high-throughput microarrays of Illumina, substrates must be encoded for subsequent deconvolution. A similar approach has been applied recently for protease activity analysis utilizing PNA-encoding [17-19]. We have developed a DNA-tag based microarray readout assay for both kinase [39] and protease activity determination. Our protease assay utilizes biotin depletion (see e.g. [40]) to decrease the assay background. Biotin depletion is also the basis of our single-well single-substrate assay described earlier [41].

Our assay readout is based on the Illumina BeadArray™ platform, which employs silica beads derivatized with specific oligonucleotide sequences and randomly assembled in micro-wells etched on fiber optic bundles [42, 43]. These microarrays have been successfully applied to a variety of biological assays, such as SNP-genotyping and gene expression (reviewed in [16, 44, 45]). The high density of these arrays allows multiple analyses to be performed from a sin-

*Address correspondence to this author at the Illumina, Inc., 9885 Towne Centre Drive, San Diego, CA 92121-1975, USA;
E-mail: mlebl@illumina.com

1386-2073/07 $50.00+.00 © 2007 Bentham Science Publishers Ltd.

gle sample in a very small volume. The presence of multiple copies of each bead type on the microarray enables signal averaging and consequently, reduces assay variability.

This paper describes application of the BeadArray™ platform for readout of a multiplexed protease assay performed entirely in solution. Array readout is enabled by labeling the substrate peptide with a DNA tag, allowing association of peptide structure with DNA sequence. A universal pentahistidine label serves as the visualization tag, *via* an appropriate fluorescently-labeled antibody. The fact that this assay is performed in solution with high multiplexing, combined with the extremely small consumption of substrate solution, makes this assay very economical and therefore attractive for a multiplicity of biological studies involving proteases (association studies, evaluation of drug safety, development of novel diagnostics, etc.).

MATERIALS AND METHODS

Synthesis of Peptide-DNA Conjugates

Fmoc amino acids, DIC, BOP reagent, and biotin resin (0.29 mmol/g) were purchased from Novabiochem (EMD Biosciences, San Diego, CA). Solvents were from VWR International (West Chester, PA). 4-Methylpiperidine was purchased from Sigma-Aldrich (Milwaukee, WI, USA).

During the development of our high throughput protease profiling assay, a custom built high throughput peptide synthesizer was utilized that is capable of simultaneous synthesis of 768 peptides in one batch in a flat-bottom polypropylene microtiter plates [46, 47]. Synthesis of peptide substrates was performed on biotin-modified resin (Fig. **1**). Peptides were synthesized using Fmoc chemistry and DIC or BOP as the coupling reagent and 4-methylpiperidine as the deprotection reagent [48, 49].

Biotin resin (300 mg) was added into a mixture of DMF and DCM (10 ml total) to form a nonsedimenting suspension which was distributed into the wells of one flat bottom polypropylene microtiterplate (Evergreen Scientific, Los Angeles, CA). Plates were placed into centrifugal synthesizer. Additional DMF was added into the plate wells (beads sedimented) and the plate was centrifuged with a tilt of 6 degrees. A standard synthetic protocol was used for Fmoc protecting group removal, with the exception that 4-methylpiperidine was used instead of piperidine [49]. Individual Fmoc protected amino acids (0.3 M solution in 0.3 M HOBt in DMF) and a solution of DIC (1 M in DMF) were pipetted to the wells. The plate was oscillated five times and rested for 50 s. During oscillation, the plate was rotated at a speed at which the liquid did not overflow the wall of the well and the solid support moved towards the outer side of the well. When the rotation was stopped, liquid returned to the horizontal position and the beads distributed at the well bottom, thereby mixing the contents of the well. This procedure was repeated 30 times. The plate was centrifuged, and the addition of amino acids and reagents was repeated. After another 30 cycles of oscillation and resting, the reagents were removed by centrifugation. Washing and deprotection was repeated to prepare the plate for the next cycle of synthesis.

The amino terminus of each peptide molecule contained a sequence of five histidines and hydroxylaminoacetic acid (Ham). This sequence was assembled either by stepwise synthesis or by coupling presynthesized block Boc-Ham-His(Trt)-His(Trt)-His(Trt)-His(Trt)-His(Trt)-Gly-OH. At the end of the synthesis, plates were dried in vacuo and 150 μL of mixture K [50] (TFA/thioanisol/water/phenol/EDT: 82.5:5:5:5:2.5 v/v) was added. Plates were capped and shaken on a plate shaker for 3 h. The suspension was transferred using a multichannel pipettor into filter plates (Orochem Technologies, Lombard, IL). The eluate in the deep well plate (VWR) was precipitated with ether (600 μL), and after refrigeration for 2-h, the plate was centrifuged. The supernatant was removed using a surface suction device [51], and the pellet was resuspended in ether (600 μL) and centrifuged once again. The ether wash and centrifugation step were repeated three times. The product was dried in a Speedvac (ThermoSavant, Waltham, MA), dissolved in 200 μL of water, or 50% dimethylsulfoxide (DMSO)/50% water. A small aliquot of the sample was diluted 1:10 in water and used for HPLC analysis (Waters, Milford MA, μBondapak, C18, 10 μm particle, 125 Å pore, 3.9 x 150 mm, gradient 0,05% trifluoroacetic acid in water to 70% acetonitrile, 0.05% trifluoroacetic acid in 15 min, flow rate, 1.5 mL/min, detection by UV at 217 nm). Peptide purity was further tested by mass spectrometric analyses, performed at HT-Labs (San Diego, CA).

The tag DNA part of the molecule was synthesized utilizing the high-throughput oligonucleotide synthesizers developed by Illumina [46, 52]. Oligonucleotides were modified by attachment of either an amino linker at 5' end of the DNA molecule, followed by condensation of 4-carboxybenzaldehyde, or by coupling a phosphoramidite containing a formylindole group (Link Technologies, Bellshill, Scotland) as the last building block in DNA synthesis.

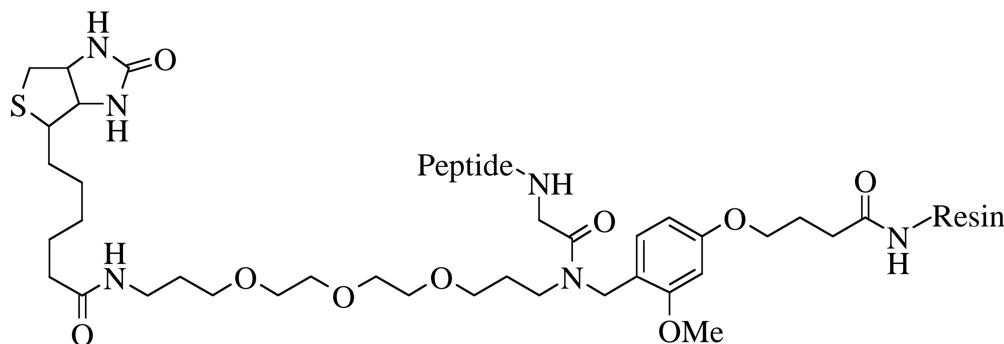

Fig. (1). Biotin tag modified resin used for the synthesis of peptide substrates.

Hydroxylamine-modified peptide substrates and aldehyde-modified oligonucleotides were incubated in 0.2 M sodium citrate buffer. pH 5.0, to form an oxime linkage. We have evaluated alternative chemistries for generating peptide-DNA conjugate and found that the oxime linkage resulting from the use of aminooxyacetic acid was the most stable [53].

Purification of Peptide-DNA Conjugates

First, 192 individual conjugates were purified by using gel electrophoresis and then characterized to create a pool. However, individually purifying each of 1000+ peptide-DNA conjugates that go into a 1000-plex substrate pool is an untenable task. Therefore, after individual peptide-DNA conjugation reactions were carried out, the reactions were quenched with 8 M urea, and the products were mixed in batches and purified by PAGE and pooled. The peptide-DNA conjugates migrate slower than the DNAs alone and therefore are amenable to gel purification (Fig. **2**). The advantage of this method is that purification of conjugates is greatly simplified. The disadvantage of this method is that the individual concentration of each of the peptide-DNA conjugates in the pool is unknown. This can potentially be a serious problem downstream for comparing substrates because not all peptide-DNA conjugated may be in comparable concentrations. One solution is to hybridize the entire pool to the DNA microarray and measure the maximum possible signal that could be generated from each conjugate. This signal would automatically account for both differences in concentrations of the conjugates in the pool and also differences in hybridization efficiency for the different DNA oligonucleotides.

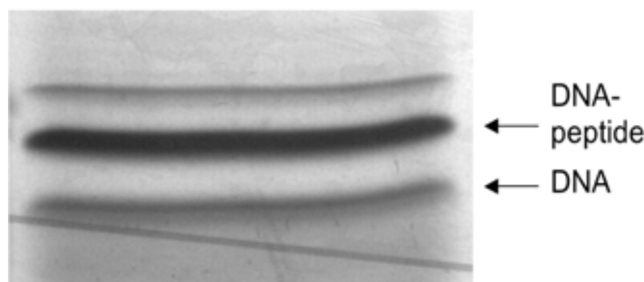

Fig. (2). Preparative purification of a pool of peptide-oligonucleotide conjugates.

Enzymatic Assay

Following the batch purification of the peptide-DNA conjugates, the pools were aliquoted and stored such that the concentration of each substrate in the master pool would be approximately 5 nM (based on spectrophotometeric measurements). Prior to an assay, the master substrate pool was diluted 1:10 (final ~ 0.5 nM per substrate) in the appropriate buffer for the enzymes being assayed. The buffers used for substrate dilutions were, for caspases, 0.7M sodium citrate, 100 mM sodium chloride, 0.1% CHAPS, 1 mM EDTA, and 1 mM DTT, pH 7.2; and for all other proteases, 100 mM Tricine, pH 7.4, 100 mM sodium chloride, 10 mM calcium chloride, 50 μM zinc chloride, and 0.05% Brij-35.

For the assay, 20 μL of the diluted protease (at 2X final concentration), or cell extract, and 20 μL of the diluted substrate pool (at 2X final concentration) were mixed so that the final substrate pool concentration was approximately 0.25 nM per substrate of the 1000-plex. In each assay, a negative control and a positive control were included. The negative control comprised 20 μL of the substrate pool mixed 20 μL of the buffer only, i.e. not treated with any protease. The positive control consisted of 20 μL of the substrate pool mixed with 20 μL of a protease mixture containing 40 mM each of trypsin and chymotrypsin. The mixtures were incubated at 37°C for 45 min and then heated to 95°C for 10 min to heat inactivate the proteases.

The amount of substrate used per reaction was 20 μL at 0.25 nM per substrate of a 1000-plex pool, i.e. ~ 5 pmol of biotin. SA- magnetic beads were used at a 20-fold molar excess for the post-assay pulldown of the unwanted biotinylated components, i.e. SA = 100 pmol/reaction. SA-beads were washed once with the assay buffer and resuspended in assay buffer such that each tube would receive the 100 pmol of SA in a 10 μL volume. Upon completion of the protease reaction followed by the heat inactivation, 10 μL of the washed SA-beads were added to every reaction except the positive control. The positive control tube received 10 μL of the assay buffer instead. The samples were agitated with the SA beads for 15 min at room temperature and then transferred to a magnetic plate. The 50 μL supernatants were transferred to new tubes and stored until hybridization.

Hybridization

Sentrix Array Matrices (SAMs) comprising 96-fiber optic bundles, each with 1624 different bead types, were used for hybridization and data measurement. Every bead type of the 1624 contains a unique oligonucleotide sequence. The SAM was removed from the packaging and were washed for 1 min each with 95% formamide and with the hybridization buffer containing 100 mM potassium phosphate, 1 M sodium chloride and 20% formamide, pH 7.6. The samples from the protease assay were each diluted 1:10 in hybridization buffer and 45 μL of the diluted samples were aliquoted into appropriate wells in a 384-well plate for hybridization to one of the 96-bundles of the SAM. Each protease assay product was diluted 1:10 into hybridization buffer and hybridized in triplicate to three separate bundles. The SAM and the hybridization plate were assembled into the controlled-humidity hybridization chamber, and samples were allowed to hybridize overnight at 48°C.

Post-Hybridization Signal Generation

Following the ON hybridization, all washes were carried out in 96-well plates and antibody incubations were done in 384-well plates. The SAM was first washed twice in the 200 μL hybridization buffer and twice in 1XPBS containing 0.1% Tween (1XPBST). The bundles were then submerged in 384-plate wells containing 45 μL each of a solution containing 0.1 μg/ml of mouse-anti-pentaHis-antibody (Qiagen) diluted in 1XPBST containing 0.1% casein (1XPBST+ casein) for 2 h at room temperature (primary antibody). After the primary antibody incubation, the bundles were washed 4-times with 1XPBST, and then the bundles were submerged into 45 μL of the secondary antibody solution, containing 1 μg/ml Alexa555-conjugated goat-anti-mouse IgG in 1XPBST+casein for 1 h at room temperature. Upon completion of the secondary Ab incubation, the bundles were washed twice each in PBST and once with PBS and then

transferred to a new black-bottomed 384-well plate containing PBS for array scanning.

SAM Imaging and Data Analysis

The processed SAM was scanned on the Illumina BeadArray Reader, which measured the fluorescence intensity from each bead on each bundle. The software automatically de-convoluted the signal from the randomly ordered beads and generated a csv file that matched signal intensity with specific bead types on each bundle. Since each bead type contained a unique oligonulceotide, the data were bioinformatically matched to the specific peptide-DNA conjugate. Since hybridizations for every sample, including negative and positive controls, were done in triplicate using three separate bundles, the data for each sample were averaged across the three bundles. Cleavage ratios (CRs) were calculated for each of the 1000-peptide sequences, where CR = [(Assay signal intensity – Negative signal intensity)/(Assay signal intensity – Positive signal intensity)]. In an effort to eliminate any substrates that were poorly represented in the master pool, any peptide-conjugates which showed a positive:negative signal ratio of less than three were not considered. CRs for each of the proteases were sorted in descending order to list the best cleaved substrate for any given protease under the specific conditions of the assay.

RESULTS AND DISCUSSION

Assay Design

The interaction of streptavidin (SA) with biotin-labeled molecules is the basis of many biochemical assays. Our assay is based on the depletion of biotin-containing substrates from solution by incubation with streptavidin immobilized on magnetic beads [40, 54] (Fig. **3**). The carboxy terminus of the peptide sequences are labeled with a biotin moiety. Upon protease cleavage, the biotin-containing part of the molecule is detached from the rest of peptide substrate, leaving the pentahistidine tag and the oligonucleotide tag intact. Following the enzymatic reaction, all biotin-containing species (uncleaved substrates and cleaved carboxy terminal portions of the substrate) are captured and removed by incubation with magnetic streptavidin beads. The cleaved amino terminal part of the substrate remains in solution, and following hybridization to a DNA microarray, the pentahistidine tag can be detected by interaction with the appropriately labeled antibody.

The success of this type of an assay is dependent on (i) the quality of synthesized peptides, and (ii) the complete removal of biotin-labeled molecules from the solution. If the synthetic product were to contain a large proportion of molecules lacking a biotin tag, the background of the assay would

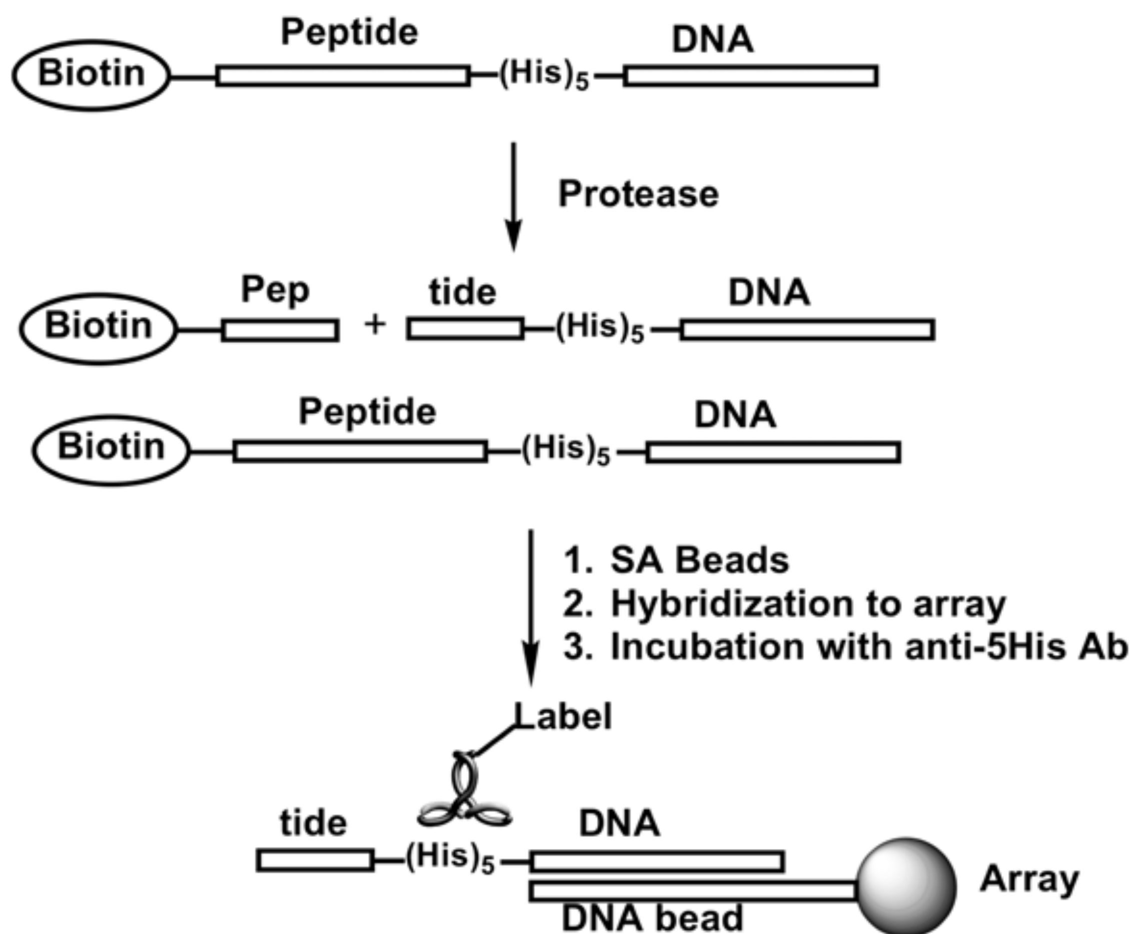

Fig. (3). Schematic depiction of the multiplexed protease assay. The complex peptide-DNA conjugate substrate pool is incubated with the protease of interest in solution. Prior to analysis, the mixture is purified by removing all biotin-containing species with streptavidin magnetic beads. The purified mixture contains only the DNA tags for conjugates which were cleaved. These cleaved substrates are hybridized to a DNA microarray and the penta-histidine tag portion of the peptide substrate is detected by a fluorescently-labeled antibody.

be unacceptably high. We have found that cleavage of the synthetic product from the resin using the cocktail of scavengers ("Mixture K") was cleaving the linkage to biotin as well. However, the extent of this reaction was found to be acceptable, since under 0.2% of the peptide material was missing biotin. The loss of biotin was not observed when TFA with 5% water was used for the deprotection. This condition, however, is not compatible with certain arginine- and tryptophan-containing sequences (see e.g. [55]). In high-throughput synthesis, conditions cannot be individually optimized for different sequences, and therefore, we decided to accept a small loss of biotin tag, resulting in higher quality of synthetic product.

To assess both peptide quality and efficiency of biotin removal, two conditions called "negative" and "positive" controls were included along with the "experimental" condition(s). The "negative" control is not treated with the protease(s). This generates a sample where none of the substrates are cleaved, and all the conjugates still have a biotin on them. Upon SA-magnetic bead pulldown, all of the substrates are pulled out of the solution phase, leaving little or nothing in the supernatant to hybridize to the array. The data from this "negative control" therefore represents the "background" of the assay for each of the substrates. This negative control automatically accounts for any substrates which may have lost their biotin either during peptide synthesis or post-synthesis handling of the peptide. On the other hand, the positive control sample was not treated with the protease(s) and also not subjected to SA pulldown. In this condition, all of the biotinylated conjugates were available for hybridization to the array. In theory, this condition should give the maximum possible signal that can be generated from any given conjugate. The amount of substrate cleaved proteolytically is represented as a cleavage ratio (CR) of cleaved product to the total amount of that particular substrate present in the pool (CR = [(experimental-negative)/(positive-negative)]). For example, if a particular conjugate is completely cleaved, it would give a CR of 1; if only one half of it were cleaved, the CR would be 0.5 and so on. This eliminates the need to know the exact starting concentration of each substrate in the pool.

However, in doing many assays we encountered a recurring problem. The experimental plan and data analysis method described above worked well for certain classes of proteases (e.g., caspases) whose peptide substrates have a preponderance of negatively charged or neutral residues. However, with many of the substrates in the pool, the supposedly maximum possible signal as derived from the positive control condition was often lower than the signal generated from the exact same conjugate in the protease-treated condition. We found that the binding of anti-5His Ab, and thus assay signal generation, was dependent on the peptide sequence. Several peptide sequences appeared to be more amenable to anti-5His Ab binding if they were cleaved by a protease than when not cleaved. Specifically, conjugates whose peptide sequences had positively charged residues (such as lysines and arginines) most often gave higher intensity upon protease cleavage than as a whole. Moreover, we noticed that conjugates with peptides that had a proline residue immediately adjacent to the 5His tag routinely gave poor intensities in all conditions. This presented an undesirable situation because often times the assay would produce data where for some proteases CRs for some of the substrates were >1.

We implemented two solutions that addressed these issues. We surmised that one possible reason for poor positive control intensities with peptides containing positively charged residues could be that the peptides fold back on the oligotag due to the charge attraction, thus blocking the 5His tag. However, when these substrates were cleaved with a protease, the blockage was eliminated, resulting in proper presentation of the 5His-tag for anti-5His Ab binding. Therefore, we altered the assay so that the positive control condition was one where the substrate pool was cleaved with a high concentration mixture of trypsin and chymotrypsin (20 nM each) and hybridized to the array without the SA-pulldown. This variation consistently provided positive control signal intensities that were substantially higher than the equivalent experimental signal.

Additionally, we monitored the positive:negative ratio (P/N) for each of the conjugates; ideally the higher this ratio the better. We only consider conjugates with a P/N > 3 for the data analysis. This eliminates two sets of problem conjugates from consideration – 1) those conjugates that have a proline residue immediately adjacent that the 5His tag and hence routinely gave poor P/N ratios regardless of enzymatic treatment; and 2) those conjugates that are present in very low concentrations in the pool. In the next generation assay, an alternative solution might be to insert a Gly between the His tag and peptide sequence.

Assay Validation

To demonstrate the feasibility of this profiling assay, we have synthesized over 1000 sequences representing substrates for several protease groups (see http://www.5z.com/substrates.pdf). In designing the multiplexed protease assay, we used two sources of structures of protease substrates – (i) the publicly available MEROPS database of characterized proteases (http://merops.sanger.ac.uk/), and (ii) synthetic libraries generated by randomizing critical residues in the consensus substrate sequences for select protease families (caspases and matrix metalloproteases). Several sequences were synthesized in duplicate to show "in synthesis" reproducibility of the synthetic and biological results. All synthetic products were analyzed by HPLC, and some structures were spot checked (about 10%) using mass spectrometry.

The peptide-DNA substrate pools were assayed with a variety of proteases including trypsin, chymotrypsin, endoproteinase Asp-N, endoproteinase Glu-C, endoproteinase Lys-C, endoproteinase Arg-C, thrombin, factor Xa, HRV1, caspases, MMPs, enterokinase, and tobacco etch virus protease. The results are summarized in Table **1**. Use of the "classical" enzymes trypsin and chymotrypsin resulted in predictable cleavage of substrates containing basic or aromatic residues. Further, we were able to distinguish between highly sensitive substrates and define sequences which are less susceptible to the cleavage. Trypsin clearly prefers substrates with arginine over lysine and especially arginine flanked by small side chain amino acids (glycine, alanine and serine). In the top 50 best trypsin substrates, the sequence RG occurred 15 times, RA 13 times, RS 13 times, and GR 18 times. Poor

Table 1. Cleavage Ratios (in %) of Selected Substrates for the Panel of Proteases

CT	TP	Lys-C	Arg-C	Glu-C	Asp-N	HRV1	TEV	EK	FXa	Thr	C3	Sequence	Annotation
99	76	77	27	0	3	4	1	22	11	19	0	LAYGLRSK	M10.005-stromelysin 1 (MMP-3)
71	28	14	10	0	44	16	4	5	27	0	0	YILKRDSY	M16.005-nardilysin
50	87	2	21	0	1	1	0	32	8	0	0	SPFRSSRI	S01.160-kallikrein hK1
48	7	7	40	0	0	10	3	16	16	0	0	GFSPFRGG	M13.002-endothelin-converting enzyme 1
46	42	4	14	0	0	8	2	4	7	0	0	EGRRGAAE	S07.001-flavivirin
43	6	6	1	0	1	1	0	0	1	0	0	PYILKRGS	A01.012-rhizopuspepsin
43	5	29	4	0	9	8	3	0	0	0	0	AVSWLLTA	M10.003-gelatinase A (MMP-2)
42	11	7	33	0	0	14	4	41	21	0	0	NITYRGTG	C02.001-calpain-1
40	1	14	13	0	37	12	6	2	0	0	0	DVAEYSLF	M10.005-stromelysin 1 (MMP-3)
40	76	5	33	0	27	10	2	26	22	2	0	RSYSDRGE	M12.005-procollagen C-proteinase
CT	**TP**	**Lys-C**	**Arg-C**	**Glu-C**	**Asp-N**	**HRV1**	**TEV**	**EK**	**FXa**	**Thr**	**C3**	**Sequence**	**Annotation**
2	**100**	6	69	0	0	8	1	4	41	0	0	GRGHARLV	M12.142-atrolysin A
3	**99**	10	39	0	0	16	4	10	75	8	0	GRASLNHG	M10.005-stromelysin 1 (MMP-3)
4	**98**	7	68	0	0	9	2	5	39	0	0	LGLARGGG	M14.004-carboxypeptidase N
7	**97**	9	63	0	0	16	5	13	30	1	0	GGFARGGG	M14.005-carboxypeptidase E
2	**97**	61	21	0	1	2	0	1	8	0	0	GPLGLKAR	M10.002-collagenase 2 (MMP-8)
1	**92**	13	36	0	42	13	3	17	71	6	0	GLARSNLD	S01.196-complement factor B
2	**89**	7	37	0	0	8	2	5	33	0	0	GHARLVHG	M12.142-atrolysin A
0	**89**	1	34	0	1	2	0	0	5	0	0	SSNQLQRR	M10.010-envelysin
0	**89**	0	78	0	1	0	0	11	30	0	0	RVSRRSRG	A26.001-omptin
26	**87**	2	100	0	0	5	0	4	40	24	0	SGRSPRRFTA	MMP-L95
CT	**TP**	**Lys-C**	**Arg-C**	**Glu-C**	**Asp-N**	**HRV1**	**TEV**	**EK**	**FXa**	**Thr**	**C3**	**Sequence**	**Annotation**
0	2	**100**	8	0	64	9	4	28	0	0	91	DEVDNKVK	C14.003-caspase-3
0	58	**93**	13	0	0	16	6	8	0	0	0	GPGGVHAK	M10.029-matrilysin-2 (MMP-26)
0	2	**90**	17	0	9	8	4	46	0	0	0	TAEEKAAV	A01.009-cathepsin D
0	31	**89**	5	0	6	3	0	2	0	0	0	GAKRGGGG	S01.015-tryptase beta 1 (Homo sapiens)
0	15	**86**	0	0	3	4	0	1	3	0	0	GAAKAAAG	A01.012-rhizopuspepsin
2	82	**82**	28	0	0	11	2	4	0	0	0	KKKRKVGG	S01.146-granzyme K
3	52	**80**	23	0	0	21	3	8	23	90	0	QSPRSFQK	S01.217-thrombin
3	57	**78**	43	0	1	10	4	41	14	0	0	IMAENRKS	M10.005-stromelysin 1 (MMP-3)
1	10	**78**	1	61	2	4	0	4	0	0	0	TLNEKAVK	S08.034-subtilisin BPN'
99	76	**77**	27	0	3	4	1	22	11	19	0	LAYGLRSK	M10.005-stromelysin 1 (MMP-3)
34	62	**76**	36	0	0	12	2	4	0	0	0	SWIKKRQQ	C25.001-gingipain R
CT	**TP**	**Lys-C**	**Arg-C**	**Glu-C**	**Asp-N**	**HRV1**	**TEV**	**EK**	**FXa**	**Thr**	**C3**	**Sequence**	**Annotation**
9	53	6	**100**	0	0	15	4	8	38	0	0	SGRSPSRLTA	MMP-L100
0	89	0	**78**	0	1	0	0	11	30	0	0	RVSRRSRG	A26.001-omptin
2	73	8	**72**	69	0	15	5	13	94	0	0	HRGRTLEI	S01.199-complement factor I
3	55	8	**72**	35	16	13	5	17	15	0	0	EAAEARRG	M12.221-ADAMTS4 endopeptidase
2	100	6	**69**	0	0	8	1	4	41	0	0	GRGHARLV	M12.142-atrolysin A
4	98	7	**68**	0	0	9	2	5	39	0	0	LGLARGGG	M14.004-carboxypeptidase N
16	85	46	**66**	0	2	12	3	94	81	0	0	EERARAKW	C02.001-calpain-1
0	69	6	**65**	0	4	4	1	2	100	0	0	GRGRGGGG	S01.214-coagulation factor IXa
7	97	9	**63**	0	0	16	5	13	30	1	0	GGFARGGG	M14.005-carboxypeptidase E
4	81	48	**53**	0	3	7	1	8	16	0	0	NTKRNRNN	S08.073-proprotein convertase 2

(Table 1) contd….

CT	TP	Lys-C	Arg-C	Glu-C	Asp-N	HRV1	TEV	EK	FXa	Thr	C3	Sequence	Annotation
5	33	11	44	**100**	1	11	4	5	20	0	0	VSQELGQR	M12.221-ADAMTS4 endopeptidase
2	22	10	8	**75**	0	12	3	10	1	0	0	RVSEMEMA	M10.005-stromelysin 1 (MMP-3)
2	73	8	72	**69**	0	15	5	13	94	0	0	HRGRTLEI	S01.199-complement factor I
1	10	78	1	**61**	2	4	0	4	0	0	0	TLNEKAVK	S08.034-subtilisin BPN'
3	0	11	15	**60**	0	10	6	10	0	0	1	LEGSHLVE	S01.131-neutrophil elastase
11	0	11	1	**60**	5	9	1	1	0	0	0	MKYEIASE	M63.001-gpr protease
9	1	9	1	**55**	3	4	2	1	0	0	0	EALYLVAG	Insulin B-chain
1	34	12	47	**50**	4	13	5	5	89	0	0	ELGRGGGG	S01.214-coagulation factor IXa
4	0	8	14	**48**	6	12	5	13	0	0	38	AETFYVDG	A02.001-HIV-1 retropepsin
3	0	64	15	**39**	27	11	5	8	0	0	0	LESDYFGK	C14.001-caspase-1
CT	**TP**	**Lys-C**	**Arg-C**	**Glu-C**	**Asp-N**	**HRV1**	**TEV**	**EK**	**FXa**	**Thr**	**C3**	**Sequence**	**Annotation**
2	0	12	11	0	**100**	12	4	7	0	0	1	ADAHDG	CAS-L3
5	63	8	36	0	**89**	11	3	22	18	0	0	DSRRAQDF	A01.020-phytepsin
0	14	27	7	7	**86**	12	5	8	5	0	0	LDKREAEA	A01.031-yapsin 2
12	4	14	7	7	**86**	12	4	20	3	0	0	AYSDMREA	M10.005-stromelysin 1 (MMP-3)
1	0	10	15	0	**79**	10	4	5	0	0	90	GDEVDGA	C14.003-caspase-3
1	1	47	0	0	**78**	3	1	0	0	0	0	PQGDAAQK	C01.073-endopeptidase 1 (mite)
0	0	6	4	0	**75**	4	1	7	0	0	100	DSVDAKPD	C14.004-caspase-7
3	26	28	17	0	**74**	12	3	13	3	0	0	EDKRHSQG	S08.073-proprotein convertase 2
5	2	14	15	0	**73**	13	5	8	0	0	0	DTDLYDYY	S01.133-cathepsin G
1	0	12	18	0	**72**	11	6	10	0	0	35	DEVDGGGG	C14.003-caspase-3
CT	**TP**	**Lys-C**	**Arg-C**	**Glu-C**	**Asp-N**	**HRV1**	**TEV**	**EK**	**FXa**	**Thr**	**C3**	**Sequence**	**Annotation**
1	2	20	0	0	6	**99**	1	4	0	0	0	PPEELKFQ	M10.008-matrilysin (MMP-7)
5	1	8	1	0	6	**98**	1	2	0	0	0	AIFQGPID	C03.007-rhinovirus picornain 3C
0	0	10	10	0	2	**98**	4	8	0	0	0	TGVSAQVQ	C02.001-calpain-1
4	2	3	1	0	3	**96**	1	0	1	0	0	ALFQGPLQ	C03.001-poliovirus-type picornain 3C
2	0	8	2	14	5	**94**	1	2	0	0	0	LEVLFQGP	HRV 3C Protease
5	1	8	5	30	6	**93**	3	2	0	0	0	YELQGPED	C37.001-calicivirin
2	0	5	2	0	2	**93**	1	1	0	0	0	MHLQGPED	C37.001-calicivirin
7	5	18	1	0	3	**92**	2	1	0	0	0	VYFQGKKN	C04.004-tobacco etch virus NIa protease
2	0	4	4	0	0	**91**	99	4	0	0	0	GENLYFQG	C04.004-tobacco etch virus protease
1	0	6	4	0	0	**90**	2	4	0	0	0	YAPPINHQ	M10.005-stromelysin 1 (MMP-3)
CT	**TP**	**Lys-C**	**Arg-C**	**Glu-C**	**Asp-N**	**HRV1**	**TEV**	**EK**	**FXa**	**Thr**	**C3**	**Sequence**	**Annotation**
2	0	4	4	0	0	91	**99**	4	0	0	0	GENLYFQG	C04.004-tobacco etch virus protease
7	2	6	1	0	4	11	**99**	1	0	0	0	LYFQSGTV	C04.004-tobacco etch virus NIa protease
8	1	6	2	0	24	33	**21**	4	0	0	0	IYTQSLDD	C04.004-tobacco etch virus NIa protease
0	1	3	1	0	10	14	**10**	4	0	0	0	TSEDLVVQ	M10.014-membrane-type matrix metalloproteinase 1
2	63	8	18	19	0	13	**9**	16	35	0	0	IEGRIVEG	S01.216-coagulation factor Xa
3	14	15	17	0	0	18	**8**	10	0	0	0	SGESLRNLTA	MMP-L38
0	0	10	14	0	10	12	**8**	14	0	0	0	QPGGINTS	M10.005-stromelysin 1 (MMP-3)
1	0	8	18	0	41	9	**8**	17	0	0	89	DGPDGPEE	C14.003-caspase-3
5	23	13	18	0	3	12	**7**	8	27	0	0	ARVYGEAL	A02.001-HIV-1 retropepsin
18	8	27	2	0	7	6	**7**	4	0	0	0	LFSQAKIS	C03.005-hepatitis A virus-type picornain 3C

(Table 1) contd….

CT	TP	Lys-C	Arg-C	Glu-C	Asp-N	HRV1	TEV	EK	FXa	Thr	C3	Sequence	Annotation
4	3	8	36	0	51	7	2	**100**	0	0	0	GDRVYIHP	A01.001-pepsin A
16	85	46	66	0	2	12	3	**94**	81	0	0	EERARAKW	C02.001-calpain-1
0	0	42	5	0	4	6	1	**85**	0	0	0	GDDDDKAG	S01.156-enterokinase
26	7	23	52	0	0	9	2	**82**	29	0	0	TFHKAEYR	S01.212-plasma kallikrein
16	22	10	14	0	0	9	2	**77**	2	0	0	GERGFFYT	M04.001-thermolysin
9	10	6	18	0	2	10	4	**70**	0	0	0	EGERGFFY	C25.001-gingipain R
2	0	43	10	0	24	7	3	**67**	0	0	0	DDDKIVGG	A01.011-penicillopepsin
2	2	1	4	0	1	3	1	**67**	0	0	0	VEGERGFF	A01.010-cathepsin E
3	9	5	5	4	5	5	1	**59**	2	0	0	ERGFFYTP	S01.133-cathepsin G
0	49	66	30	7	40	11	4	**53**	11	0	0	GDKELRTG	M10.029-matrilysin-2 (MMP-26)
CT	**TP**	**Lys-C**	**Arg-C**	**Glu-C**	**Asp-N**	**HRV1**	**TEV**	**EK**	**FXa**	**Thr**	**C3**	**Sequence**	**Annotation**
0	69	6	65	0	4	4	1	2	**100**	0	0	GRGRGGGG	S01.214-coagulation factor IXa
2	73	8	72	69	0	15	5	13	**94**	0	0	HRGRTLEI	S01.199-complement factor I
1	34	12	47	50	4	13	5	5	**89**	0	0	ELGRGGGG	S01.214-coagulation factor IXa
16	85	46	66	0	2	12	3	94	**81**	0	0	EERARAKW	C02.001-calpain-1
3	99	10	39	0	0	16	4	10	**75**	8	0	GRASLNHG	M10.005-stromelysin 1 (MMP-3)
1	92	13	36	0	42	13	3	17	**71**	6	0	GLARSNLD	S01.196-complement factor B
3	79	5	40	0	2	4	0	13	**69**	0	0	RAGRSKGT	S08.076-proprotein convertase 5
0	53	6	7	0	1	3	0	4	**57**	0	0	GAIEGRGA	S01.216-coagulation factor Xa
0	67	9	17	0	11	9	4	14	**45**	0	0	GSARAGDG	C25.001-gingipain R
7	67	7	18	0	1	17	3	26	**42**	100	0	IEPRSFSQ	S01.217-thrombin
CT	**TP**	**Lys-C**	**Arg-C**	**Glu-C**	**Asp-N**	**HRV1**	**TEV**	**EK**	**FXa**	**Thr**	**C3**	**Sequence**	**Annotation**
7	67	7	18	0	1	17	3	26	42	**100**	0	IEPRSFSQ	S01.217-thrombin
8	6	1	5	0	2	0	0	8	3	**98**	0	GVPRSFRG	S01.217-thrombin
3	52	80	23	0	0	21	3	8	23	**90**	0	QSPRSFQK	S01.217-thrombin
3	15	6	11	0	11	5	2	4	11	**62**	0	LSPRTFHP	S01.217-thrombin
12	21	6	17	0	3	3	1	8	14	**42**	0	LGIRSFRN	S01.217-thrombin
0	42	4	5	0	4	4	1	4	4	**40**	0	VVPRGVNL	S01.217-thrombin
1	10	2	2	0	1	2	0	0	2	**39**	0	GLVPRGSG	S01.217-thrombin
26	87	2	100	0	0	5	0	4	40	**24**	0	SGRSPRRFTA	MMP-L95
99	76	77	27	0	3	4	1	22	11	**19**	0	LAYGLRSK	M10.005-stromelysin 1 (MMP-3)
1	22	2	4	0	3	2	0	1	12	**18**	0	APRSSMSN	M04.003-vibriolysin
CT	**TP**	**Lys-C**	**Arg-C**	**Glu-C**	**Asp-N**	**HRV1**	**TEV**	**EK**	**FXa**	**Thr**	**C3**	**Sequence**	**Annotation**
0	0	6	4	0	75	4	1	7	0	0	**100**	DSVDAKPD	C14.004-caspase-7
1	0	14	21	0	66	12	6	13	0	0	**99**	DEVDGVDE	C14.003-caspase-3
0	1	7	1	0	22	2	1	2	0	0	**94**	PFDLLDFN	S01.133-cathepsin G
0	2	100	8	0	64	9	4	28	0	0	**91**	DEVDNKVK	C14.003-caspase-3
1	0	10	15	0	79	10	4	5	0	0	**90**	GDEVDGA	C14.003-caspase-3
1	0	8	18	0	41	9	8	17	0	0	**89**	DGPDGPEE	C14.003-caspase-3
4	1	10	13	0	0	11	3	5	0	0	**65**	AVEVDG	CAS-L84

A mixture of 1085 substrates (list at http://www.5z.com/substrates.pdf) was incubated with the indicated protease at a specified concentration. Substrates exhibiting a positive/negative ratio of <3 were not included in the evaluation. The following concentrations were used: chymotrypsin (CT) – 100 pM, trypsin (TP) – 100 pM, endoproteinase Lys-C – 1 nM, endoproteinase Arg-C – 10 pM, endoproteinase Glu-C – 1 nM, endoproteinase Asp-N – 100 pM, HRV1 – 10 nM, tobacco etch virus protease (TEV) – 1 U, enterokinase (EK) – 0.005 U, factor Xa (FXa) - 100 pM, thrombin (Thr) - 1 nM, caspase 3 (C3) - 100 pM. The annotation specifies the enzyme which is associated with the particular substrate in the Merops database. If the sequence of the substrate in database was not an octamer, then it was extended by the use of glycine residues. Annotation MMP-L or CAS-L denotes that these substrates are from the libraries designed for the study of matrix metalloproteases or caspases.

substrates usually contained an acidic (Asp or Glu) or hydrophobic (Phe, Tyr, Leu) residue in close proximity to arginine or lysine. Sequences with proline at the P1' position were extremely poor substrates.

Chymotrypsin substrates defined in the MEROPS database were not the best cleaved substrates in the chymotrypsin assay. We found better chymotrypsin substrates in our pool than defined in MEROPS database, particularly peptides containing an aromatic amino acid followed by a basic residue.

Endoproteinases Asp-N, Glu-C, Lys-C, and Arg-C behaved as predicted and cleaved substrates containing Asp (or Asn), Glu, Lys, and Arg, respectively. However, we found many examples of peptides containing these residues that were not cleaved efficiently by these endoproteinases, but we were unable to identify consensus sequences which would allow us to predict poor substrate quality of the given peptide.

Enterokinase cleaves after lysine at the cleavage site Asp-Asp-Asp-Asp-Lys (e.g. [56, 57]). However, it also cleaves at other basic residues, depending on the conformation of the protein substrate. In our assay, the "classical" substrate DDDDK performed well, but numerous peptides were found to be better substrates. Most of enterokinase substrates contain the sequences ER (or DR or DK), EXR or EXXR. Moreover, we found some sequences that did not follow this rule, e.g., NSNRAFGV (gingipain R substrate), GGWRGGGG (coagulation factor IXa substrate), NITYRGTG (calpain 1 substrate), SLRYYTAG (MMP-2 substrate), or LGIRSFRN (thrombin substrate). The last sequence, LGIRSFRN, was found previously to be cleaved by enterokinase using an alternative assay [41].

In contrast to the broad specificity enzymes mentioned earlier, tobacco etch virus (TEV) protease has a unique specificity which is used for removal of affinity tags from recombinant proteins [58]. We did not expect to find substrates of TEV protease from the large mixture of other protease substrates. However, besides the specific substrates GENLYFQG and LYFQSGTV and alternative substrate IYTQSLDD, several other sequences were cleaved to a minor extent. Interestingly, at low enzyme concentrations, TEV specifically cleaved only substrates containing an LYFQ motif from among the 1000 peptide mixture.

Factor Xa cleaves after an arginine residue in its preferred cleavage site Ile-(Glu or Asp)-Gly-Arg. It will sometimes cleave at other basic residues, depending on the conformation of the protein substrate (for its use see e.g. [59, 60]). The most common secondary site is Gly-Arg. Factor Xa will not cleave a site followed by proline or arginine. These conclusions were largely confirmed by our study. In addition we have found that substrates with sequences Pro-Arg and Ala-Arg were also cleaved by factor Xa.

The optimum cleavage sequences for alpha-thrombin are (a) P4-P3-Pro-Arg-P1'-P2', where P3 and P4 are hydrophobic amino acid and P1', P2' are nonacidic amino acids and (b) P2-Arg-P1', where P2 or P1' are Gly [61]. Our screening confirmed these consensus sequences, though we found several exemptions to these rules. There was clear preference for Ser, Thr, Ala, or Gly in P1'. However, we found number of sequences defined as thrombin substrates in the MEROPS database that were poor substrates. Fig. **4** shows titration curves for the selected substrates which we have tested in our assay.

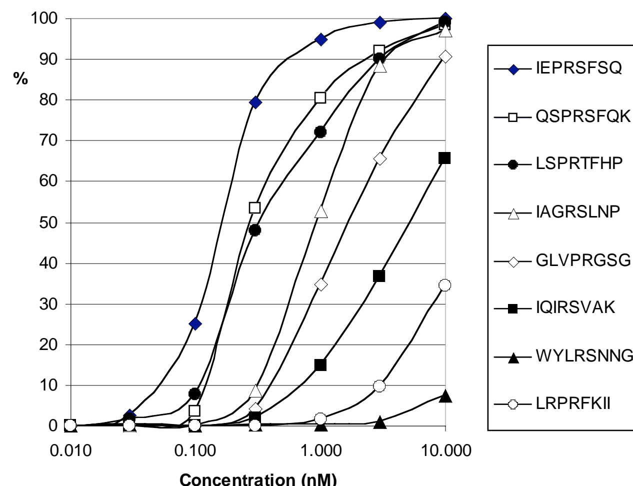

Fig. (4). Titration curves of thrombin substrates selected from the MEROPS database. Substrates were selected to illustrate the extended dynamic range of assays utilizing multiple substrates for the same enzyme.

Thrombin substrates were also investigated for their cleavage pattern by electrospray mass spectrometry. The fourteen top substrates found in our assay were individually incubated with thrombin. Table **2** shows the cleavage sites defined by fragments observed *via* mass spectrometry. Observed fragments show that the cleavage occurred at the location predicted by earlier studies. Several substrates prepared as a part of "mutation" of MMP substrate (MMP substrate library) showed high sensitivity to thrombin cleavage. Two of these substrates contained two potential sites for thrombin cleavage, and they were, indeed, cleaved in two positions.

Table 2. Top Fourteen Thrombin Substrates and Their Cleavage Sites as Defined by Using Mass Spectrometry

Substrate	Cleavage Site	Annotation
IEPRSFSQ	IEPR ↓ SFSQ	S01.217-thrombin
QSPRSFQK	QSPR ↓ SFQK	S01.217-thrombin
LSPRTFHP	LSPR ↓ TFHP	S01.217-thrombin
SGAKPRALTA	SGAKPR ↓ ALTA	MMP library
GVPRSFRG	GVPR ↓ SFRG	S01.217-thrombin
IAGRSLNP	IAGR ↓ SLNP	M10.005-stromelysin 1 (MMP-3)
LGIRSFRN	LGIR ↓ SFRN	S01.217-thrombin
VVPRGVNL	VVPR ↓ GVNL	S01.217-thrombin
GLVPRGSG	GLVPR ↓ GSG	S01.217-thrombin
IKPRIVGG	IKPR ↓ IVGG	S01.211-coagulation factor XIIa
SGRSPRRFTA	SGR ↓ SPR ↓ RFTA	MMP library
SGRAPRNLTA	SGR ↓ APR ↓ NLTA	MMP library
SGRAPSNLTA	SGR ↓ APSNLTA	MMP library
GRASLNHG	GR ↓ ASLNHG	M10.005-stromelysin 1 (MMP-3)

Assay Characterization

We used thrombin as a model enzyme for characterization of our assay metrics using the 1000-plex substrate conjugate pool (substrate concentration ~0.25 nM each). The limit of detection (LOD) of thrombin activity for the best substrate in the pool (IEPRSFSQ) was 27 pM. An analysis of assay precision found that, with this best substrate, a <1.5 fold change in enzyme concentration can be accurately measured. An important characteristic of an enzymatic assay is the linear dynamic range of substrate concentrations across which the enzyme is active and can be quantified. Multiplexed assays provide a significant improvement over traditional assays that utilize a single substrate. As can be seen in Fig. **4**, although the linear dynamic range defined by any individual substrate is limited (slightly over one order of magnitude), combined use of multiple substrates can extend the dynamic range to over three orders of magnitude.

Protease Fingerprinting

Fig. **5** shows the "star plots" of tested proteases on the 1000-plex substrate mixture. Cleavage of any given substrate by a protease is represented as a ratio of the assay signal to the maximum possible signal from that substrate conjugate (CR=Cleavage Ratio). Each substrate is represented by one of a thousand radial spokes in the radial plot. The length of the spoke is proportional to the cleavage ratio for that particular substrate by the protease. The radial plots of the CRs of several proteases show their unique cleavage signatures. As is clearly visible, our selection of substrates can distinguish between all tested proteases, and identification of particular enzyme by its "fingerprint" (see e.g. [24, 40, 62-64]) is feasible.

Application of the Assay to Biological Samples

We used our multiplexed assay to study the activity of caspases in apoptotic cellular extracts. For these studies, we synthesized dedicated pools of substrates and applied these relatively small mixtures alone or as a part of 1000-plex substrate mixture. We found that the performance of the substrates in a small mixture (110 plex) was identical to the performance in an order of magnitude bigger mixture (1000 plex) (data not shown). Detection limits for caspase substrates is given in Table **3**.

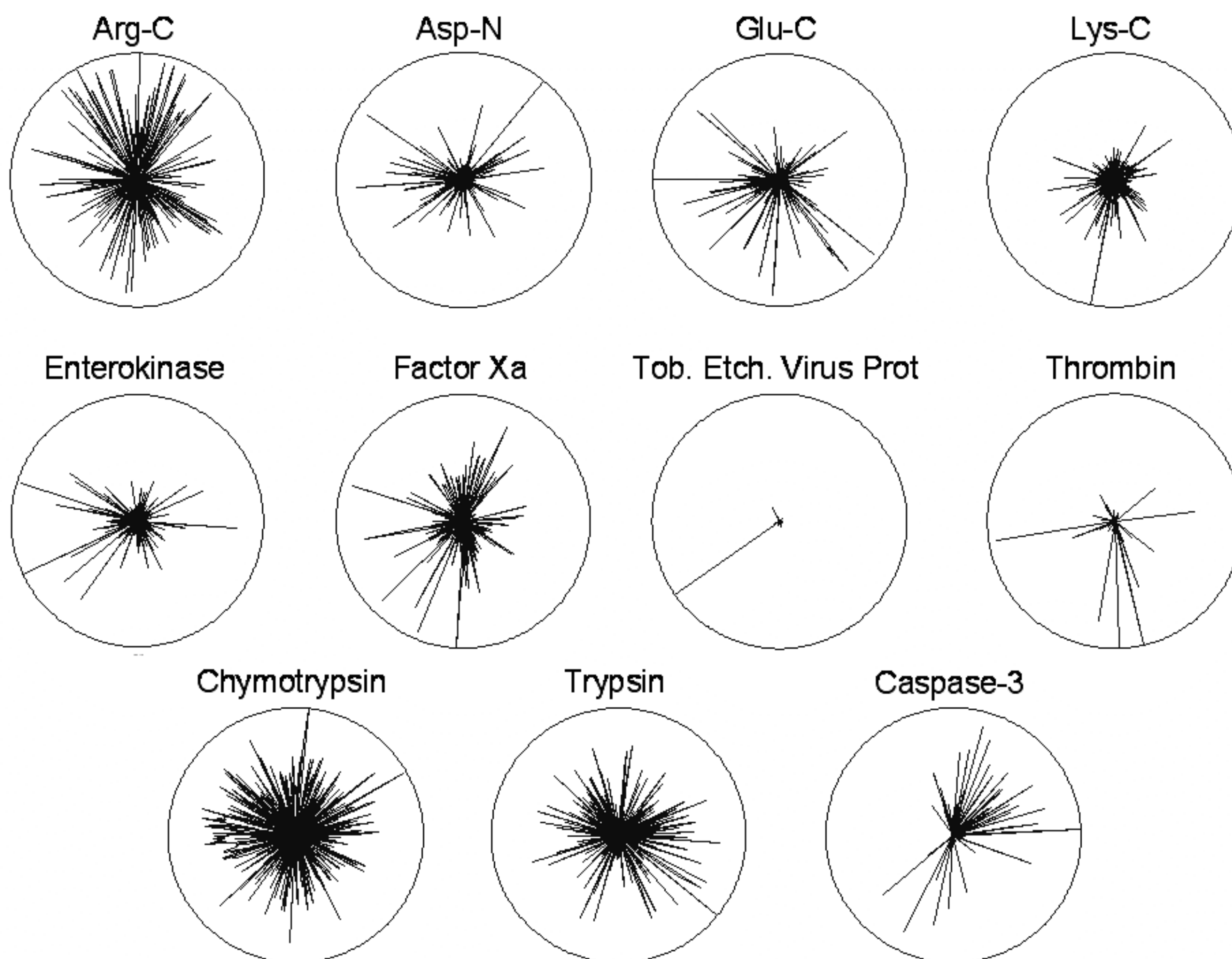

Fig. (5). Star plots of 1000-plex substrate pool tested with model proteases. The circles represent signal intensities of 100% (CR=1). Each plot represents all 1000 peptides in the same order, number 1 being at the spoke pointing straight up and ascending numbers going in the clockwise direction (peptide 250 pointing to the right, peptide 500 pointing down, etc.) Protease concentrations used to generate the data shown ranged from 10 pM to 1 nM, except tobacco etch virus protease (1 u) and enterokinase (0.005 u).

Table 3. Limit of Detection of Caspases and Native Concentrations of these Enzymes in Cells

Caspase	LOD on Array	Approx. Cell Conc.
3	< 8 pM	100 nM
6	< 1 nM	10 nM
7	< 1 nM	6 nM
8	< 2 nM	10 nM
9	< 30 nM	12 nM

Jurkat cells were cultivated and stimulated to undergo apoptosis by the anti-Fas antibody. Jurkat cells were harvested following 0, 1, 3 and 6 h of anti-Fas treatment and cell lysates prepared as previously described [65]. An aliquot of 20 μL of control or apoptotic cell lysates (representing equal starting cell number) were substituted for purified protease in the multiplexed protease assay, at the 1000-plex level. A 20 μL of 0.6 nM purified caspase3 was run alongside as a control. Only a handful of substrates from the 1000-plex pool showed lysate-induced cleavage as measured on the microarray. Most of the substrates were not cleaved by the lysates (data not shown). The data for the top five substrates cleaved by the apoptotic Jurkat lysates are shown in Fig. **6**. The top five cleaved substrates all contain the canonical caspase-cleavage motif (DXXD/G). Moreover, the extent of cleavage of these substrates follows the expected time-course corresponding to the previously known increase in caspase-3 levels in apoptotic cells. Similar apoptotic time-course experiments were conducted with cell lysates from vector-control Jurkat cells and Jurkat cells that over-express Bcl2. (Bcl2 is an oncogene whose over-expression has been shown to inhibit apoptosis in many cellular systems [66]). As expected, the control cells showed an increase in cleavage of caspase-specific substrates with time whereas this activity was mostly abrogated in the lysates from Jurkat cells that overexpress Bcl2 (data not shown).

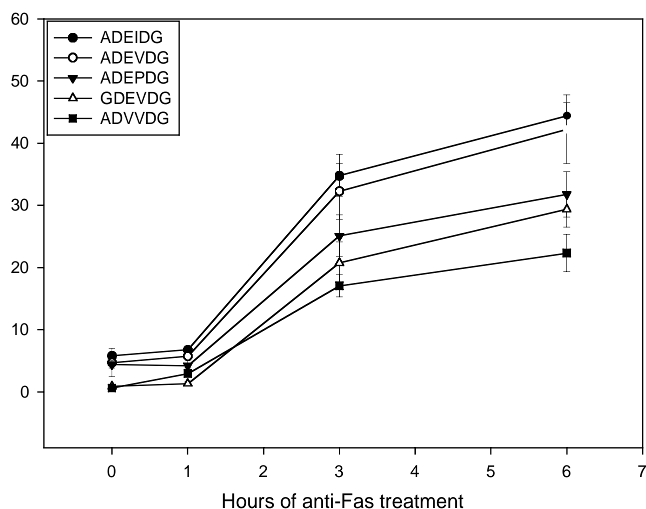

Fig. (6). Time-course of caspase activity in apoptotic Jurkat lysates. The data shown are the average of three independent multiplexed protease assays, interrogating apoptotic lysates generated from two independent Jurkat/anti-Fas time-course experiments. Error bars indicate the standard deviation of three hybridization replicates.

CONCLUSIONS

We have developed a solution-based multiplexed protease assay utilizing enzymatic cleavage of a portion of the substrate conjugate containing biotin. A DNA tag (used for the assay readout by DNA microarray) and a histidine tag (used for visualization of substrate on the array) were shown not to interfere with the enzymatic processing of the peptide substrate. The assay can be performed at close to physiological conditions, and multiplexing of more than a 1000 substrate conjugates does not interfere with individual performance of the substrates. The assay is sensitive, allowing detection of enzymes at cellular concentrations, and has a large dynamic range due to the multiple substrates used for the same enzyme.

The described assay system is capable of fast and accurate evaluation of protease cleavage of a multiplicity of peptides. More importantly, it can be used for characterization of proteases in biological samples. We have proven its effectiveness with both broad- and narrow-specificity proteases. This assay can be used for "fingerprinting" biological samples, which is especially useful in disease association studies. It can be very useful in drug candidate screening of protease inhibitors and in evaluation of drug candidates for their potential harmful side effects such as causing toxicity by interfering with essential protease activities.

ACKNOWLEDGEMENTS

The project was partially supported by NIH SBIR grant R44 AI056869-02. We thank Prof. Salvesen and Prof. Smith (The Burnham Institute, San Diego) for their kind gift of caspases and matrix metalloproteases, Martin Latterich, Gali Steinberg-Tatman, and Eugene Chudin for helpful discussions, Bahram Kermani for the help with data analysis, Kellie Wise, Veronica Schevchenko and Kerri Tardio for reagent preparation, and Cynthia Allred for assistance with grant applications.

REFERENCES

[1] Vitzthum, F.; Behrentz, F.; Anderson, N.L.; Shaw, J.H. *J. Proteom. Res.*, **2005**, *4*, 1086-1097.
[2] Lopez-Otin, C.; Overall, C.M. *Nat. Rev. Mol. Cell Biol.*, **2002**, *3*, 509-519.
[3] 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.; Tempst, P. *J. Clin. Invest.*, **2006**, *116*, 271-284.
[4] Villanueva, J.; Martorella, A.J.; Lawlor, K.; Philip, J.; Fleisher, M.; Robbins, R.J.; Tempst, P. *Mol. Cell Proteomics*, **2006**, *5*, 1840-1852.
[5] Briz, V.; Poveda, E.; Soriano, V. *J.Antimicrob.Chemother.*, **2006**, *57*, 619-627.
[6] Yin, P.D.; Das, D.; Mitsuya, H. *Cell Mol. Life Sci.*, **2006**, *63*, 1706-1724.
[7] Lebon, F.; Ledecq, M. *Curr. Med. Chem.*, **2000**, *7*, 455-477.
[8] Richardson, P.L. *Curr. Pharm. Des.*, **2002**, *8*, 2559-2581.
[9] Maly, D.J.; Huang, L.; Ellman, J.A. *Chembiochem*, **2002**, *3*, 16-37.
[10] Goddard, J.P.; Reymond, J.L. *Curr. Opin. Biotechnol.*, **2004**, *15*, 314-322.
[11] Meldal, M.; Svendsen, I.; Breddam, K.; Auzanneau, F.I. *Proc. Natl.Acad. Sci. USA*, **1994**, *91*, 3314-3318.
[12] Knight, C.G. *Methods Enzymol.*, **1995**, *248*, 18-34.
[13] Kuroda, K.; Kato, M.; Mima, J.; Ueda, M. *Appl. Microbiol. Biotechnol.*, **2006**, *71*, 127-136.
[14] Giepmans, B.N.; Adams, S.R.; Ellisman, M.H.; Tsien, R.Y. *Science*, **2006**, *312*, 217-224.
[15] He, Y.; Tian, Y.; Ribbe, A.E.; Mao, C. *J. Am. Chem Soc.*, **2006**, *128*, 12664-12665.

[16] Oliphant, A.; Barker, D.L.; Stuelpnagel, J.R.; Chee, M.S. *BioTechniques,* **2002**, *32*, S56-S61.

[17] Urbina, H.D.; Debaene, F.; Jost, B.; Ole-Feysot, C.; Ason, D.E.; Kuzmic, P.; Harris, J. L.; Winssinger, N. *Chembiochem,* **2006**, *7,* 1790-1797.

[18] Winssinger, N.; Damoiseaux, R.; Tully, D.C.; Geierstanger, B.H.; Burdick, K.; Harris, J.L. *Chemistry & Biology,* **2004**, *11*, 1351-1360.

[19] Diaz-Mochon, J.J.; Bialy, L.; Bradley, M. *Chem.Commun.,* **2006**, 3984-3986.

[20] Uttamchandani, M.; Huang, X.; Chen, G.Y.; Yao, S.Q. *Bioorg. Med. Chem. Lett.,* **2005**, *15*, 2135-2139.

[21] Uttamchandani, M.; Walsh, D.P.; Yao, S.Q.; Chang, Y.T. *Curr. Opin. Chem. Biol.,* **2005**, *9*, 4-13.

[22] Uttamchandani, M.; Chan, E.W.; Chen, G.Y.; Yao, S.Q. *Bioorg. Med. Chem. Lett.,* **2003**, *13*, 2997-3000.

[23] Salisbury, C.M.; Maly, D.J.; Ellman, J.A. *J. Am. Chem. Soc.,* **2002**, *124*, 14868-14870.

[24] Reymond, J.L.; Wahler, D. *Chembiochem,* **2002**, *3*, 701-708.

[25] Lesaicherre, M.L.; Uttamchandani, M.; Chen, G.Y.; Yao, S.Q. *Bioorg. Med. Chem. Lett.,* **2002**, *12*, 2079-2083.

[26] Lesaicherre, M.L.; Uttamchandani, M.; Chen, G.Y.; Yao, S.Q. *Bioorg. Med. Chem. Lett.,* **2002**, *12*, 2085-2088.

[27] Houseman, B.T.; Huh, J.H.; Kron, S.J.; Mrksich, M. *Nat. Biotechnol.,* **2002**, *20*, 270-274.

[28] Harris, J.; Mason, D.E.; Li, J.; Burdick, K.W.; Backes, B.J.; Chen, T.; Shipway, A.; Van Heeke, G.; Gough, L.; Ghaemmaghami, A. *Chemistry & Biology,* **2004**, *11*, 1361-1372.

[29] Livache, T.; Bazin, H.; Caillat, P.; Roget, A. *Biosensors and Bioelectronics,* **1998**, *13*, 629-634.

[30] Reineke, U.; Volkmer-Engert, R.; Schneider-Mergener, J. *Curr.Opin.Biotechnol.,* **2001**, *12*, 59-64.

[31] Houseman, B.T.; Mrksich, M. *Trends Biotech.,* **2002**, *20*, 279-281.

[32] Reimer, U.; Reineke, U.; Schneider-Mergener, J. *Curr. Opin. Biotechnol.,* **2002**, *13*, 315-320.

[33] Reineke, U.; Ivascu, C.; Schlief, M.; Landgraf, C.; Gericke, S.; Zahn, G.; Herzel, H.; Volkmer-Engert, R.; Schneider-Mergener, J. *J. Immunol. Method,* **2002**, *267*, 37-51.

[34] Cooley, P.; Wallace, D.; Antohe, B. *J. Ass. Lab. Autom.,* **2002**, *7*, 33-39.

[35] Caputo, E.; Moharram, R.; Martin, B.M. *Anal. Biochem.,* **2003**, *321*, 116-124.

[36] Heine, N.; Ast, T.; Schneider-Mergener, J.; Reineke, U.; Germeroth, L.; Wenschuh, H. *Tetrahedron,* **2003**, *59*, 9919-9930.

[37] Min, D.H.; Mrksich, M. *Curr. Opin. Chem. Biol.,* **2004**, *8*, 554-558.

[38] Weiser, A.A.; Or-Guil, M.; Tapia, V.; Leichsenring, A.; Schuchhardt, J.; Frommel, C.; Volkmer-Engert, R. *Anal. Biochem.,* **2005**, *342*, 300-311.

[39] Shults, M.D.; Kozlov, I.A.; Nelson, N.; Kermani, B.G.; Melnyk, P.C.; Shevchenko, V.; Srinivasan, A.; Musmacker, J.; Hachmann, J.P.; Barker, D.L.; Lebl, M.; Zhao, C. *Chembiochem,* **2007**, 933-942.

[40] Craig, D.B.; Wong, J.C.; Polakowski, R.; Dovichi, N.J. *Anal. Chem.,* **1998**, *70*, 3824-3827.

[41] Kozlov, I.A.; Melnyk, P.C.; Zhao, C.; Hachmann, J.P.; Shevchenko, V.; Srinivasan, A.; Barker, D.L.; Lebl, M. *Comb. Chem. High Throughput Screen.,* **2006**, *9*, 481-487.

[42] Michael, K.L.; Taylor, L.C.; Schultz, S.L.; Walt, D.R. *Anal. Chem.,* **1998**, *70*, 1242-1248.

[43] Walt, D.R. *Account. Chem. Res.,* **1998**, *31*, 267-278.

[44] Steemers, F.J.; Gunderson, K.L. *Pharmacogenomics,* **2005**, *6*, 777-782.

[45] Steemers, F.J.; Ferguson, J.A.; Walt, D.R. *Nat. Biotech.,* **2000**, *18*, 91-94.

[46] Lebl, M. *J. Assoc. Lab. Autom.,* **2003**, *8*, 30-36.

[47] Lebl, M. *Bioorg. Med. Chem. Lett.,* **1999**, *9*, 1305-1310.

[48] Hachmann, J.P.; Lebl, M. *Biopolymers (Pept.Sci.),* **2006**, *84*, 340-347.

[49] Hachmann, J.P.; Lebl, M. *J. Comb. Chem.,* **2006**, *8*, 149.

[50] King, D.S.; Fields, C.G.; Fields, G.B. *Int. J. Peptide Prot. Res.,* **1990**, *36*, 255-266.

[51] Lebl, M.; Krchnak, V.; Ibrahim, G.; Pires, J.; Burger, C.; Ni, Y.; Chen, Y.; Podue, D.; Mudra, P.; Pokorny, V.; Poncar, P.; Zenisek, K. *Synthesis-Stuttgart,* **1999**, 1971-1978.

[52] Lebl, M.; Burger, C.; Ellman, B.; Heiner, D.; Ibrahim, G.; Jones, A.; Nibbe, M.; Thompson, J.; Mudra, P.; Pokorny, V.; Poncar, P.; Zenisek, K. *Collect. Czech. Chem. Commun.,* **2001**, *66*, 1299-1314.

[53] Kozlov, I.A.; Melnyk, P.C.; Hachmann, J.P.; Barker, D.L.; Lebl, M.; Zhao, C. *Nucleosides, Nucleotides & Nucleic Acids,* **2007**, *in press*.

[54] Turk, B.E.; Cantley, L.C. *Methods,* **2004**, *32*, 398-405.

[55] Stierandova, A.; Sepetov, N.F.; Nikiforovich, G.V.; Lebl, M. *Int. J. Peptide Prot. Res.,* **1994**, *43*, 31-38.

[56] Collins-Racie, L.A.; McColgan, J.M.; Grant, K.L.; DiBlasio-Smith, E.A.; McCoy, J.M.; Lavallie, E.R. *Biotechnology (NY),* **1995**, *13*, 982-987.

[57] Maroux, S.; Baratti, J.; Desnuelle, P. *J. Biol. Chem.,* **1971**, *246*, 5031-5039.

[58] Stevens, R.C. *Struc. Fold. Des.,* **2000**, *8*, R177-R185.

[59] Quinlan, R.A.; Moir, R.D.; Stewart, M. *J. Cell Sci.,* **1989**, *93 (Pt 1)*, 71-83.

[60] Nagai, K.; Perutz, M.F.; Poyart, C. *Proc. Natl. Acad. Sci. USA,* **1985**, *82*, 7252-7255.

[61] Chang, J.Y. *Eur. J. Biochem.,* **1985**, *151*, 217-224.

[62] Wahler, D.; Badalassi, F.; Crotti, P.; Reymond, J.L. *Chemistry,* **2002**, *8*, 3211-3228.

[63] Wahler, D.; Badalassi, F.; Crotti, P.; Reymond, J.L. *Angew. Chem. Int. Ed. Engl.,* **2001**, *40*, 4457-4460.

[64] Grognux, J.; Reymond, J.L. *Chembiochem,* **2004**, *5*, 826-831.

[65] Armstrong, R.C.; Aja, T.; Xing, J.; Gaur, S.; Krebs, J.F.; Hoang, K.; Bai, X.; Korsmeyer, S.J.; Karanewsky, D.S.; Fritz, L.C.; Tomaselli, K.J. *J. Biol. Chem.,* **1996**, *271*, 16850-16855.

[66] Srinivasan, A.; Foster, L.M.; Testa, M.-P.; Ord, T.; Keane, R.W.; Bredesen, D.E.; Kayalar, C. *J. Neurosci.,* **1996**, *16*, 5654-5660.

Received: May 17, 2007 **Revised: June 29, 2007** **Accepted: July 2, 2007**