Nanoparticle-Chaperoned Urinary 'Synthetic Biomarkers' for Profiling Proteases in Cancer

by

Omar O. Abudayyeh

Submitted to the
Department of Mechanical Engineering
In Partial Fulfillment of the Requirements for the Degree of

Bachelor of Science in Engineering as Recommended by the
Department of Mechanical Engineering
at the
Massachusetts Institute of Technology

June 2012

Signature of Author ..
Department of Mechanical Engineering
May 11, 2012

Certified by .....
Sangeeta Bhatia
John J. and Dorothy Wilson Professor of Health, Sciences, and Technology &
Electrical Engineering and Computer Science
Thesis Supervisor

Certified by .....
Scott Manalis
Professor of Biological and Mechanical Engineering
Thesis Supervisor

Accepted by .....
John H. Lienhard V
Samuel C. Collins Professor of Mechanical Engineering
Undergraduate Officer

Nanoparticle-Chaperoned Urinary 'Synthetic Biomarkers' for Profiling Proteases in Cancer

by

Omar O. Abudayyeh

Submitted to the Department of Mechanical Engineering
on May 11, 2012 In Partial Fulfillment of the
Requirements for the Degree of

Bachelor of Science in Engineering as Recommended by the
Department of Mechanical Engineering

Abstract

Many biomarker-based diagnostics have poor predictive value because of their dependence on naturally occurring endogenous biomolecules to indicate disease. This work presents a diagnostic platform that uses nanoparticles to profile underlying proteolytic signatures of diseases. In this thesis, work is presented on long circulating peptide-nanoparticle probes that can survey, sense, and remotely report on dysregulated protease activities in cancer. In this strategy, iron oxide nanoparticles are utilized as chaperones to deliver protease-specific peptide libraries to tumors whereupon selective cleavage by active proteases releases peptide fragments that are cleared by the renal system into the urine. These peptide fragments are predesigned with internal photolabile triggers that uncage isobaric peptide mass tags optimized for multiplexed LC MS/MS quantification. Results demonstrate that such peptide 'synthetic biomarker' panels uncover unique proteolytic signatures that can be correlated with disease states, allowing for the detection of cancer and potential long-term monitoring of disease using an implantable form. This concept of administering prodiagnostic reagents and analyzing remote reporters is amenable to a broad range of protease-dependent complex diseases, such as liver fibrosis and coagulopathies, and infectious disease.

Thesis Supervisor: Sangeeta Bhatia
Title: John J. and Dorothy Wilson Professor of Health, Sciences, and Technology &
Electrical Engineering and Computer Science

Thesis Supervisor: Scott Manalis
Title: Professor of Biological and Mechanical Engineering

Acknowledgements

With much gratitude and respect, I would like to acknowledge Dr. Gabe Kwong for his mentorship, encouragement, and the long discussions that inspired me and improved the way I think about science. I would like to express my appreciation for his patience, time, and friendship.

I admiringly acknowledge Professor Sangeeta Bhatia for providing me the opportunity to join her team in the Laboratory for Multiscale Regenerative Technologies at MIT and for supporting my work with her invaluable advice and interactions, which have contributed greatly to my intellectual maturity. I am very grateful for this amazing educational and research experience.

Finally, many thanks go to all the members in the laboratory who created an intellectually stimulating and engaging research environment.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Biomarkers in Disease

One of the major goals associated with noninvasive diagnostics of complex diseases is the development of sensitive and specific biomarker assays for early detection and monitoring (1,2). Biomarkers serve as important measurable indicators of physiological states, particularly signifying the presence or stage of a disease. Various forms of biomarkers have demonstrated much potential for screening, diagnosing, and monitoring diseases in addition to guiding personalized therapies and assessing the efficacy of therapies (9). Biomarkers that have been identified and validated for use, include circulating tumor cells (3), proteins (2,4), peptides (5), metabolites (2), cell-free DNA and RNA (2,6), microRNAs (7), and exosomes (8). Enzymatic activities have also been used as biomarkers to predict disease. Alkaline and acid phosphatase activities were used successfully to predict the presence of bone disease and prostate cancer [4], respectively. Enzymes such as these were attractive options for biomarkers because of the high sensitivity associated with detecting their cleaved products.

Biomarkers have been successfully discovered for a variety of diseases, but many problems are associated with accurately detecting them. Renal diseases, such as acute renal failure, have been targeted for biomarker discovery because of their lack of characterization, difficulty to be diagnosed early, and their preferred diagnostic method being the biopsy, a highly invasive procedure. Similar issues have prompted the need for liver disease biomarkers as well and so studies have focused on finding direct markers for liver fibrogenesis that release into the peripheral blood, including various proteins such as cytokines and enzymes (collagenases and

metalloproteinases) (28). Additionally, proteomic approaches have been applied for analysis of biomarkers for lung, breast, ovarian, colon, and prostate cancers for early detection and monitoring (2). However, there remains significant challenges with naturally occurring biomarkers, including the inability to detect early stage disease, making more invasive diagnostic procedures, such as biopsies, necessary. The development of diagnostics that incorporate effective combinations of biomarkers with improved sensitivity and specificity will be needed to circumvent the current problems with biomarker tests.

1.2 Detecting Biomarkers

The ability of blood biomarkers to dynamically reflect multiple disease states and its relative ease of collection from patients have made blood central to many diagnostic assays (4). For most serum analytes, robust detection in plasma is challenging since most biomarkers are shed from diseases tissues at low levels and are further diluted upon entering systemic circulation. If the average primary tumor size is ~0.05mL then biomolecules shed into the blood, which has a total volume of ~5L, will be diluted by a factor of 10^5. Further, these signals of interest can easily become masked by the complexity of the plasma proteome, which contains hundreds of proteins present at high concentrations such as Albumin (4). Biomolecules can also be degraded *in vivo* and *ex vivo*.

Mass spectrometry has emerged as a technology for analyzing protein and peptide biomarkers, but is especially hindered by the inability to sensitively detect low signal due to complex samples. While these limitations can only be overcome by extensive sample processing, fractionation, chromatography, and data analysis, which is costly and time consuming (9), alternative biological fluids, particularly urine, offer a proteomically simpler and non-invasive

source for detecting disease (9, 10). However, urine has drawbacks including a molecular weight cut-off (<60kDa) (9,11), which would inhibit the detection of enzymes and other informative biomarkers released from the disease site. While in some cases of kidney disease and cancer larger biomarkers, including proteins, are released into the urine and become detectable (11), this is generally not amenable to a broad range of diseases.

Many of the challenges associated with blood and urine diagnostics could be overcome by the administration of exogenous biomarkers. Exogenous agents have been designed to report on the underlying physiology of specific diseases, such as activity-based probes that fit into active sites within proteases and report on their activity *in vivo* (12,13). Other examples of exogenous diagnostics are molecules for imaging, such as FDG-PET, for determining sites of glucose metabolism. Challenges with these approaches, however, include the inability to multiplex signals (limitations with reporter output such as fluorophore spectral overlap) and other difficulties that are associated with specifically delivering the diagnostic to the disease site and avoiding rapid renal clearance.

1.3 Proteases in Disease

Proteases are essential for the control of many biological processes and for the function of most biological systems across all living organisms. Within this role, proteases have been found to activate many proteins, regulate protein-protein interactions, interact within cellular pathways of processing information, and create and amplify molecular signals (14). These functions translate to influencing DNA replication/transcription, angiogenesis, wound repair, stem cell movement, blood clotting, immunity, inflammation, activation of zymogens, remodeling of the extracellular matrix (ECM), and apoptosis (14). Because of all these roles, aberrant protease activity is related to a variety of diseases, including cancer, neurodegenerative

disorders, inflammatory conditions, and cardiovascular diseases. A positive correlation between MMP expression and tumor progression has been found as MMPs play an active role in tumor invasion and metastasis by degrading the ECM and in basement membrane turnover for invasion of cancerous cells into the underlying organs and vasculature (29). Chronic obstructive pulmonary disease and emphysema have been shown to be caused by destruction of the alveoli by excessive neutrophil elastase activity. Hepatic fibrosis can also be caused by aberrant elastase activity and matrix degradation caused by metalloproteinases. When the balance of proteases and their inhibitors is tipped, architectural elements and function of the liver are drastically altered. Dysregulated proteases play a role in the progression of a variety of diseases and are an attractive target for developing a diagnostic platform.

1.4 Detection of Proteases

The nature of proteases to act on a variety of substrates and their role in health and disease makes them amenable for real-time, non-invasive monitoring for diagnosing disease. Because of the promiscuity of proteases for endogenous substrate degradation, multiplexed quantification of protease activity would elucidate protease network interactions into abnormal biological states. Activity-based probes have been developed in multiple studies for analyzing protease activity. When these probes bind to the active size of the enzymes of interest, they covalently bind. Because they are labeled with fluorophores or biotin, they allow for visualization or affinity purification of the active enzymes, respectively. This approach has been adopted for labeling numerous biomedically relevant enzymes, including serine hydrolases, cysteine proteases, and oxidoreductases (12). Additionally, probes have been designed to study the MMPs in live cells and in animals for the purpose of identifying MMPs upregulated in cancer

cells and for aiding in the development of inhibitors for clinical use. Near-infrared florescent molecular probes have also been developed for detecting MMP, cathepsin, and caspase activity through live imaging in tumor-bearing animals (30).

A major issue with most of these probing mechanisms is that there is much promiscuity between enzymes and their substrates. As many of these detection methods, including zymography, mass-spectrometry based systems, and activity based probes, do not simultaneously allow for non-invasiveness, specificity, and high throughput, new approaches are being tested, such as the 'Proteolytic Activity Matrix Analysis' (PrAMA) method. This method addressed many of these issues by using FRET-substrates for dynamic, multiplexed quantification of specific protease activities. Through using previously determined cleavage patterns of substrates by the purified enzymes, PrAMA deconvolutes specific protease activity from complex protease signatures. PrAMA has been validated for identifying protease activity within complex networks *in vitro* with a strong potential for systems biology to screening inhibitors (31). This is an important advancement because current fluorogenic approaches do not provide the same level of specificity and multiplexing abilities.

Fluorescent-based approaches are highly limited by spectral overlap and thus do not allow for simultaneous measurement of multiple enzyme activities while mass spectrometry has many advantages because it allows for a broad range of masses to be detected accurately and thus providing multiplexed reporters. Quantitative approaches involving mass spectrometry for tracking multiple cleavage events would improve the accuracy of current diagnostics and perhaps broaden the range of diseases that could be detected with a set of probes.

1.5 Quantitative Mass Spectrometry for Detecting Protease Activity

Using insight attained from quantitative MS, protease detection assays with multiplexed capabilities could be developed. Because of the pressing need to characterize the proteome for analyses of biological systems and pathways, quantitative MS technologies have been developed. Approaches for proteome analysis usually begin with two-dimensional gel electrophoresis and selected proteins are further analyzed with quantitative MS or tandem MS/MS (57-61). Issues associated with these traditional methods for proteome quantification include the labor, sensitivity issues, and the masking of less abundant proteins. A technique for the general labeling of proteins and their quantified detection by MS has been developed using isotope-coded affinity tags (ICAT) (62) (Fig. 1-1). ICAT reagents contain a specific chemistry group for the reaction of interest, which in this study, was a thiol reactive group for cysteine modification. They further consist of a linker, which is isotope encoded, and an affinity tagged. The isotope-encoding scheme involves heavy (eight deuteriums) and light forms (no deuteriums) of the linker, allowing for quantification of duplexed peaks of the same protein from different samples (Fig. 1-2). After the ratio of proteins from different cell states are calculated from the relative peak heights in the MS, fragmentation of these peaks in the MS/MS enables the identification of the proteins, allowing for a complete picture to be obtained of the protein expression differences between two cells or states.

Further development of isotope-encoding using isobaric rags has resulted in greater multiplexing with isobaric tags for relative and absolute quantitation (iTRAQ) (32, 63). Currently, iTRAQ allows for 4-plex (Fig. 1-3) and 8-plex experiments in measuring proteins. Because iTRAQ uses isobaric tags, proteins are only differentiated by fragmentation peaks that

vary by 1 mass unit in the MS/MS. The iTRAQ reagents are balanced such that before fragmentation all the variants are of equal mass, meaning that a protein from different samples will display as a single parent peak in the MS. Only upon fragmentation are the reporter ions, which differ by one mass unit, exposed (Fig. 1-4). Isobaric approaches offer more multiplexing ability and greater sensitivity than other approaches. However, the isobaric technologies developed so far have leveled out at a maximum of 8 different tags (64), which is limiting for a widespread proteome analysis of cells or *in vivo* disease signatures. Because mass spectrometry has much promise in highly multiplexed, quantitative, analyses of biological systems, adapting isobaric techniques using peptide standards that ionize efficiently and yield high signal is desired. By using peptides for isotopic-encoding, more probes can be generated than is achievable with small molecule tags, allowing for a more multiplexed analyses to be performed for measuring proteases *in vivo*.

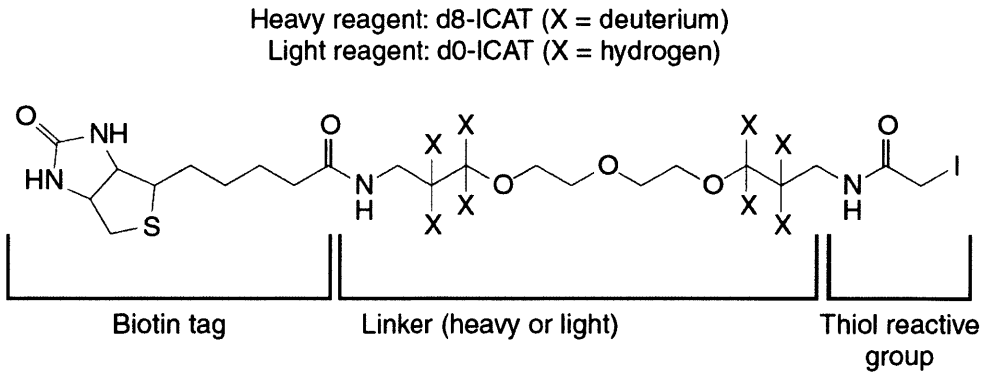


Figure 1-1: iCAT reagent used for labeling proteins with an isotope-encoded tag that allows for quantification of protein signatures by tandem mass spectrometry between two cell states. Reproduced from (62).

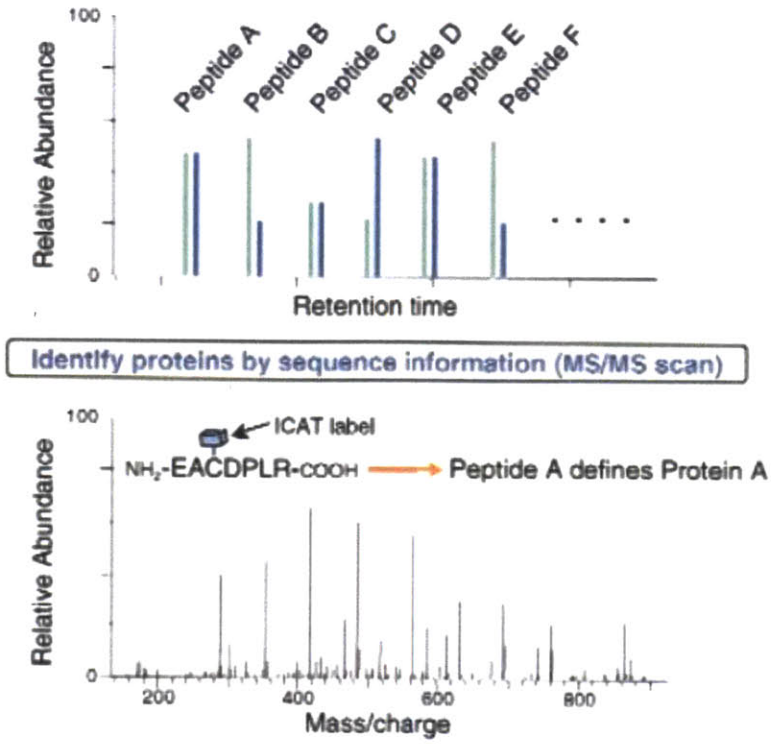


Figure 1-2: The tandem mass spectrometry scheme for quantifying iCAT labeled proteins. Proteins between two cell states are first measured by MS parent peaks and displayed as two peaks separated by a few mass units. After fragmentation, the quantified protein peaks can be identified using the fragmentation spectrum in MS/MS. Reproduced from (62).

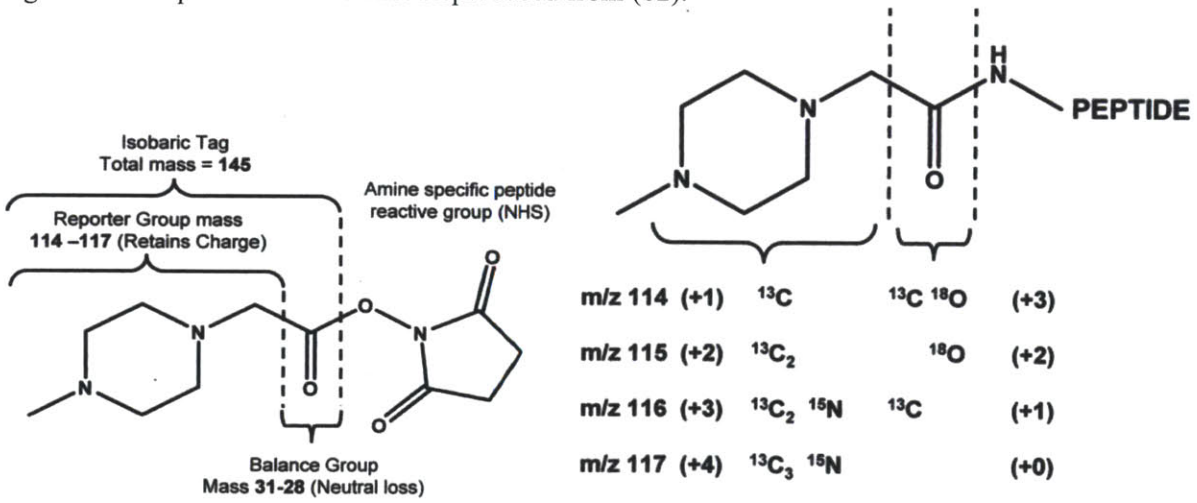


Figure 1-3: The iTRAQ reagent used for labeling proteins with isobaric mass tags. The scheme for developing a 4-plex reagent set is shown. The reagent is balanced around the carbonyl bond such that all four variants are of equal mass until fragmentation releases the reporter ions, which sequentially differ by one mass unit. Reproduced from (63).

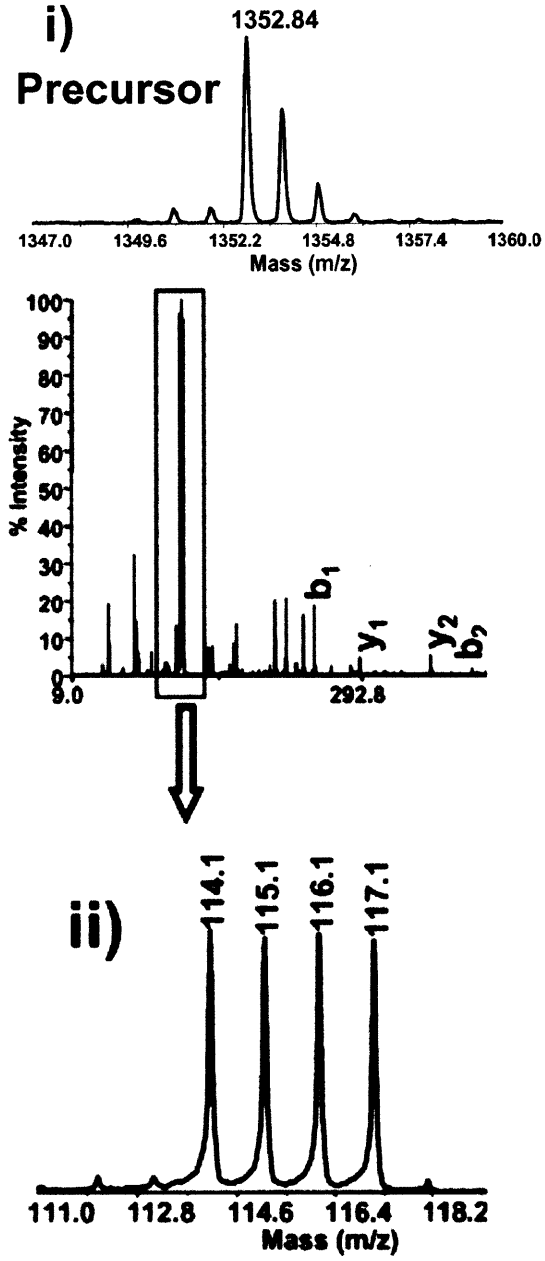


Figure 1-4: Tandem mass spectrometry detection of iTRAQ tags. Because the iTRAQ reagents are isobaric, the tagged protein appears as a single parent peak in MS. Upon fragmentation, the multiplexed nature of the iTRAQ tags becomes evident as reporter ions separated by one mass unit. Reproduced from (63).

1.6 Detection and Monitoring of Cancer

Early detection of cancer is critical for improving the poor survival rates of cancer patients. Unfortunately, current clinical technologies are deficient in their ability to detect early cancer growth and to fully eradicate certain tumors after removal (40). Two of the most common cancers in the United States, breast and prostate cancers, are particularly difficult to diagnose. Common diagnostic tools, such as Magnetic Resonance Imaging (MRI), are not sensitive or specific enough to identify early cancerous growth, especially when nonmalignant lesions are also present. As most tumors are usually not detectable until they have reached a size of approximately 1cm, by then they are more likely to have metastasized. Scientists have dealt with these deficiencies by developing creative assays that allow for the early detection and targeting of tumors.

Targeted approaches, such as monoclonal antibodies, are in development to better detect cancer. Antibodies designed to detect tumor-associated antigens have been generated and are capable of homing in on tumors and detecting angiogenesis mice and in clinical trials (49-53). Noise and specificity issues pose major limitations, however, as significant antibody signal levels in the tumor have been difficult to achieve over the background. Strategies to improve targeting include using tumor-specific peptides (52), improved labeling techniques (54), enhancing tumor vessel permeability for better extravasation (55), and upregulating the expression of tumor antigens *in vivo* (56). Protease detection approaches have also been used with the goal of increased sensitivity due to activity-based diagnosis. One study delivered near-infrared fluorescence (NIRF) probes using a polymer scaffold that allowed for long circulating times and improved delivery (57). After accumulation of the NIRF probes at the tumor site, protease cleavage of the probes freed the fluorophores from their quenched state and a 12-fold increase in

signal could be detected. Issues with specificity plague an activity-based approach, which could be solved by probing the protease profile of the tumor using multiplexed probes.

Nanoparticle-based devices also have the potential to impact cancer treatment significantly since accessing the local cancer environment within the body is key to early detection and better treatment outcome (40). While most diagnostic tools and treatments ignore the local environment and rather test the serum, these novel devices are able to precisely target cancer cells and thus sense the local tumor environment. Treatments of various cancer types, such as epithelial ovarian cancer, could highly benefit from the sensing of the local environment through the identification of cancer-related biomolecules (43,44). While nanoparticles that can target tumors to improve MRI imaging or deliver drugs are certainly possible solutions, more novel diagnostic schemes and devices are needed to detect early signs of cancer. One such device uses nanoparticle magnetic relaxation switches to monitor the beta subunit of human chorionic gonadotrophin (hCG-β), a cancer biomarker (40). The device, implanted at sites where tumors have been removed, allows MRI to detect changes in the magnetic relaxation time when the nanoparticles cluster around hCG-β, thus allowing for the local monitoring of tumor reoccurrence.

Serum biomarker assays have also been proposed for cancer. Most clinically used biomarker diagnostics lack the ability to detect small tumors and many biomarkers have low sensitivity and specificity, such as CA-125 for ovarian cancer and PSA for prostate cancer. Panels of biomarkers for diagnosing cancer would have much more potential for offering early detection that is accurate across cohorts of patients. For most serum analytes, however, robust detection in plasma is challenging since most biomarkers are shed from diseases tissues at low levels and are further diluted in the blood. Better results could perhaps be achieved by probing the tumor environment with panels of exogenous biomarkers that can efficiently be assessed in a non-invasive manner.

Such an approach would avoid the problems with serum complexity and would provide more accuracy due to the multiplexed assessment of the tumor environment.

1.7 Purpose of Inquiry

The use of 'synthetic biomarkers' that are designed to respond through specific mechanisms to complex disease processes and report non-invasively through the urine is proposed in this thesis. By using exogenous biomolecules as indicators of disease, the platform can be tailored to respond to specific molecular processes and signals within the disease. Aberrant protease activities are present in many complex disease processes, especially in cancer, thrombosis, and liver fibrosis (14). Despite the importance of proteases in disease, few technologies exist for the detection and multiplexed monitoring of aberrant protease activities *in vivo*. In this work, the aberrant matrix metalloproteinase (MMP) activity present within cancer is focused on using synthetic biomarkers consisting of long circulating peptide-nanoparticle probes designed to intravenously home in on the disease site and profile enzyme activity by releasing mass reporters locally upon proteolysis (Fig. 1-5). These cleaved mass-encoded reporters are designed to then filter into the urine for identification by using liquid chromatography tandem mass spectrometry (LC MS/MS). A description of how the administration of prodiagnostic reagents, through quantified protease activities, can improve the low sensitivity and specificity associated with current biomarker tests is presented in this thesis.

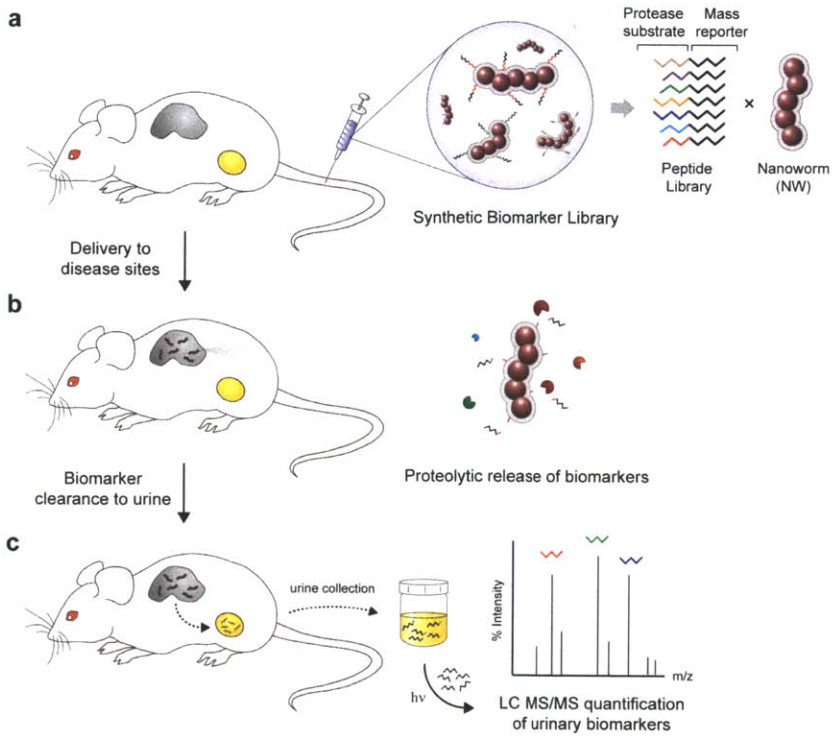


Figure 1-5: Design Schematic. a) The 'synthetic biomarker' library conjugated to the surface of the nanoworms is intravenously injected into mice. b) The nanoworms accumulate at the tumor site and proteolytic cleave occurs. c) The mass-encoded biomarkers released by cleavage freely filter into the urine and can be quantified by tandem mass spectrometry.

1.8 Goals

The idea of 'synthetic biomarkers' redefines the current state of biomarker tests by allowing for the non-invasive administration of agents that profile local disease activity and, through multiplexed reporting, provide signature of the underlying disease. The is aim to use the nanoparticle-chaperoned synthetic biomarkers to detect and monitor cancer by measuring local protease activities. Specifically, the goals of this study are:

1. To develop a set of nanoparticle-chaperoned reporters that proteolytically respond to a panel of disease-relevant enzymes
2. To generate a multiplexed set of isobaric peptide reporters for quantification of protease profiles in disease
3. To measure urinary fluorescence and mass signatures of *in vivo* protease activity in animal models of cancer

Chapter 2

Methodology

2.1 Peptide-nanoworm synthesis

Paramagnetic, dextran-coated iron-oxide nanoworms (NWs) with a longitudinal size of ~70nm were synthesized as previously described and derivatized with N-Succinimidyl iodoacetate (SIA) and flurophore, Alexa Fluor 488 (41). The peptide-nanoworm library was prepared from polyethylene glycol 20,000 (PEG) and the Glufib reporter peptides (95:20 mol ratio) by SIA conjugation. The size and shapes of the NWs were analyzed by transmission electron microscopy (TEM) and dynamic light scattering (DLS). Circulation half-life in the blood was characterized in mice. NWs in PBS (5μM FeO) were intravenously injected into nude mice and blood was drawn at various times from the periorbital plexus using heparinized capillary tubes (VWR). Blood samples were treated with 10mM EDTA (in PBS) to prevent coagulation and centrifuged to pellet red blood cells. The samples were analyzed by fluorescence using a fluorescence microplate reader (SpectraMax Gemini EM, Molecular Devices).

The Glufib reporter peptides were synthesized using Fmoc chemistry (Swanson Biotechnology Center) and their composition was verified by mass spectrometry (Table 2.1). Cysteine residues were added to the C-terminus of the peptides to allow for conjugation to the nanoworms. The number of peptides or Alexa 647 dyes per NW was determined from the absorbance spectrum.

2.2 In vitro nanoworm protease activity assays

Forty-three Peptide substrates were screened for specificity to the MMPs (see Figure 2). The substrates were synthesized and conjugated to NWs and these peptide NWs were screened against recombinant MMP-2/8/9 (R&D Systems), MMP-7/14 (AnaSpec, Inc.), Thrombin, Tissue Factor, Factor X_a, and Cathepsin B (Haematologic Technologies, Inc.). The substrates (2.5 μM) were incubated with each of the recombinant enzymes (32nM) in a 96-well plate at 37°C. The increased fluorescence due to proteolytic dequenching of the NWs was measured in 1 min intervals for 3 h using a fluorescence microplate reader.

Ten peptide-NWs were selected from the screen for further experimentation (Table 2.1). The 10-plex NW library in PBS (100nm) was incubated with recombinant matrix metalloproteinases (MMPs) and thrombin for 3.5 h at 37°C. The liberated Glufib reporters were purified by centrifugation with a 10K MW-cutoff filter (Microcon, Millipore). After exposure to UV-radiation (380nm) for 30 min, the purified reporters were dried in a speed vacuum centrifuge. The proteolytic signatures were evaluated by analyzing the reporters using liquid chromatography tandem mass spectrometry (LC-MS/MS).

2.3 In vivo imaging

Tumor-bearing (MDA-MB-435) NCR immuno-deficient mice (Charles Liver Laboratories), aged 3-4 weeks, were injected with NWs loaded with Alexa Fluor 488 or peptides derivatized with VivoTag-680. Bladder accumulation of the reporter peptides and tumor uptake of the NWs were detected using an IVIS imaging system (Xenogen, USA).

For histological analysis, frozen sections of the tumors were prepared. The sections were fixed with 4% paraformaldehyde. Tumor sections were stained for blood vessels (PCAM, Santa

Cruz Biotechnologies) and nuclei (Hoechst). For pathological analysis and detection of NW accumulation, the slices were mounted and imaged by fluorescence microscopy.

2.4 In situ zymography assays

In situ zymography was performed on fibrotic and tumor tissue sections on glass slides. A mixture of 90μL heated 0.5% agarose in PBS, 10μL peptide-NWs (1 mg/mL FeO), and 5μL of 20x Hoechst was applied to dried sections. After the slides were incubated at room temperature over night, the proteolytic liberation of the peptide reporters was evaluated by florescence microscopy.

2.5 Collection and purification of urinary peptides

Immuno-deficient mice were implanted bilaterally with MDA-MB-435 human melanoma tumors and monitored for tumor progression. Fibrotic and tumor-bearing mice were intravenously injected with the 10-plex peptide-NW library (5μM FeO) and background protease inhibitor (Roche complete inhibitor, EDTA free). At 30 min post-injection for the tumor-bearing mice and 2.5 h post-injection for the fibrotic mice, urine was collected for reporter readout. Each cohort included 10 ~ 15 mice. Tumor-bearing mice were injected pre-tumor implantation and at 2 and 4 wk.

To prevent background proteolytic cleavage, collected urine samples were injected with complete protease inhibitor (Roche). After exposing the urine to UV radiation (380nm) for 30 min to liberate the Glufib reporters, an aliquot of it was purified by using anti-FITC magnetic beads (Dynabeads, Invitrogen) and total reporter fluorescence was measured using a fluorescence microplate reader. Remaining urine was purified by TCA precipitation with 100% (w/v) TCA. The Glufib reporters were extracted from the urine by filtration through a macro-

spin column (The Nest Group, Inc.) at an elution fraction of 80% Acetonitrile/0.1% Formic Acid and further analyzed by LC-MS/MS.

2.6 LC MS/MS analysis

Purified urine samples were dried in a speed vacuum centrifuge, resuspended in a mixture of 0.1% formic acid/5% acetonitrile, and analyzed by mass spectrometry (Swanson biotechnology center, MIT). Peptides were isolated using a C18 nanoflow HPLC column (75 micron internal diameter Magic C18 AQ, Michrom BioResources, Inc.) at a flow rate of 300nL/min using a water/acetonitrile mixture with 0.1% formic acid. Analysis by MS was performed using a QSTAR Elite Q-TOF mass spectrometer (AB Sciex).

2.7 Statistics

Euclidean clustering was performed on normalized y_6 ion intensities in R. Class prediction units of self-similarity (USS) k-nearest neighbor (k-NN) leave-one-out-cross validation was performed using GenePattern software (Broad Institute of MIT-Harvard). The Pearson's product-moment correlation coefficients between the different protease signatures were determined using MATLAB. GraphPad 5.0 was used for the ANOVA analyses. SigmaPlot was used for determining the Logistic regressions and ROC curves.

Table 2.1. The iCORE peptide library. Peptide substrates are shown along with their associated mass codes and the unique reporters that are generated for LC MS/MS quantification.

	Probe[a, b]	Substrate	Isobaric mass code[c, d]	y_6 reporter	$[y_6+H^+]$
G1	$e^{+3}G^{+6}$VndneeGFfsAr-X-K(FAM)GGPQGIWGQC-NW	PQGIWGQ	$e^{+3}G^{+6}$VndneeGFfsAr	GFfsAr	683.4
G2	$e^{+2}G^{+6}$Vndnee^{+1}GFfsAr-X-K(FAM)GGLVPRGSGC-NW	LVPRGSG	$e^{+2}G^{+6}$Vndnee^{+1}GFfsAr	$^{+1}$GFfsAr	684.4
G3	$e^{+1}G^{+6}$Vndnee^{+2}GFfsAr-X-K(FAM)GGPVGLIGC-NW	PVGLIG	$e^{+1}G^{+6}$Vndnee^{+2}GFfsAr	$^{+2}$GFfsAr	685.4
G4	eG^{+6}Vndnee^{+2}GFfs^{+1}Ar-X-K(FAM)GGPWGIWGQGC-NW	PWGIWGQG	eG^{+6}Vndnee^{+2}GFfs^{+1}Ar	$^{+2}$GFfs^{+1}Ar	686.4
G5	eG^{+5}VndneeGFfs^{+4}Ar-X-K(FAM)GGPVPLSLVMC-NW	PVPLSLVM	eG^{+5}VndneeGFfs^{+4}Ar	GFfs^{+4}Ar	687.4
G6	$e^{+3}G^{+1}$Vndnee^{+1}GFfs^{+4}Ar-X-K(FAM)GGPLGLRSWC-NW	PLGLRSW	$e^{+3}G^{+1}$Vndnee^{+1}GFfs^{+4}Ar	$^{+1}$GFfs^{+4}Ar	688.4
G7	e^{+3}GVndneeG^{+6}FfsAr-X-K(FAM)GGPLGVRGKC-NW	PLGVRGK	e^{+3}GVndneeG^{+6}FfsAr	G^{+6}FfsAr	689.4
G8	e^{+2}GVndneeG^{+6}Ffs^{+1}Ar-X-K(FAM)GGf(Pip)RSGGGC-NW	f(Pip)RSGGG	e^{+2}GVndneeG^{+6}Ffs^{+1}Ar	G^{+6}Ffs^{+1}Ar	690.4
G9	e^{+1}GVndnee^{+2}G^{+6}FfsAr-X-K(FAM)GGfPRSGGGC-NW	fPRSGGG	e^{+1}GVndnee^{+2}G^{+6}FfsAr	$^{+2}$G^{+6}FfsAr	691.4
G10	eGVndnee^{+3}G^{+6}FfsAr-X-K(FAM)GGf(Pip)KSGGGC-NW	f(Pip)KSGGG	eGVndnee^{+3}G^{+6}FfsAr	$^{+3}$G^{+6}FfsAr	692.4

[a] X (3-amino-3-(2-nitrophenyl)propionic acid), FAM (carboxyfluorescein), Pip (pipecolic acid), NW (nanoworm)
[b] lower case = d-amino acid
[c] photocleaved C-terminus = $CONH_2$
[d] mass = 1589.8 Da

Chapter 3

Results

3.1 Nanoparticle-conjugated fluorogenic 'synthetic biomarkers' proteolytically released into urine

A set of biomarkers that can detect the protease profiles present in disease have been designed. While this work is focused on biomarkers that detect the MMP family, which is usually upregulated in cancer (15), also included are other enzymes commonly found in disease such as the cathepsins, and in blood such as thrombin, Factor Xa, and Tissue Factor (TF). A library of approximately 50 fluorescein-labeled peptides were created to cleave to the enzymes of interest (16-20) and conjugated to iron-oxide nanoworm (NW) (21) nanoparticles. The iron-oxide NWs were shown to have an average diameter of approximately 38nm and have a longer half-life *in vivo* due to the surface-conjugated PEG (Fig. 3-1a,b,c). To characterize the kinetic cleavage of each peptide biomarker, the peptide-NWs were incubated with recombinant versions of the MMPs, cathepsins, and blood enzymes. Proteolytic cleavage was measured by the fluorescence increase associated with proteolytic liberation from the homoquenching of adjacent fluorophores on NWs (Fig. 3-2). From each kinetic profile, the initial velocity of the enzyme reaction was extracted and used to characterize each peptide biomarker (Fig. 3-3. From this library of peptide substrates, 10 peptides associated with intense cleavage by the enzymes implicated in cancer were chosen (Table 2.1).

With a set of 10 surface-conjugated peptide NWs, the ability of these peptides to selectively clear into the urine after release from the NWs in a disease setting was then investigated. For the disease setting, a MDA-MB-435 xenograft murine model of cancer was

chosen. To investigate peptide trafficking, the peptide glutamate-fibrinopeptide B (Glu-fib, EGVNDNEEGFFSAR) was conjugated to the peptide substrates to serve as a reporter for proteolytic cleavage because Glu-fib is biologically inactive and has been found in urine after coagulation, indicating its ability to pass renal filtration. While fluorogenically-labeled Glu-fib intravenously (i.v.) injected into mice freely trafficked into the urine (Fig. 3-4a), the larger nanoworms were unable to pass renal filtration and accumulated in the liver in a liver fibrosis model with dysregulated MMP activity (Fig. 3-4b). The Glu-fib reporters were designed to be able to pass the 2nm glomerular filtration limit, allowing the NWs to serve as a filter and only permitting cleaved reporters to accumulate in the urine. The success of this design was demonstrated with fluorogenically-labeled Glu-fib conjugated to the peptide substrates on the NWs trafficking into the urine upon proteolytic cleavage in the disease setting, (Fig. 3-4c).

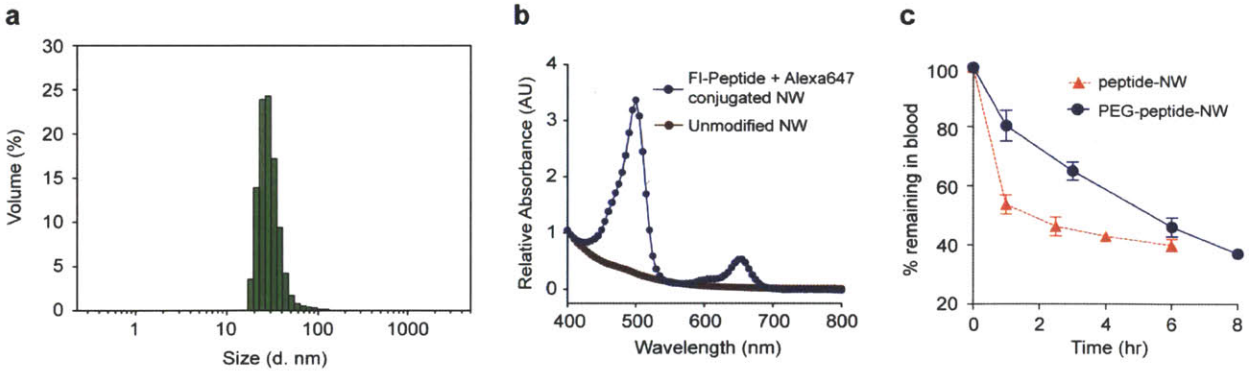


Figure 3-1: Characterization of the iron-oxide nanoworm chaperones: a) Size distribution of the nanoworms as demonstrated by dynamic light scattering. b) The absorbance spectrum of the nanoworms conjugated to fluorogenic peptides (~500nm) +Alexa 647 (blue) and unmodified nanoworms (red). c) Circulation time of the nanoworms conjugated to the peptide biomarkers with (blue) and without PEG (red).

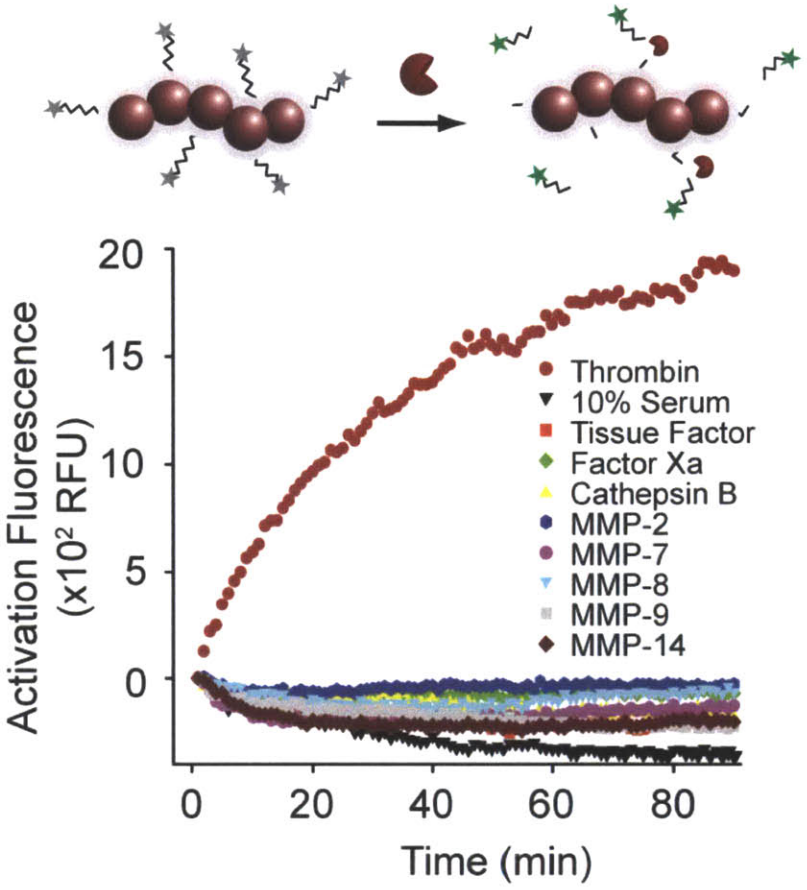


Figure 3-2: Kinetics of proteolytic cleavage of peptide substrates. Peptide biomarkers designed to cleave by proteolysis were conjugated to nanoworms and kinetics were determined by incubation with recombinant proteases.

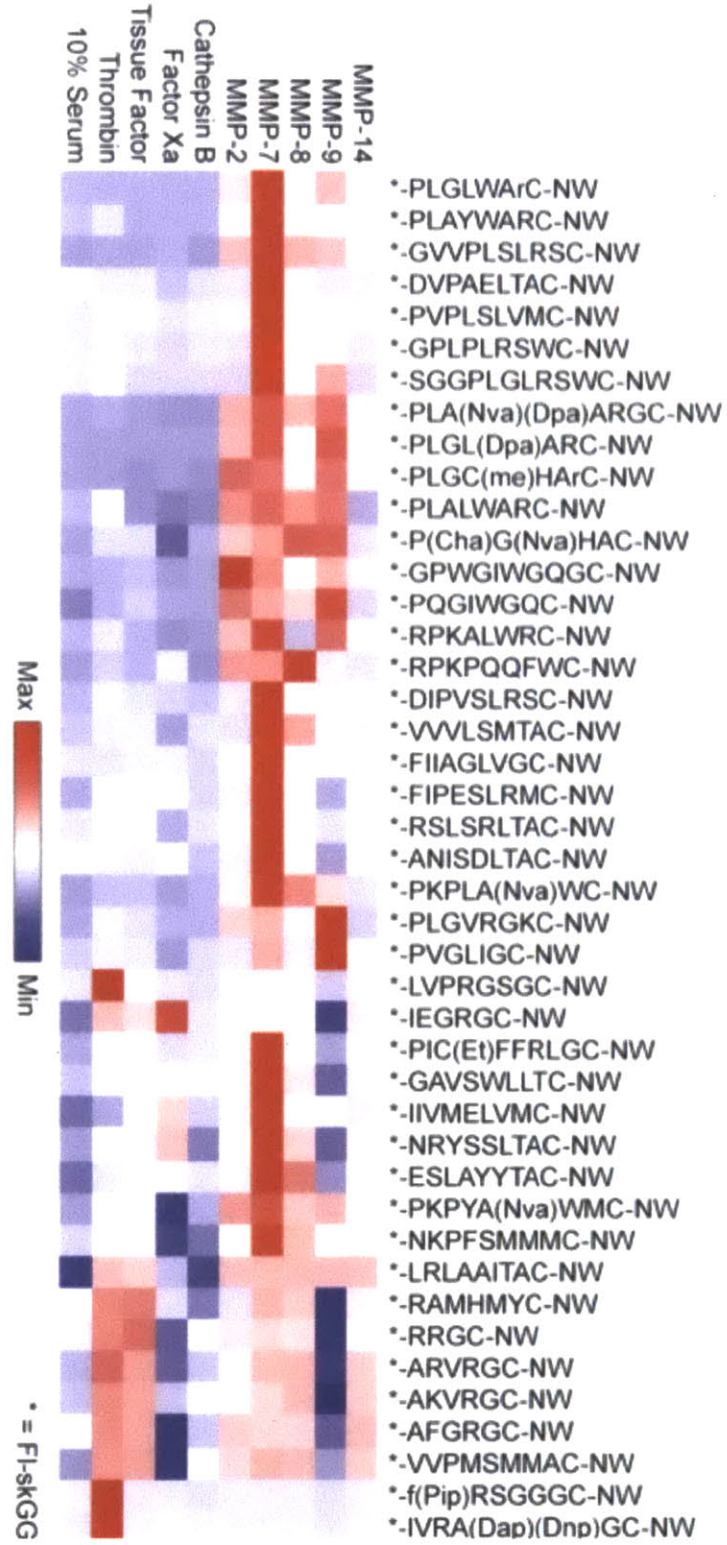


Figure 3-3: Heatmap of the cleavage efficiency of all protease (left)-substrate (top) combinations possible with the chosen library. Cleavage intensities were measured by taking the initial slope of the cleavage kinetics profile (see Fig. 7).

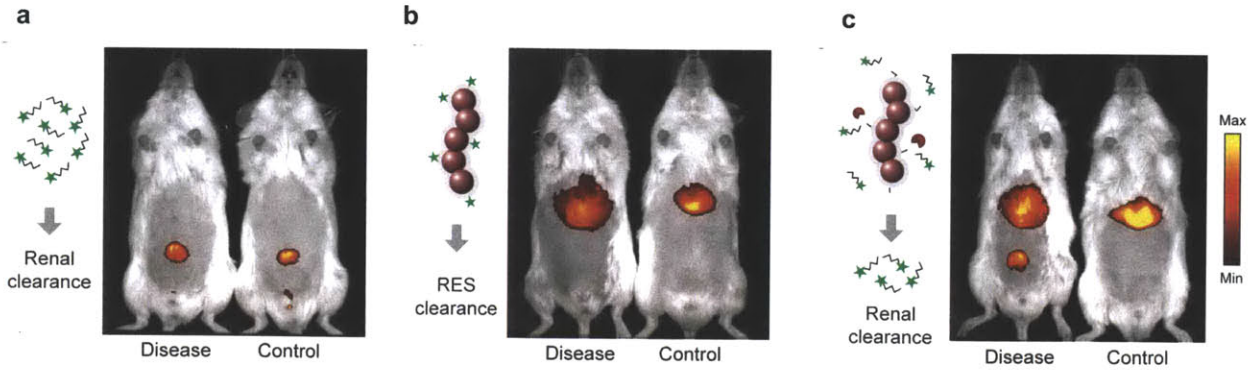


Figure 3-4: Designer biomarkers arrive at the site of disease and accumulate in the urine. a) Unconjugated VivoTag-680-labeled peptide substrates freely clear into the urine as determined by *in vivo* live animal imaging in both the cancer and control animals. b) Biomarker-free VivoTag-680-labeled nanoworms are unable to pass the renal filter and accumulate in the urine in both fibrotic and control animals. c) Nanoparticle-chaperoned peptide biomarkers are cleared into the urine only in cancer animals.

3.2 Multiplexed Protease Profiles in Disease using Mass Spectrometry

While fluorescence analysis of Glu-fib in the urine correlated with disease, this work sought to multiplex the protease profile of disease using the 10 peptide-substrates. Glu-fib is traditionally used as a standard in mass spectrometry because of its high ionization ability. When the doubly charged Glu-fib species is subjected to collision induced dissociation (CID), it fragments into b-type ions and y-type ions (Fig. 3-9a). In the MS spectrum, Glu-fib (MW: 1570.6) appears as a doubly charged parent ion peak (Fig. 3-5a), but upon fragmentation a spectrum of y-type ions and b-type ions appear (Fig. 3-5b). As confirmed by MS/MS, Glu-fib has numerous ion fragments with high signal-to-noise ratios, making it an appropriate choice as a mass reporter because high sensitivity and quantification will be possible. To render Glu-fib resistant to proteolytic activity, Glu-fib was synthesized using d-isomer amino acids and these reporters were conjugated to the N-termini of our peptide substrates using a photo-labile linker consisting of a nitrophenyl group (Fig. 3-6). A photolabile linker is integral to the reporting

scheme because upon cleavage by a protease, the mass reporter is still conjugated to the remaining cleaved substrate portion. Upon UV irradiation, the Glu-fib reporter is freed from the cleaved substrate and is able to then be quantified by mass spectrometry. The scheme was validated by exposing Glu-fib linked to a peptide substrate through the photolabile molecule (compound I, Fig. 3-7a) to UV irradiation. After exposure to UV light, a doubly charged, acetamide-terminated Glu-fib is generated from peptide cleavage at the linker group (compound II, Fig. 3-7b). To test the robustness of the approach to use Glu-fib as mass reporters in animals, the Glu-fib mass tags attached to NWs were injected into tumor-bearing animals and their urine was collected after 2.5 hours. After the urine was irradiated to uncage the Glu-fib mass reporters, LC MS/MS was performed on the isolated mass tags. The MS spectrum revealed multiple peaks, but the Glu-fib parent peak was present with sufficient intensity and revealed that the Glu-fib trafficked into the urine without degradation (Fig. 3-9a). Fragmentation of the Glu-fib tags yielded a MS/MS spectrum with the expected y_6 ions, demonstrating the feasibility of using Glu-fib peptides as mass tags *in vivo*.

To develop Glu-fib into a system that could allow for the multiplexed detection of when the 10 peptide-substrates (Table 2.1) are cleaved, isobaric mass tag technology was incorporated (23,34). Glu-fib was encoded with heavy amino acids such that 10 variants of Glu-fib were generated where each had a parent ion with identical mass, but were designed upon fragmentation by tandem mass spectrometry to yield C-terminal y_6 ions that were each 1Da apart in mass (Fig. 3-8a,b). Of all the C-terminal y-type ions, the y_6 ions demonstrated the highest signal and were thus chosen for the reporter ion. The mass-encoding technology was developed around the y_6 ion fragment (GFFSAR) by creating isotopic versions that incorporated heavy amino acids, yielding a set of 10 Glu-fib reporters that each contained a y_6 ion fragment of

different mass (Fig. 3-10). For each Glu-fib fragment to have equal mass, the y_6 ion was balanced by the remainder of the Glu-fib peptide (EGVNDNEE). This allows each peptide-substrate to be associated with a Glu-fib y_6 ion of unique mass. This approach of creating mass encoded multiplexed reporters was named "isobar Coded Reporters" (iCORE). This approach was validated experimentally by creating an equimolar mixture of the 10-plex iCORE library (Fig. 3-11) and subjecting it to tandem mass spectrometry. As expected, the entire library ionized as a single parent peak (789.95 $\pm$ 0.5 m/z, Fig. 3-10), but upon fragmentation, appeared as 10 single peaks corresponding to the individual mass encoded reporters (683.4-692.4 m/z, Fig. 3-10). To confirm the ability of the peptide substrates to generate a signature upon proteolytic cleavage, an equimolar mixture of the 10-plex iCORE library conjugated to the peptide substrates via the photo-labile linker was incubated with recombinant MMP-9. The cleavage products were isolated by size filtration and subjected to UV irradiation to release the photo-caged Glu-fib reporters. Tandem mass spectrometry showed varying intensities of Glu-fib y_6 ions, revealing a unique iCORE profile for MMP-9 activity (Fig. 3-12). Because naturally occurring isotopes can make quantification difficult for the proteolytic signatures, a specific fragmentation unit mass window was centered on the parent peak to prevent leakage of the isotope peaks. This correction reduced the isotope peak in the MS/MS to approximately 5% of the parent peak. By performing this necessary correction, quantification could then be accomplished through the iCORE technology.

To experimentally demonstrate the potential for this approach, a mixture of 10-plex reporters at defined ratios: 1:2:3:5:10:10:5:3:2:1 was created and could be visualized with the correct ratios by LC MS/MS (Fig. 3-13a). With and without the naturally occurring isotope correction, a linear correlation was observed between the ratios spiked in and the peak intensities

measured (Fig. 3-13a,b). Despite the linear relationship in both cases, all future signatures were corrected for naturally occurring isotopes.

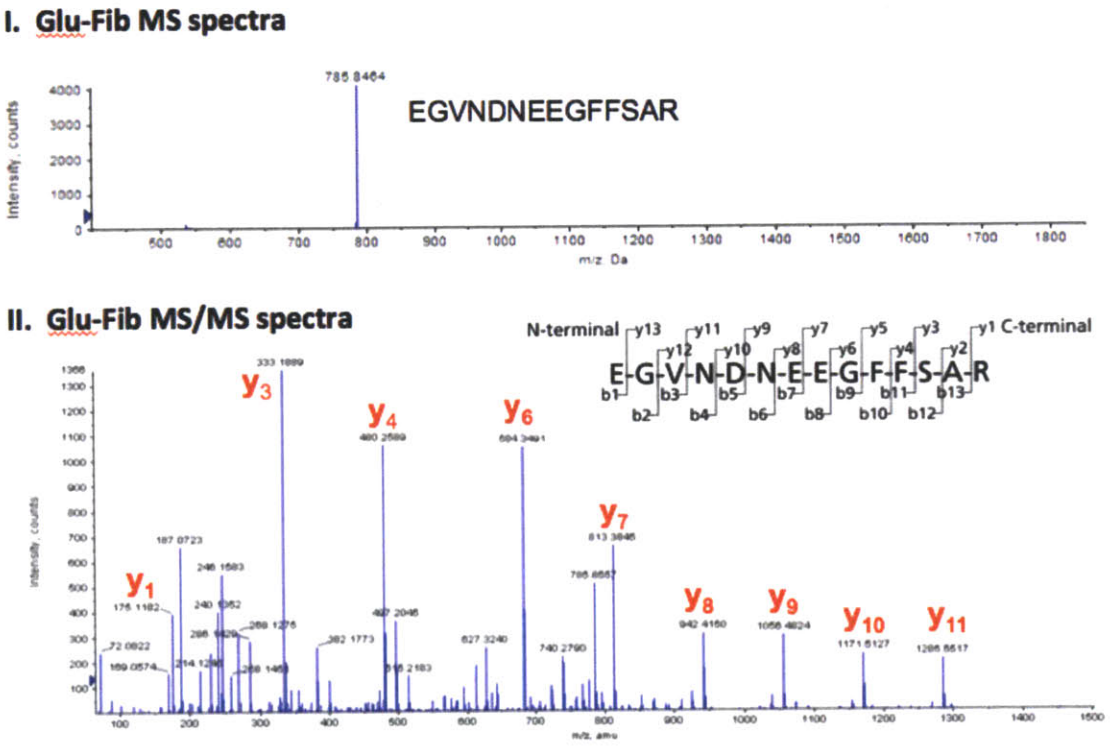


Figure 3-5: Fragmentation pattern of the Glu-fib peptide into y-type ions during LC MS/MS after administration. a) The MS spectrum of Glu-fib showing the parent ion located at 785.8464 m/z. b) The MS/MS spectrum of the fragmented Glu-fib parent peak. The y-type ions are shown. The amino-acid sequence of Glu-fib is shown with the fragmentation positions for the y-type and b-type ions.

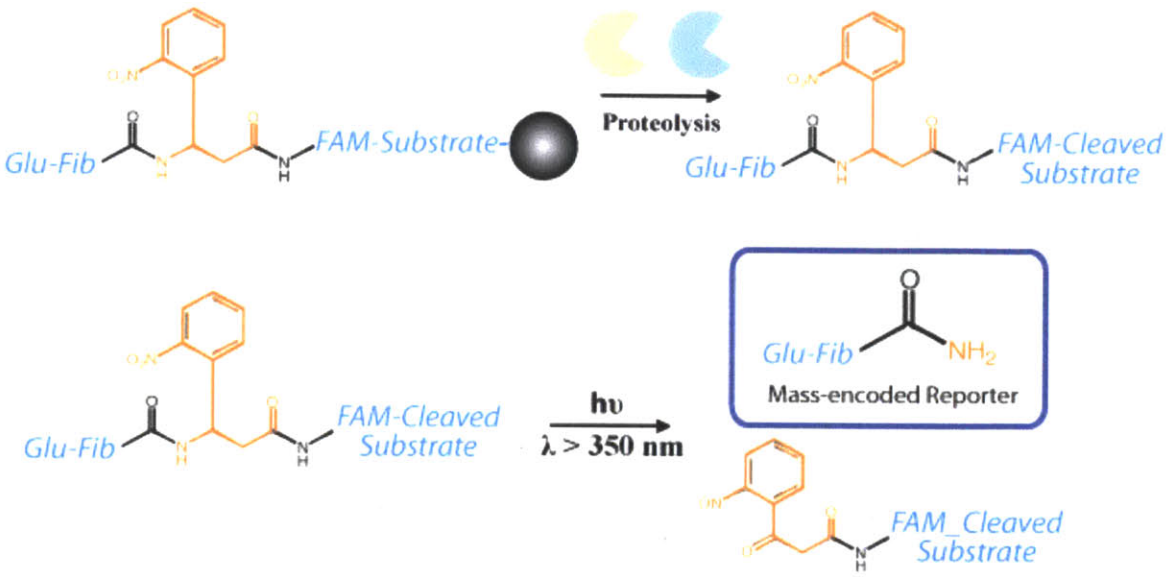


Figure 3-6: The scheme used for connecting the Glu-fib reporter to the peptide-substrate through a photolabile linker. To separate the mass reporter from the cleaved peptide-substrate after proteolysis, UV light is used to cleave the photo-labile linker, uncaging the mass reporter for MS quantification.

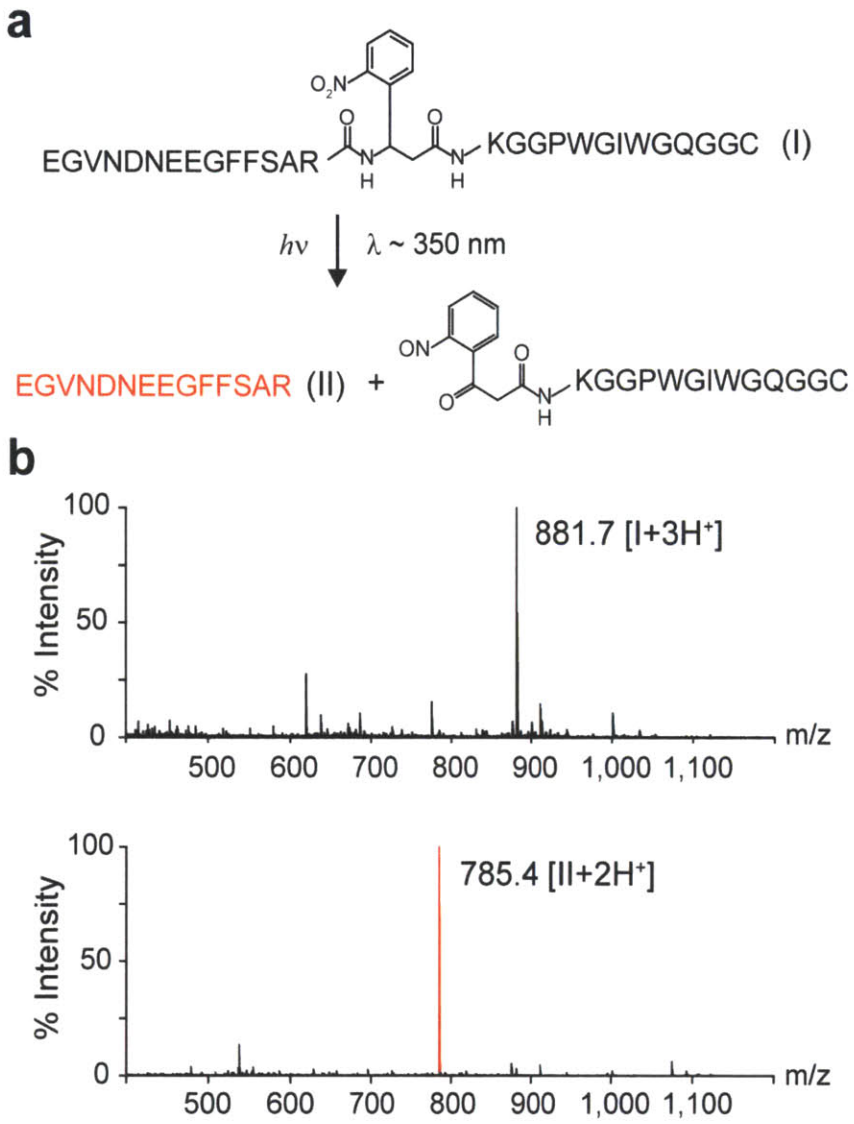
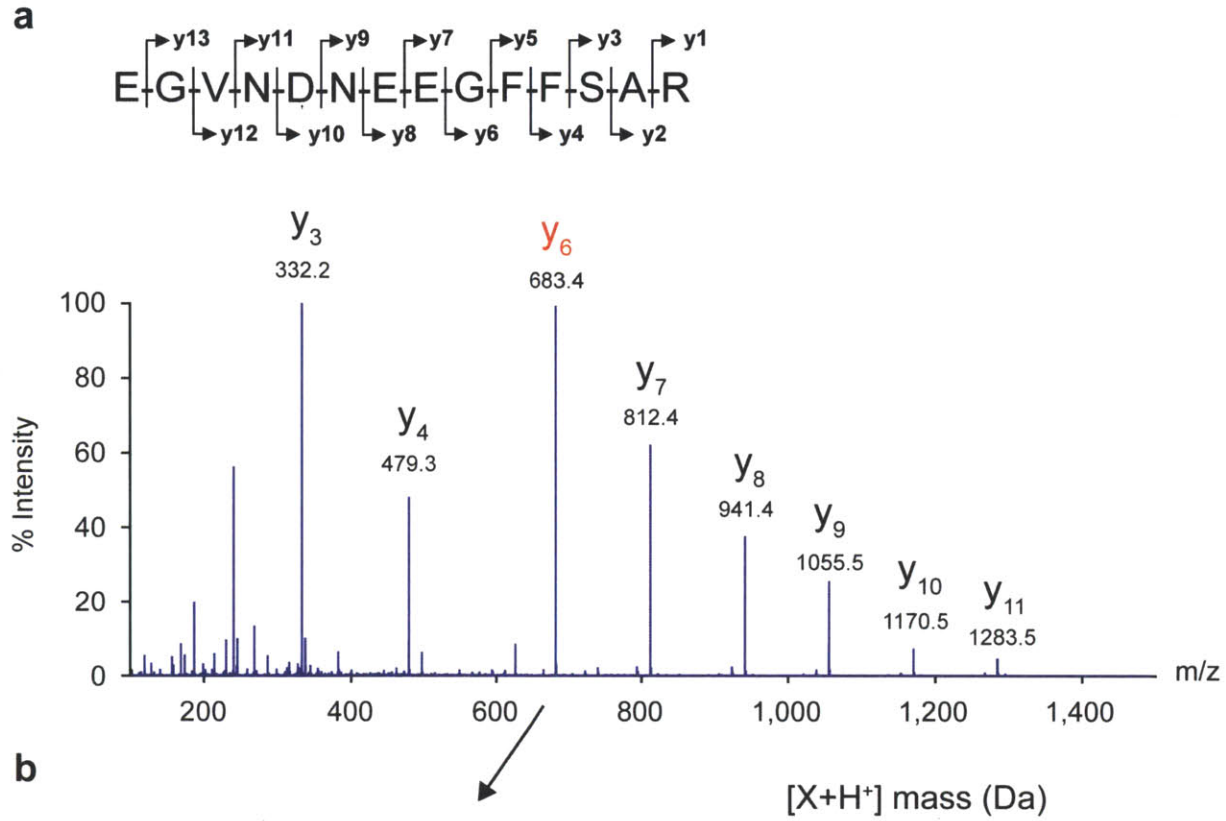


Figure 3-7: UV exposure uncages the mass reporter for MS identification. a) The molecular mechanism for cleavage of the photolabile linker. The mass encoded Glu-fib reporter is attached to the peptide substrate by a photolabile linker (compound I). UV irradiation of compound I activates peptide cleavage and release of the Glu-fib reporter (compound II, red). b) Confirmation by LC MS/MS that the photolabile linker frees the mass reporter conjugated to the cleaved peptide substrate (black, 881.7 m/z) to yield the original mass reporter (red, 785.4 m/z).

a



b

[X+H⁺] mass (Da)

Balance Reporter (y_6 ion)	Balance	Reporter	Total
E^{+3}G^{+6}VNDNEE-GFFSAR	895.3	683.3	1578.7
E^{+2}G^{+6}VNDNEE-$^{+1}$GFFSAR	894.3	684.3	1578.7
E^{+1}G^{+6}VNDNEE-$^{+2}$GFFSAR	893.3	685.3	1578.7
EG^{+6}VNDNEE-$^{+2}$GFFS^{+1}AR	892.3	686.3	1578.7
EG^{+5}VNDNEE-GFFS^{+4}AR	891.3	687.3	1578.7
E^{+3}G^{+1}VNDNEE-$^{+1}$GFFS^{+4}AR	890.3	688.3	1578.7
E^{+3}GVNDNEE-G^{+6}FFSAR	889.3	689.3	1578.7
E^{+2}GVNDNEE-G^{+6}FFS^{+1}AR	888.3	690.3	1578.7
EG^{+1}VNDNEE-$^{+2}$G^{+6}FFSAR	887.3	691.3	1578.7
EGVNDNEE-$^{+3}$G^{+6}FFSAR	886.3	692.3	1578.7

Figure 3-8: Isobaric Coded Reporter (iCORE) mass encoding scheme. A) The tandem mass spectrometry spectrum after collision induced dissociation of Glu-fib. The observed intensities are the C-terminal y6 ion fragment intensities. B) The y_6 ion fragment (red) is the basis for mass encoding 10 variants of Glu-fib that each differ by 1 mass unit. These variants were produced by incorporating heavy amino acids in the y_6 ion and then balancing this in the remaining portion such that each of the 10 variants had equal mass.

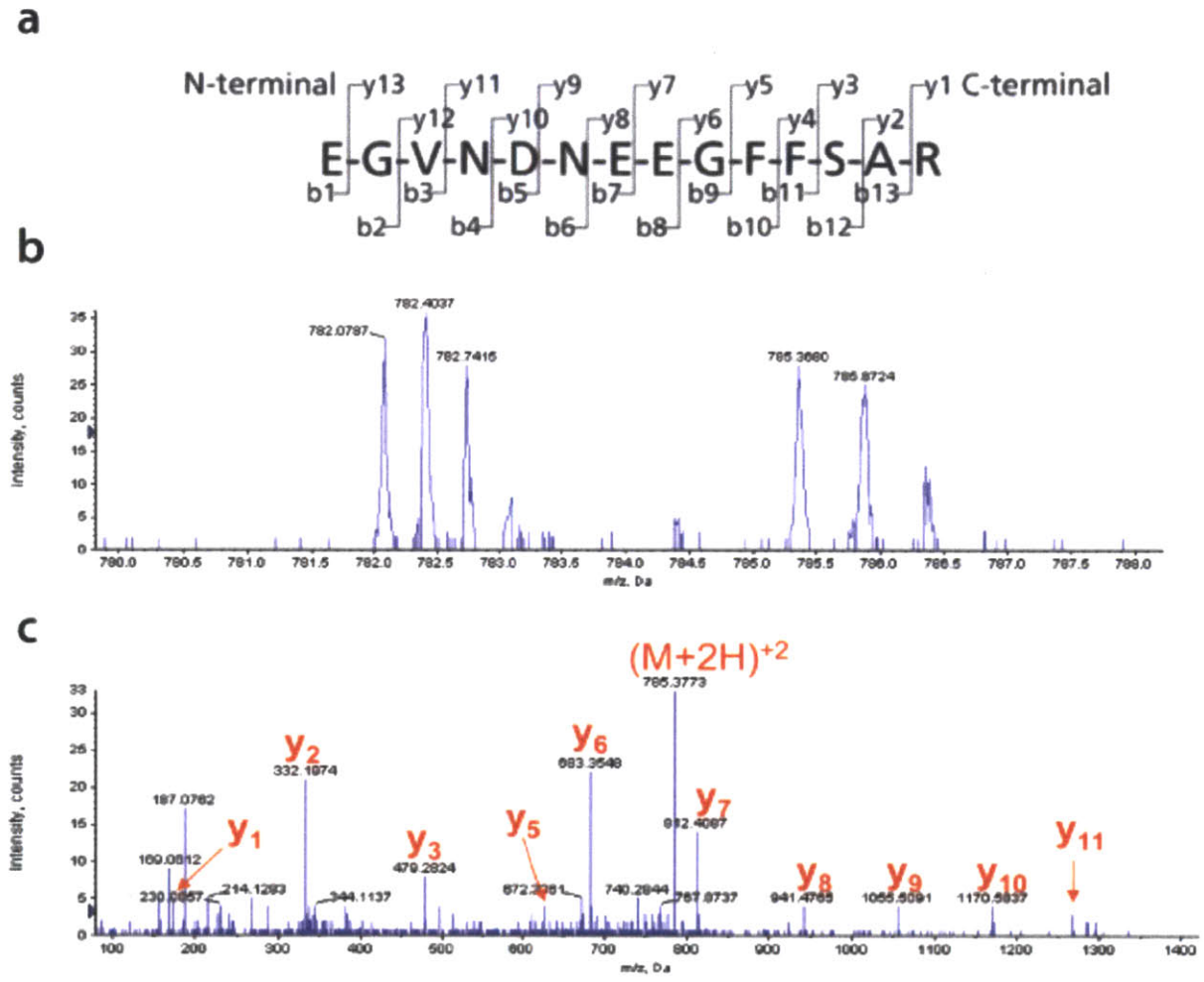


Figure 3-9: Fragmentation pattern of the Glu-fib peptide into y-type ions during LC MS/MS after administration *in vivo* and retrieval through the urine. a) The amino-acid sequence of Glu-fib shown with the fragmentation positions for the y-type and b-type ions. b) The MS spectrum of Glu-fib showing the parent ion located at 785.3690 m/z. c) The MS/MS spectrum of the fragmented Glu-fib parent peak. The y-type ions are shown.

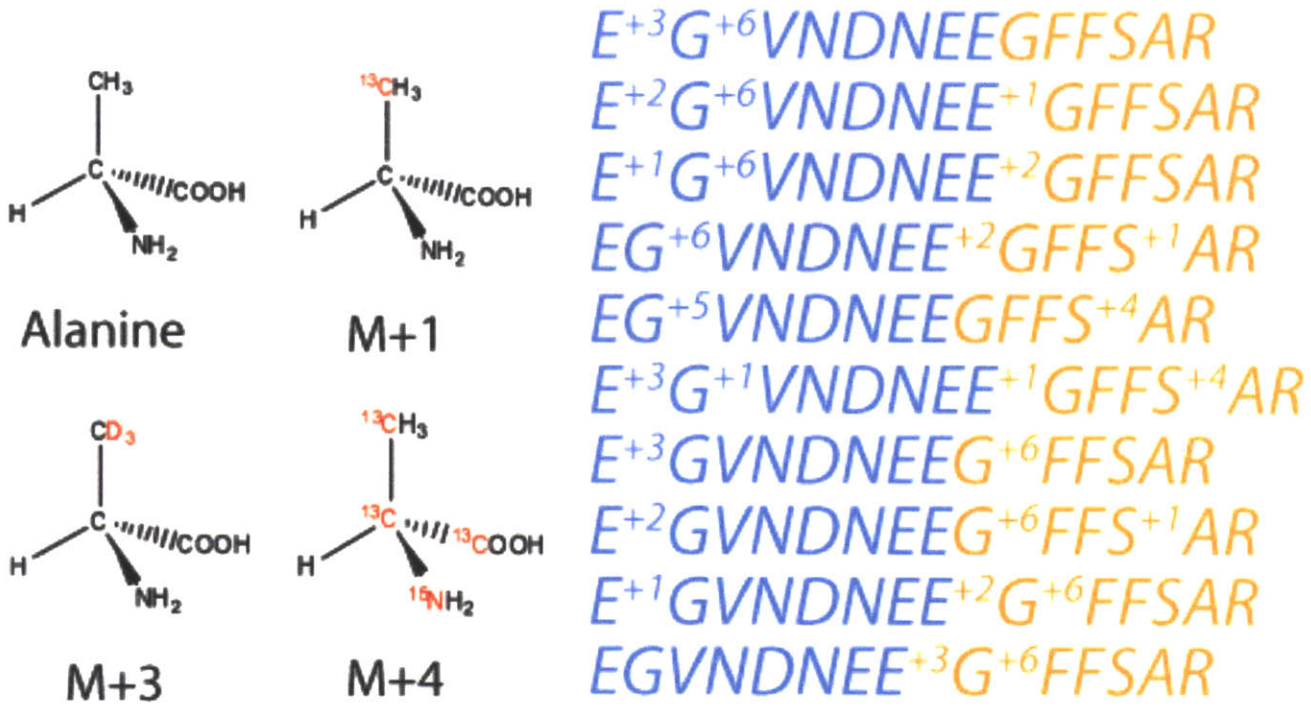


Figure 3-10: Heavy atoms on the amino acids can be used for generating isobaric mass tags.

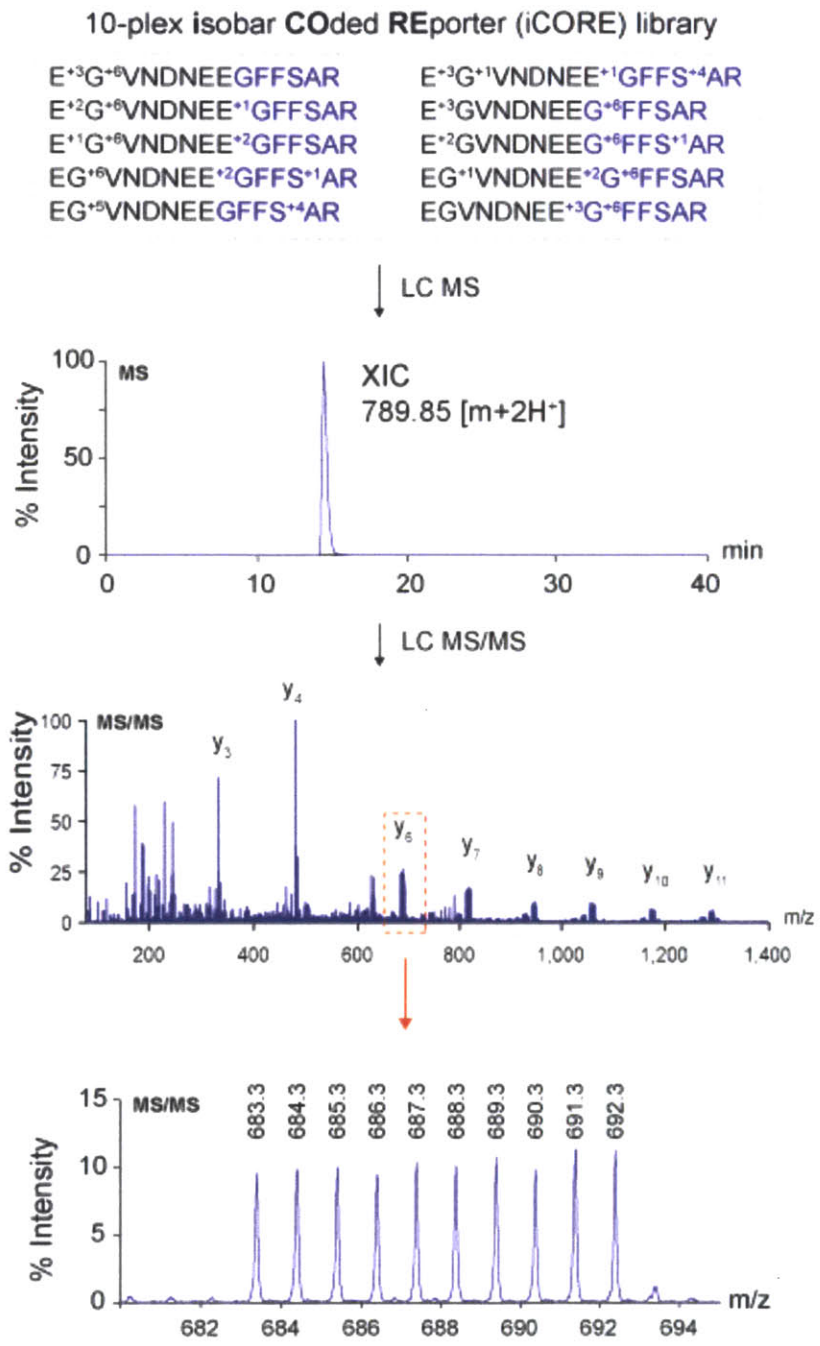


Figure 3-11: The MS/MS spectrum of the 10-plex synthetic biomarkers. The Glu-fib biomarkers fragment into y-type ions with the y_6 fragments of interest. The parent peak is fragmented to yield multiplexed signatures in the MS/MS spectrum and the expanded y_6 ion region is shown. The reporter ions are designed to differ by one mass unit (683.3-692.3m/z)

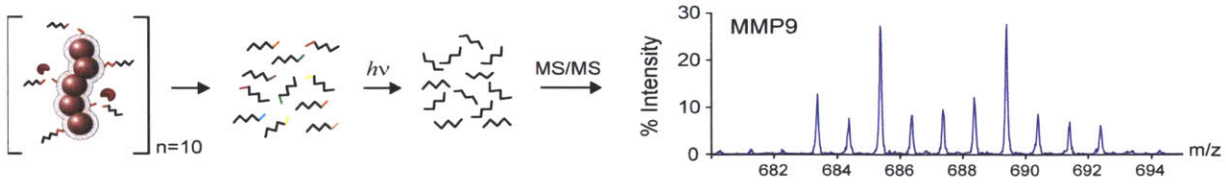


Figure 3-12: The iCORE peptide library allows for multiplexed quantification of protease profiles using liquid chromatography tandem mass spectrometry (LC MS/MS). A profile is shown of the 10-plex library conjugated to nanoworms cleaved by recombinant MMP-9.

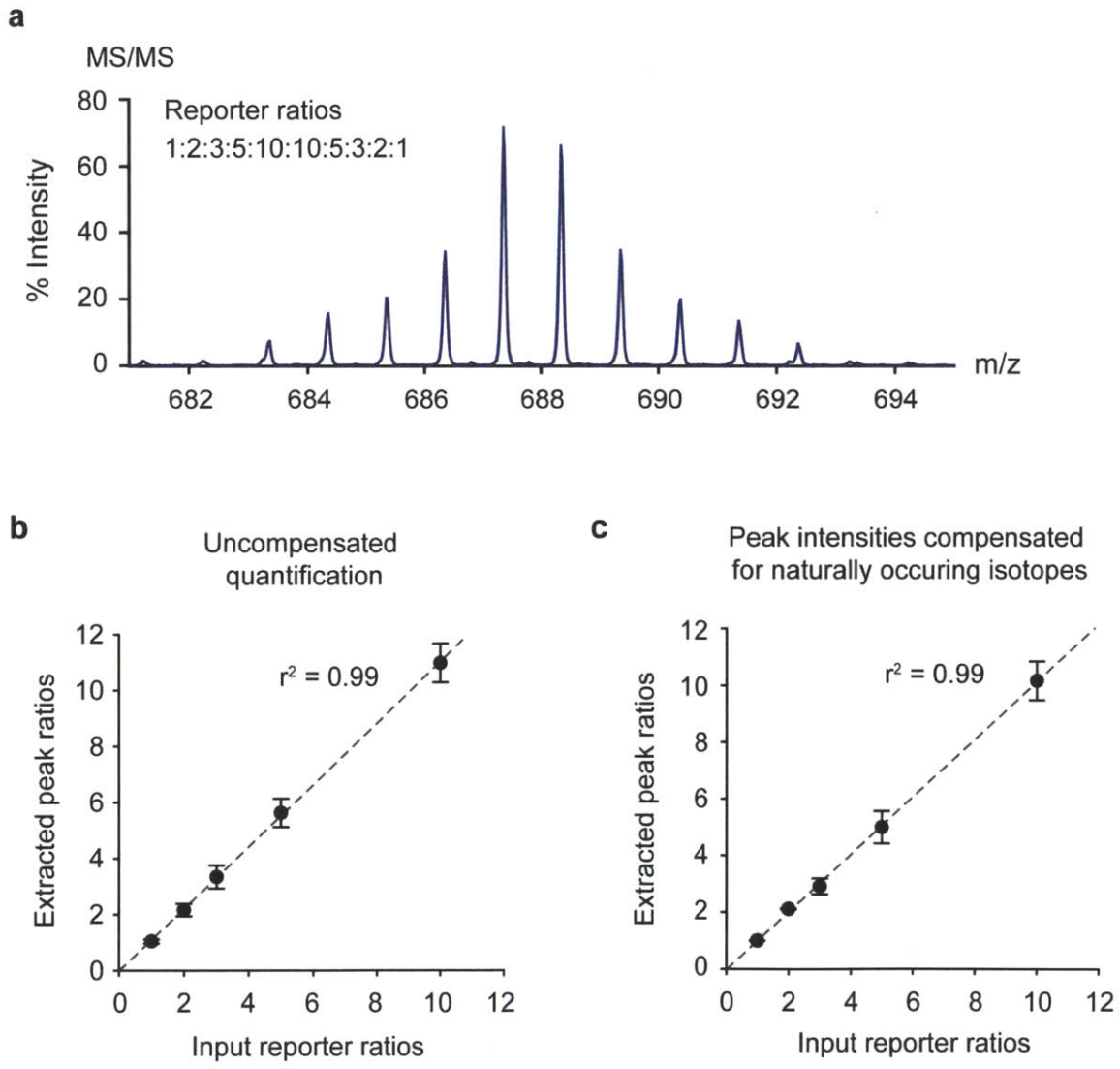


Figure 3-13: Quantification of the 10-plex iCORE library signatures. a) MS/MS spectrum of a 10-plex library mixture with the probes spike in the following intensities: 1:2:3:5:10:10:5:3:2:1. b) In both the uncorrected and c) the corrected naturally occurring isotope case, the quantification of peak intensities correlated well with the input ratios. ($r^2 = 0.99$, n = 3, error = SEM).

To demonstrate how iCORE could yield signatures for different enzymes, proteolytic signatures using the 10-plex library were generated *in vitro*. The 10-plex probe library consisting of photo-caged iCORE mass reporters conjugated to NWs were exposed in equimolar concentrations to recombinant MMP-2, MMP-9, MMP-12, and Thrombin. After incubation, the cleaved reporters were isolated by size filtration and subjected to UV-irradiation to free the photo-caged iCORE mass tags. As expected, the substrate activities were correlated to distinct iCORE signatures consisting of differing y_6 ion intensities (Fig. 3-14a). The signatures for each of the four proteases tested were not correlated in any way (Pearson's correlation analysis, Fig. 3-14b), signifying the potential for the iCORE system to characterize enzymes by mass tag signatures. Because the substrates are designed to cleave to different proteases (Fig. 22), unique signatures could be obtained by proteolysis and quantification by LC MS/MS. The uniqueness of signatures is important because the presence of disease is detected by measuring the different proteases expressed. If unique protease signatures were not possible to obtain using the iCORE methodology, then it would be difficult to achieve accurate diagnosis by the platform.

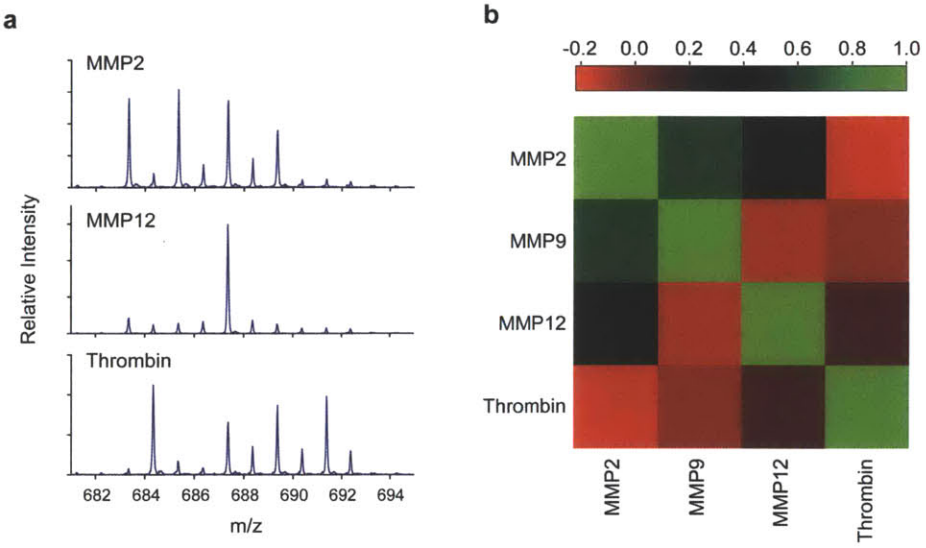


Figure 3-14: iCORE signatures for proteases. a) The 10-plex iCORE library was subjected to recombinant proteases *in vitro* and the MS/MS spectra were generated. b) A heat map of the Pearson's correlation coefficients between the protease signatures reveals the uniqueness of the iCORE signatures.

3.3 Synthetic biomarkers for early and accurate detection of cancer

Early detection of cancer is critical for improving poor survival rates of cancer patients. Most clinically used biomarker diagnostics lack the ability to detect small tumors and many biomarkers have low sensitivity and specificity, such as CA-125 for ovarian cancer and PSA for prostate cancer. Many approaches are shifting to multiplexed analysis of biomarkers to overcome these limitations (4). This work proposes that the 10-plex iCORE panel of 'synthetic biomarkers' could provide the potential to profile proteases in cancer and lead to improved sensitivity and specificity over current biomarkers for early detection of tumors.

For the disease model, melanoma (MDA-MB-435 cell line) tumor xenografts in athymic nude mice were investigated. Because nanoparticles have been shown to passively accumulate at the site of tumors due to the enhanced permeability and retention effect (22) (EPR), the NWs serve as chaperones for delivery of the peptides to the disease site to allow for an amplified signal and earlier detection. For our murine melanoma model, MDA-MB-435 xenografts were established in the hind flanks and tumor progression was monitored at 0, 2, and 4 weeks by injection of the 10-plex iCORE library NWs (Fig. 3-15a). While no significant levels of protease activity were measured at 2 weeks by urine fluorescence of the reporters, levels of the reporters in urine were significant at 4 weeks and these levels were correlated with tumor burden ($p < 0.05$; $n = 10$; Fig. 3-15b,c). The signal observed was confirmed to be due to NW accumulation at the tumor site due to the enhanced permeability and retention (EPR) effect (Fig. 3-16). Urine samples from the 0 and 4 week time points were further analyzed by mass spectrometry to generate a 10-plex signature of the disease. LC MS/MS allows quantification of the signature intensities from diseased and healthy animals using the iCORE platform (Fig. 3-17a). The signatures generated from the mice were analyzed by unsupervised hierarchical clustering using Euclidean distance (Fig. 3-17b). Unsupervised learning is a powerful tool because it is capable of

organizing data into groups of members based on similar characteristics. The clustering algorithm does not require any prior information, basing the grouping purely on observed similarities between data members. In the analysis, distance between probe intensities as the metric to group animals into diseased or healthy groups was used. Successful clustering of most of the diseased and pre-disease animals into two different groups using four probes (G1-G4) with 79% accuracy (n=19) was achieved. Visually it is clear that certain probes were cleaved more (G4) than others (G1) in the disease state. It is clear that varied behavior of the probes is indicative enough to classify diseased and control cohorts of animals correctly.

To further characterize the ability of these biomarker signatures to indicate disease, a larger group of animals was tested (n=36). With this cohort of animals, diseased and control groups were appropriately classified by unsupervised clustering (Fig. 3-18). The predictive nature of the platform was tested for this cohort of animals using a k-nearest neighbor (KNN) leave-one-out-cross-validation analysis (Fig. 3-19). The KNN algorithm bases classification of a member on its n nearest neighbors. The algorithm observes locally a certain member and classifies it to the majority trait of the nearest data members. Leave-one-out cross validation trains the classifier by leave out a data point and using the rest of the data as a training set to classify that one element left out. The process of training the data in such a manner yields a higher accuracy and is used for iCORE approach. As the number of biomarkers classifiers used in the KNN algorithm increased towards 4, the sensitivity and specificity of disease prediction increased, with a maximal prediction of 92% (33/36 animals correctly classified as tumor or pre-disease, sensitivity = 83%, specificity = 100%) at 4 probes (peptide substrates). Including more than 4 biomarkers in the class prediction reduced sensitivity and specificity, indicating that non-informative probes added noise to the data. An important aspect of the KNN algorithm is the

usage of many neighbors to classify each data member. Testing a range of different neighbor size for a 10-probe and 4-probe analysis (Fig. 3-20) showed that 3-nearest neighbors yielded the best accuracy. The ability for multiplexed biomarkers, specifically 4 probes, to better predict disease than single probes shows the promise of screening for 'synthetic biomarkers' that are able to diagnose disease based on proteolytic cleavage.

To demonstrate the advantage of multiple probes for sensitivity and specificity, the iCORE signatures were quantified by mass spectrometry and analyzed by receiver operating characteristic (ROC) using risk score functions determined by logistic regressions for the biomarkers. A logistic regression fits a binomial function to the biomarker data, providing a predicted probability that an animal is diseased or healthy. These predictions can then be used to generate a ROC curve, which is a graphical plot of sensitivity vs. specificity and indicates the tradeoffs between the two characteristics. A ROC curve can be used as a standardized method for evaluating the effectiveness of a diagnostic. Ideal tests will have specificity and sensitivity as high as possible, which is signified by an ROC curve that follows the border along the upper left corner. The predictive value of the diagnostic can be evaluated using the area under the curve (AUC), which indicates the probability that the test will classify a diseased animal correctly.. The higher the AUC the more accurate the diagnostic is in predicting the state of the animal. ROC curves were generated for each of the single biomarkers for the tumor-bearing animals (Fig. 3-21) with the G4 (AUC= 0.76), G8 (AUC = 0.62), and G9 (AUC=0.64) single biomarkers having the most predictive value. A ROC analysis of combinations of biomarkers showed that double (G1 + G4, AUC = 0.80) (Fig. 3-22) and triple combinations (G4+G10+G6, AUC = 0.74) (Fig. 3-23) had improved predictive performance over the single biomarkers. Not all the combinations (Fig. 3-22,3-23) had improved AUC values. It is important to select the right

combinations of biomarkers as some will only incorporate noise into the analysis and reduce the accuracy of prediction. The ability for combinations of certain biomarkers to yield better accuracy demonstrates the potential for multiplexed biomarkers to better diagnose disease.

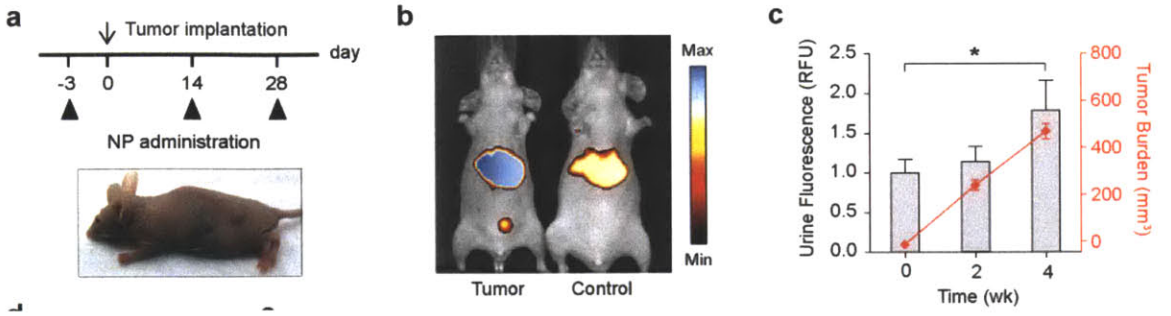


Figure 3-15: Administration of iCORE NW library to mice monitors disease progression. a) The scheme of tumor implantation and iCORE NW injections. b) The trafficking of iCORE reporters into the urine in the diseased animals. c) Correlation of the fluorogenic peptide biomarkers in the urine to the increasing tumor burden over 4 weeks. Significant increase in reporter fluorescence at 4 weeks ($p < 0.05$, ANOVA).

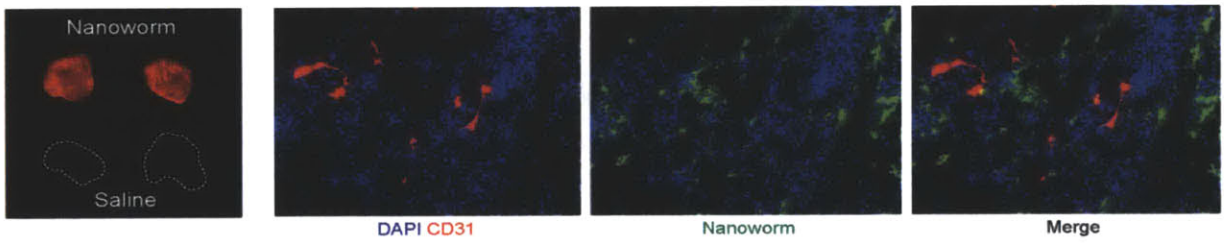


Figure 3-16: Trafficking of the NWs into the tumor site. Fluorescent analysis of tumors removed from mice injected with fluorescently-labeled NWs and saline control show that NWs traffic into the tumor site. (left). Histological analysis of the tumor reveals that NWs (green) extravasate from the blood vessels (red) and enter the tumor tissue.

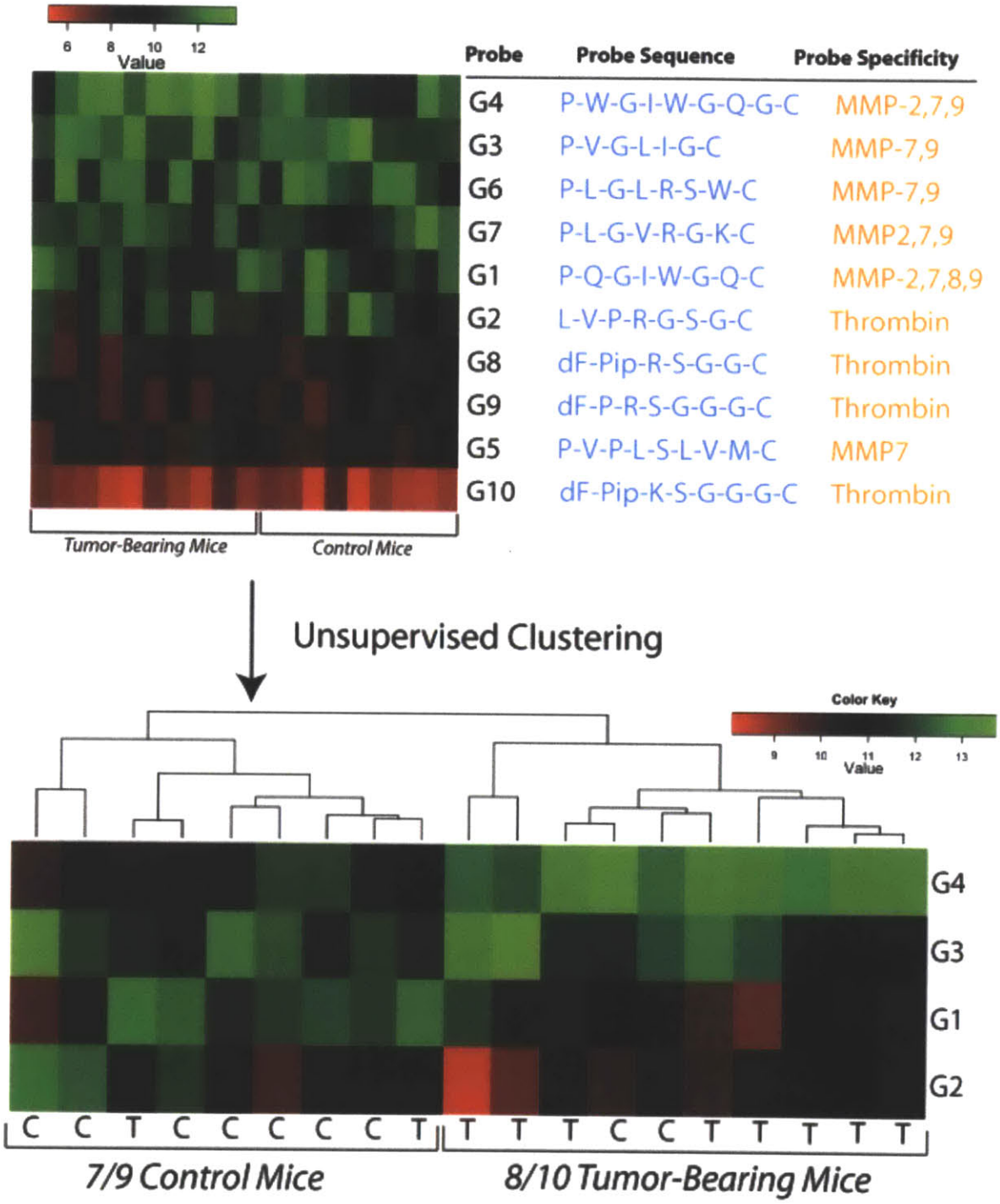


Figure 3-17: Unsupervised clustering of protease signatures enables identification of diseased vs. healthy animals. *Top:* Heat map of proteolytic signatures of the iCORE library from tumor and control animals quantified using LC MS/MS. *Bottom*: Training analysis by unsupervised clustering reveals that four probes (G1-G4) best differentiate the tumor and control animals.

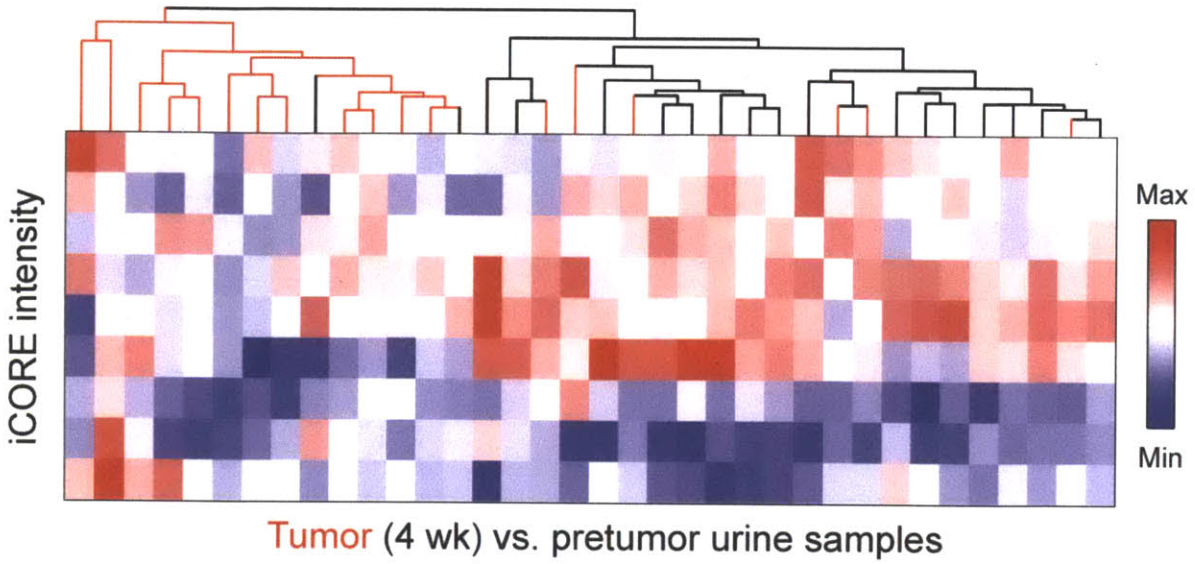


Figure 3-18: Protease signatures accurately classify disease and non-disease groups. Unsupervised clustering of iCORE reporter signatures of an expanded cohort of animals. Rows show probe intensities and each column represents an animal.

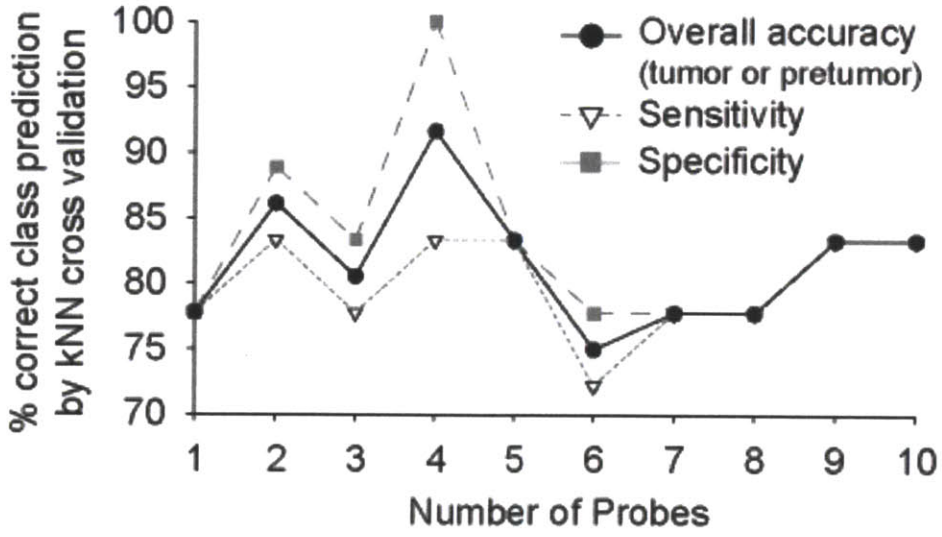


Figure 3-19: k-NN leave-one-out-cross-validation classifies disease vs. non-diseased animals. Accuracy, sensitivity, and specificity are plotted vs. number of probes used.

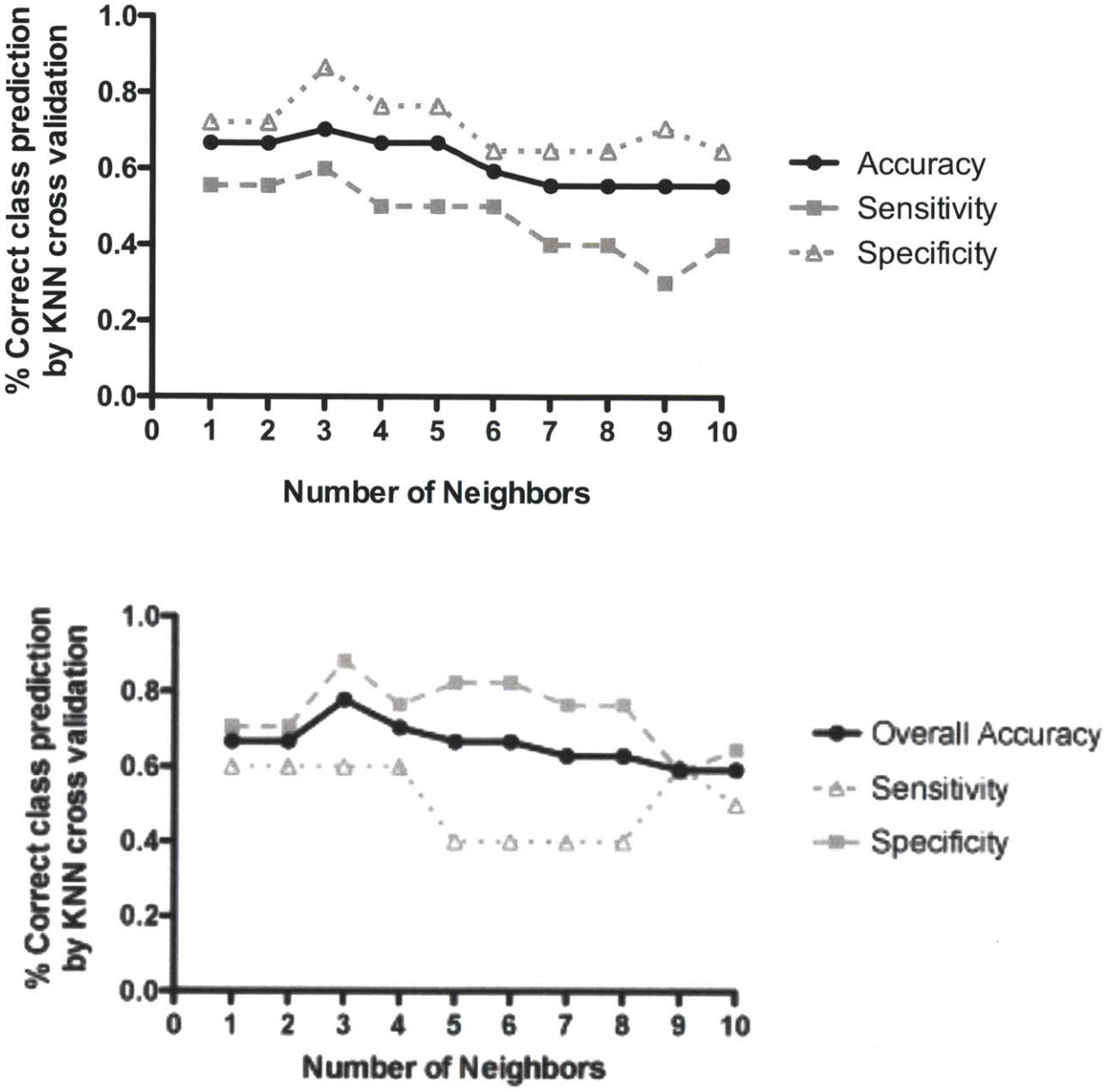


Figure 3-20: k-NN leave-one-out-cross-validation with the number of neighbors varied. *top:* Accuracy, sensitivity, and specificity are plotted vs. the number of neighbors used in the analysis for 10 probes. *Bottom:* Accuracy, sensitivity, and specificity are plotted vs. the number of neighbors used in the analysis for 4 probes. In both situations, 3-nearest neighbor yields the best predictive results.

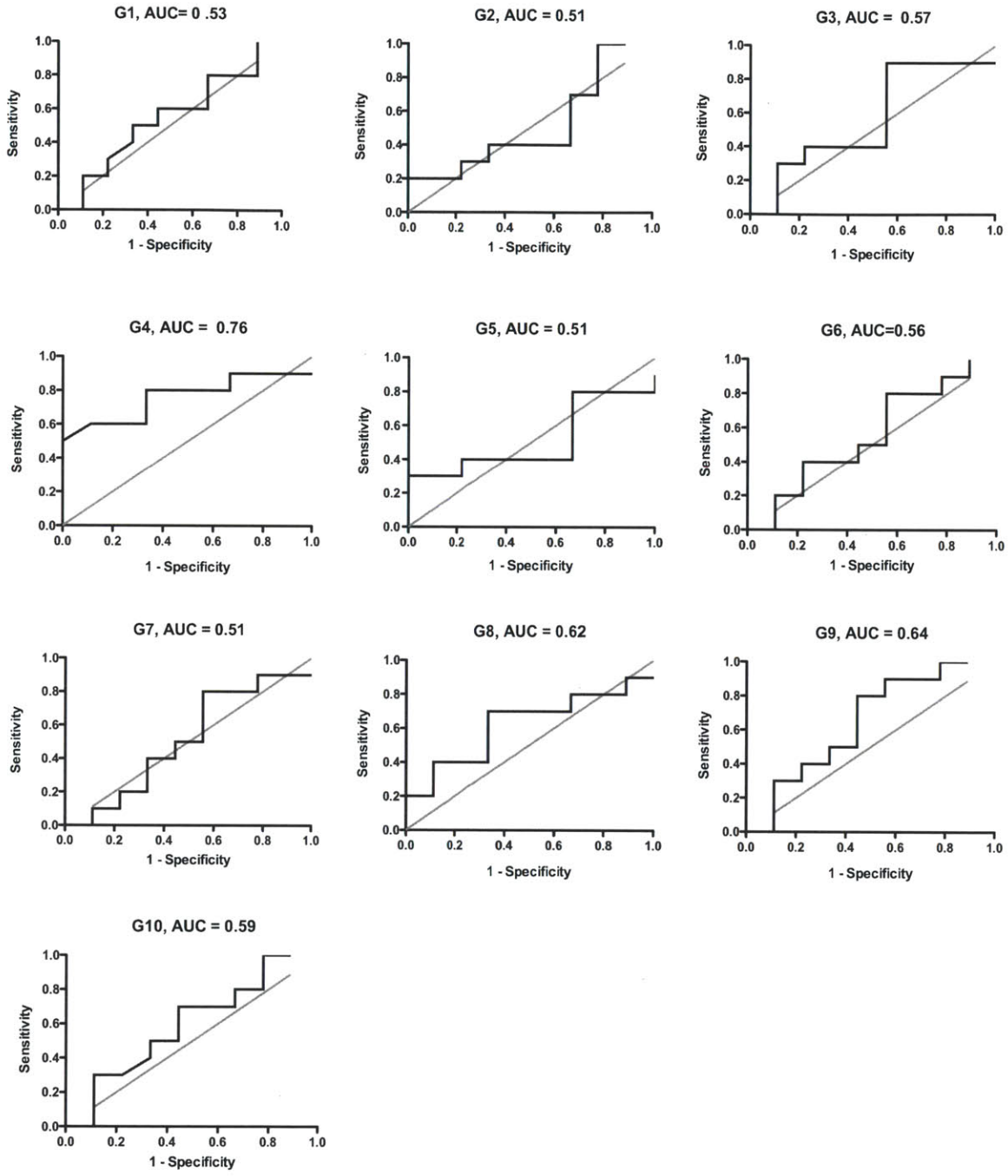


Figure 3-21: The ROC curves for each of the synthetic biomarkers with associated AUC value.

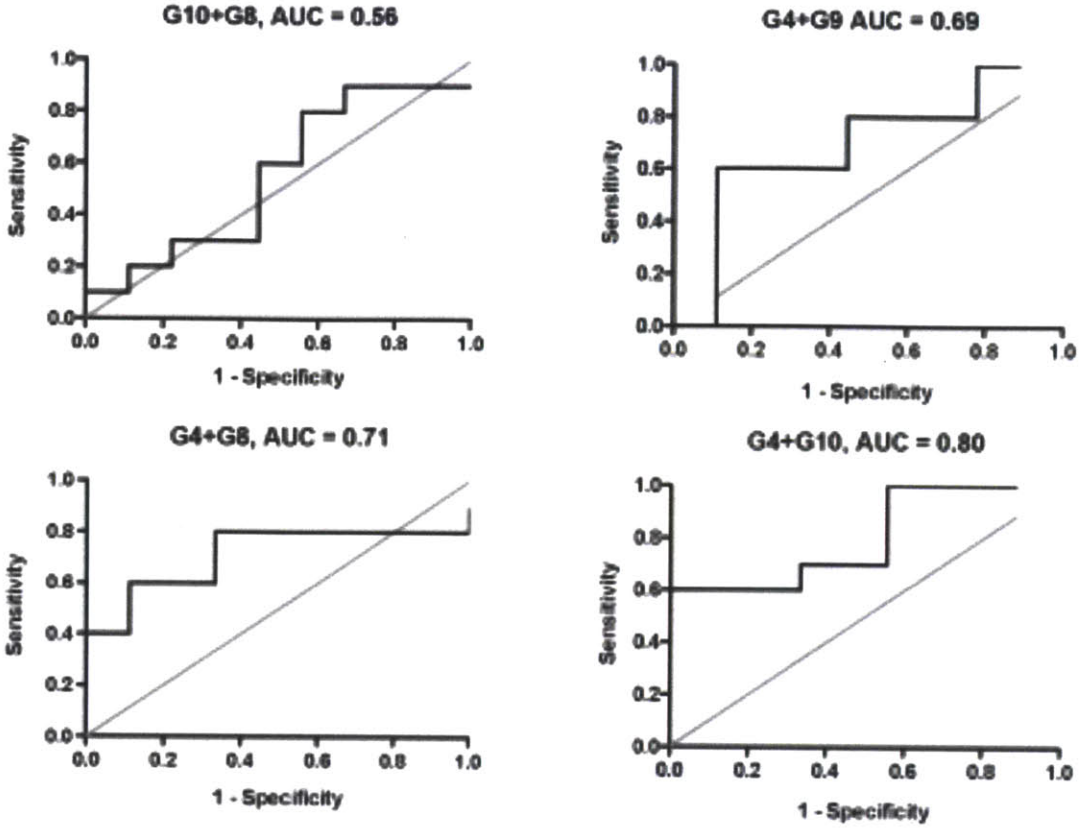


Figure 3-22: The ROC curves for double combinations of the most predictive synthetic biomarkers with associated AUC value.

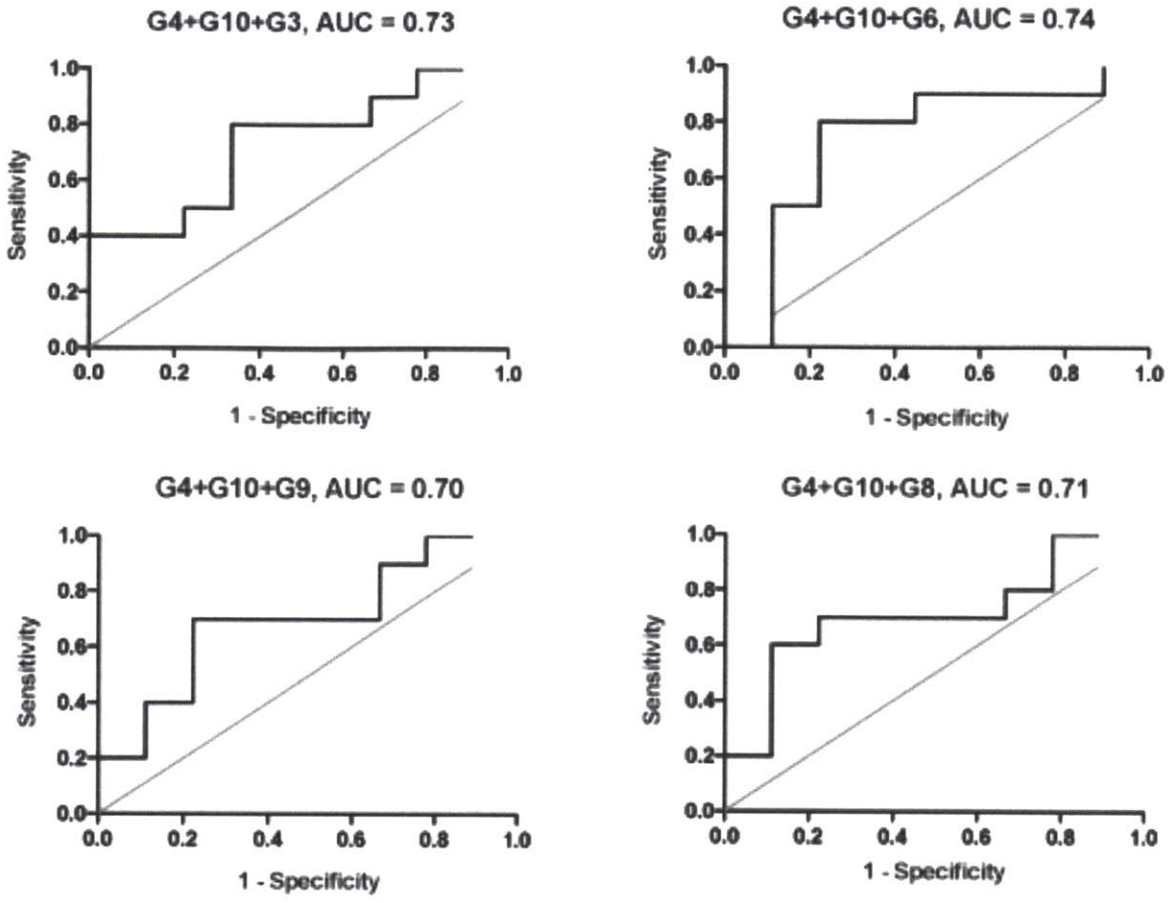


Figure 3-23: The ROC curves for triple combinations of the most predictive synthetic biomarkers with associated AUC value.

Chapter 4

Discussion

Because of the challenges associated with endogenous biomarkers that have limited the reliability and ability for early detection of current diagnostics, a multiplexed library of 'synthetic biomarkers' that are able to traffic into the site of disease and report identifying signatures was developed. The synthetic biomarkers are designed to cleave to disease-relevant proteases, such as the MMPs, to allow for detection and monitoring of cancer. The goals of this project are:

1. To develop a set of nanoparticle-chaperoned reporters that proteolytically respond to a panel of disease-relevant enzymes
2. To generate a multiplexed set of isobaric peptide reporters for quantification of protease profiles in disease
3. To measure urinary fluorescence and mass signatures of *in vivo* protease activity in animal models of cancer

4.1 Development of Nanoparticle-Chaperoned Reporters

Iron oxide NWs were surface functionalized with 'synthetic biomarkers' designed to be cleaved *in vivo* using the relevant proteases and to release reporters that could be detected fluorogenically in the urine and by identifying signatures using mass spectrometry. The NWs are optimal delivery vehicles because they serve a bi-functional purpose: to increase the circulation time for the synthetic biomarkers and to accumulate at the site of disease allowing for proper proteolysis. Because the NWs naturally extravasate into tumors due to the enhanced permeability and retention (EPR) effect, they are a natural choice for diagnosing cancer. Additionally, the

NWs only permit cleaved reporters to clear into the urine because of the 2nm glomerular filtration limit, which prevents the larger NWs from passing through and hinders any uncleaved reporters from being detected. The synthetic biomarkers were chosen by screening through a set of 50 peptide substrates against a panel of recombinant proteases, which included MMPs and clotting cascade enzymes such as thrombin. From this *in vitro* screen, a set of 10 peptide substrates that represented cleavage by the MMPs and serum proteases were chosen to comprise a 10-plex biomarker library. These peptide substrates were then conjugated through a photo-labile linker to a fluorogenic Glu-fib, the mass reporter chosen due to its efficient renal clearance and high ionization efficiency. When Glu-fib is subjected to collision induced disassociation (CID) during LC MS/MS, the Glu-fib fragments into C-terminal y_6 ions. By mass encoding these ions, a 10-plex set of mass encoded reporters can be developed to detect cleavage of specific peptide substrates.

4.2 Designing Reporters for Multiplexed Profiling of Proteases

With a set of peptide substrates for MMPs and other serum proteases, a reporting scheme needed to be designed for the multiplexed detection of substrate proteolysis. While the ability to detect significant urinary bulk fluorescence of these reporters in the diseased versus the control animals was demonstrated, a multiplexed detection scheme was still needed to more accurately diagnose and monitor cancer stages. A mass-encoding scheme was developed around the y_6 ion fragment (EGFFSAR) because of its high ionization and signal in LC MS/MS. A 10-plex set of mass encoded reporters was designed by incorporating heavy amino acids in the y_6 ion fragment, such that each variant differed by one mass unit. The leftover fragment (EGVNDNEE) also incorporated heavy amino acids to balance the weight of the y_6 ion so that each parent fragment

had the same mass and would elute at the same retention point along the extracted ion chromatograph. This scheme allows for a 10-plex signature to be quantified in tandem mass spectrometry which presents an improvement over current protease activity detection technologies. Many of the current activity based probes for profiling proteases utilize fluorescence as the readout mechanism (31). These technologies have been successful at detecting diseases, such as coronary artery disease and tumors (45,46). Using fluorescence as the readout for multiplexing protease activities is quite restricted due to the small number of fluorophores that could be potentially used due to spectral overlap considerations. Thus, combining a fluorogenic approach with a highly multiplexed mass-encoding scheme allows for high-throughput fluorescence analysis of the urine with expanded resolution of disease signatures through mass spectrometry.

4.3 Early and Accurate Detection of Colorectal Cancer

Furthermore, these findings are extended to address the challenge of detecting tumors at early stages (< 1 cm^3) and the limited detection capabilities of endogenous biomarkers. It can be possible for certain tumors to grow for a decade to size > 2.5cm before shedding detectable levels of biomarkers. For current biomarker-based diagnostics to be effective, they will need to have drastically improved sensitivity so tumors can be detected at earlier stages before metastasis is possible.

With this 10-plex iCORE library, success was achieved in differentiating melanoma xenograft-bearing mice from the control ones using unsupervised clustering of the 10-plex signatures. By using the k-NN leave-one-out-cross-validation technique, the biomarkers could be evaluated for their role in predicting early stages of disease. Notably, the best results were

achieved by classifying based on the three nearest neighbors and with four biomarkers, which yielded a prediction rate of 92%. When more biomarkers were included, the predictive value of the signatures was reduced, indicating that the remaining biomarkers introduced higher levels of noise.

Additionally, the diagnostic was evaluated in terms of ROC curves for analyzing the specificity and sensitivity of the iCORE platform. While a few single biomarkers had predictive success, combinations of double and triple biomarkers had even higher accuracy, indicating the need for multiplexed diagnostic technologies. It is important to have multiple biomarkers for profiling the range of proteases present in disease because of their marked variation and promiscuity.

Because of the platform's basis on dysregulated enzyme activities and on a powerful nanoscale delivery vehicle, the iCORE diagnostic can be tailored for many diseases and applications. Many diseases have aberrant protease levels such as liver fibrosis and thrombosis. Liver fibrosis is in fact a disease that could benefit the most from a synthetic biomarker diagnostic because the standard diagnostic procedure is the biopsy. The iCORE platform offers the ability to noninvasively diagnose the disease and even monitor progression and response to therapy. Additionally, the ability to detect smaller tumors at remote sites suggests that our diagnostic can be used for sampling small unknown disease sites that are unreachable with current diagnostics, such as liver metastasis.

Furthermore, the iCORE scheme could be a mechanism for discovering novel biomarkers, allowing for an expanded panel of 'synthetic biomarkers'. The iCORE diagnostic platform is readily expandable to 100s of probes for the detection of nuanced diseases, such as hepatocellular carcinoma, which is difficult to detect, and for extension to other diseases, such as

liver fibrosis and thrombosis, because of its reliance on fundamental proteolytic mechanisms. While the NWs used in this study are tailored for cancer diseases because of their natural extravasation from vessels into these disease sites, extensive literature exists for designing nanomaterials that target cargo to different tissues and organs (47,48). Being able to actively target nanomaterials carrying the 'synthetic biomarkers' to any disease site will allow for this platform to extend to a multitude of enzyme-regulated diseases.

The iCORE diagnostic platform also has the potential to simultaneously detect multiple diseases at the same site. Patients with liver fibrosis or cirrhosis have a high likelihood for developing hepatocellular carcinoma. Because of the multiplexed signatures this diagnostic provides, it may be possible to differentiate a state in which the patient has both diseases. Additionally, larger libraries would allow for mathematical deconvolution of the signatures to identify the precise proteases present in the disease, which would overcome the challenges of screening for substrates that are specific for certain proteases and are not promiscuous.

4.4 Conclusions

Nanotechnology has been leveraged to create a multiplexed set of 'synthetic biomarkers' capable of profiling proteases at local tumor sites and to report non-invasively through the urine for signature quantification by mass spectrometry. The platform was designed to home in to the disease site to allow for increased sensitivity and to release isobaric mass reporters for multiplexed quantification. An analysis of the diagnostic accuracy reveals that combinations of the synthetic biomarkers yield better predictions than individual ones, emphasizing the need to multiplex protease activity for successful diagnosis. This work is a motivation for further

investigation to develop designer biomarkers that probe biological processes for diagnosing underlying disease states.

Bibliography

1. Sawyers, C. L. The cancer biomarker problem. *Nature* 452, 548-552 (2008).
2. Hanash, S., Pitteri, S., & Faca, V. Mining the plasma proteome for cancer biomarkers. *Nature* 452, 571-579 (2008).
3. Danila, D., Fleisher, M., & Scher, H. Circulating Tumor Cells as Biomarkers in Prostate Cancer. *Clin. Cancer Res.* 17, 3903 (2011).
4. Rifai, N., Gillette, M., & Carr, S. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nature Biotechnology* 24, 971-983 (2006).
5. Schulz-Knappe, P., Schrader, M., & Zucht, H. The Peptidomics Concept. *Combinatorial Chemistry & High Throughput Screening* 8, 697-704 (2005).
6. Schwarzenbach, H. et al. Cell-free Tumor DNA in Blood Plasma As a Marker for Circulating Tumor Cells in Prostate Cancer. *Clin. Cancer Res.* 15, 1032 (2009).
7. Wittman J., & Jack, H. Serum microRNAs as powerful cancer biomarkers. *Biochimica et Biophysica Acta – Rev. on Cancer* 1806, 200-207 (2010).
8. Simpson, R. et al. Exosomes: proteomic insights and diagnostic potential. *Expert Review of Proteomics* 6, 267-283 (2009).
9. Anderson, L. & Anderson, N. The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Molecular and Cellular Proteomics* 1, 845-867 (2002).
10. Zhou, H. et al. Collection, storage, preservation, and normalization of human urinary exosomes for biomarker discovery. *Kidney International* 69, 1471-1476 (2006).
11. Hewitt, S., Dear, J., & Star, R. Discovery of Protein Biomarkers for Renal Diseases. *J. Am. Soc. Nephrol.* 15, 1677-1689 (2004).
12. Saghatelian, A. et al. Activity-based probes for the proteomic profiling of metalloproteases. *Proc. Of the Natl. Acd. Of Sci.* 101, 10000-10005 (2004).
13. Fonovic, M., & Bogyo, M. Activity Based Probe as a tool for Functional Proteomic Analysis of Proteases. *Expert Rev. Proteomics* 5, 721-730 (2008).
14. Lopez-Otin, C. & Bond, J. Proteases: Multifunctional Enzymes in Life and Disease. *J. of Biol. Chem.* 283, 30433-30437 (2008).
15. Giambernardi, T. et al. Overview of Matrix Metalloproteinase Expression in Cultured Human Cell Lines. *Matrix Biology* 8, 483-496 (1998).
16. Bremer, C., Tung, C. H., and Weissleder, R. In vivo molecular target assessment of matrix metalloproteinase inhibition. *Nat. Med.* 7, 743–748 (2001).
17. Kridel, S. J. *et al.* A unique substrate binding mode discriminates membrane type-1 matrix metalloproteinase from other matrix metalloproteinases. *J. Biol. Chem.* 277, 23788–23793 (2002).
18. Lutolf, M. P. *et al.* Repair of bone defects using synthetic mimetics of collagenous extracellular matrices. *Nat. Biotechnol.* 21, 513–518 (2003).
19. Mahmood, U. and Weissleder, R. Near-infrared optical imaging of proteases in cancer. *Mol. Cancer Ther.* 2, 489–496 (2003).
20. Turk, B. E., Huang, L. L., Piro, E. T., and Cantley, L. C. Determination of protease cleavage site motifs using mixture-based oriented peptide libraries. *Nat. Biotechnol.* 19, 661–667 (2001).
21. Park, J.-H. *et al.* Systematic surface engineering of magnetic nanoworms for in vivo tumor targeting. *Small* 5, 694–700 (2009).
22. Jain, R. K. and Stylianopoulos, T. Delivering nanomedicine to solid tumors. *Nat. Rev. Clin. Oncol.* 7, 653–664 (2010).
23. Ross, P. L. *et al.* Multiplexed protein quantitation in saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.ralph. *Mol. Cell. Proteomics* 3, 1154–1169 (2004).

24. Thompson, A. *et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by ms/ms. *Anal. Chem.* 75, 1895–1904 (2003).
25. Daniel, K.D., Kim, G.Y, Vassiliou, C.C, Galindo, M., Guimaraes, A.R., Weisslder, R., Charest, A., Langer, R., Cima, M.J., *Biosensors and Bioelectronics* 24, 3252-3257 (2009).
26. Ramaswamy, S. and Perou, C. DNA microarrays in breast cancer: the promise of personalised medicine. *The Lancet* 361, 1576-1577 (2003).
27. Fernie, A. et al. Metabolite profiling: from diagnostics to systems biology. *Nature Rev. Mol. Cell Bio.* 5, 763- 769 (2004).
28. Sebastiani, G. et al. Non invasive fibrosis biomarkers reduce but not substitute the need for liver biopsy. *World J. of Gastroenterology* 12, 3682 (2006).
29. Foda, H. and Zucker, S. Matrix metalloproteinases in cancer invasion, metastasis and angiogenesis. *Drug Discovery Today* 6, 478-482 (2001).
30. Mahmood, U. Near-Infrared Optical Imaging of Protease Activity for Tumor Detection. *Radiology* 213, 866-870 (1999).
31. Miller, M. et al. Proteolytic Activity Matrix Analysis (PrAMA) for simultaneous determination of multiple protease activities. *Integrative Biology* 4, 422-438 (2010).
32. Enoksson, M. et al. Identification of Proteolytic Cleavage Sites by Quantitative Proteomics. *J. of Proteome Research* 6, 2850-2858 (2007).
33. Daniel, K.D., Kim, G.Y, Vassiliou, C.C, Galindo, M., Guimaraes, A.R., Weisslder, R., Charest, A., Langer, R., Cima, M.J. *Biosensors and Bioelectronics* 24. 3252-3257 (2009).
34. Park, J., von Matlzahn, G., Ruoslahti, E., Bhatia, S., Sailor, M. *Angewandte Chemie* 47, 7284-7288 (2008).
35. Svarovsky, S., Szekely, Z., Barchi, J.,. *Tetrahedron: Asymmetry* 16, 587-598 (2005).
36. Baron, A.T., Boardman, C.H., Lafky, J.M., Rademaker, A., Liu, D.C., Fishman, D.A., Podratz, K.C., Maihle, N.J., 2005. Cancer Epidemiology, Biomarkers & Prevention 14, 306-318.
37. Sedlaczek, P., Frydecka, I., Gabrys, M., van Dalen, A., Einarsson, R., Harlozinska, A., 2002. Cancer 95, 1886-1893.
38. Olson, E. et al. Activatable cell penetrating peptides linked to nanoparticles as dual probes for in vivo fluorescence and MR imaging of proteases. *Proc. Of the Natl. Acad. Of Sci. of the USA.* 107, 4311-4316 (2010).
39. Quillard, T. et al. Molecular imaging of macrophage protease activity in cardiovascular inflammation in vivo. *Thromb Haemost* 105, 828-836 (2011).
40. Ruoslahti, E., Bhatia, S. N., and Sailor, M. J. Targeting of drugs and nanoparticles to tumors. *J. Cell Biol.* 188, 759–768 (2010).
41. Sugahara, K. N. *et al.* Coadministration of a tumor-penetrating peptide enhances the efficacy of cancer drugs. *Science* 328, 1031–1035 (2010).
42. Baum, R.P. & Brummendorf, T.H. Radioimmunolocalization of primary and metastatic breast cancer. Q. J. Nucl. Med. 42, 33–42 (1998).
43. Teates, C.D. & Parekh, J.S. New radiopharmaceuticals and new applications in medicine. Curr. Probl. Diagn. Radiol. 22, 229–266 (1993).
44. Dessureault, S. et al. Pre-operative assessment of axillary lymph node status in patients with breast adenocarcinoma using intravenous 99mtechnetium mAb-170H.82. Breast Cancer Res. Treat. 45, 29–37 (1997).
45. Pasqualini, R., Koivunen, E. & Ruoslahti, E. Alpha v integrins as receptors for tumor targeting by circulating ligands. Nat. Biotechnol. 15, 542–546 (1997).
46. Neri, D. et al. Targeting by affinity-matured recombinant antibody fragments on an angiogenesis associated fibronectin isoform. Nat. Biotechnol. 15, 1271–1275 (1997).
47. Goodwin, D.A. & Meares, C.F. Pretargeting: general principles: October 10-12, 1996. Cancer 80, 2675–2680 (1997)

48. Muldoon, L.L. et al. Comparison of intracerebral inoculation and osmotic bloodbrain barrier disruption for delivery of adenovirus, herpesvirus, and iron oxide particles to normal rat brain. Am. J. Pathol. 147, 1840–1851 (1995).
49. Roselli, M. et al. Systemic administration of recombinant interferon alfa in carcinoma patients upregulates the expression of the carcinoma-associated antigens tumor-associated glycoprotein-72 and carcinoembryonic antigen. J. Clin. Oncol.14, 2031–2042 (1996).
50. Link, A.J., Hays, L.G., Carmack, E.B. & Yates, J.R. Identifying the major proteome components of Haemophilus influenzae type-strain NCTC 8143. Electrophoresis 18, 1314–1334 (1997).
51. Shevchenko, A. et al. Linking genome and proteome by mass spectrometry: largescale identification of yeast proteins from two dimensional gels. Proc. Natl. Acad. Sci. USA 93, 14440–14445 (1996).
52. Gygi, S.P., Rochon, Y., Franza, B.R. & Aebersold, R. Correlation between protein and mRNA abundance in yeast. Mol. Cell. Biol. 19, 1720–1730 (1999).
53. Garrels, J.I. et al. Proteome studies of Saccharomyces cerevisiae: identification and characterization of abundant proteins. *Electrophoresis* 18, 1347–1360 (1997).
54. Boucherie, H. et al. Two-dimensional gel protein database of Saccharomyces cerevisiae. *Electrophoresis* 17, 1683–1699 (1996).
55. Gygi, P. et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotech.* 17, 994-999 (1999).
56. Ross, P. et al. Multiplexed Protein Quantitation in *Saccharomyces Cerevisiae* Using Amine-Reactive Isobaric Tagging Reagents. *Molec. And Cell. Proteomics* 3, 1154-1169 (2004).
57. Thompson, A. et al. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Analytical Chemistry* 75, 1895-1904 (2003).
58. Cohen et al. Controlled delivery systems for proteins based on Poly(Lactic/Glycolic Acid) microspheres. *Pharmaceutical Research* 8, 713-716 (1991)