

Washington University in St. Louis

Washington University Open Scholarship

All Theses and Dissertations (ETDs)

January 2011

Digital Protein Analysis: Technologies for Protein Diagnostics and Proteomics through Single-Molecule Detection

Lee Tessler

Washington University in St. Louis

Follow this and additional works at: <https://openscholarship.wustl.edu/etd>

Recommended Citation

Tessler, Lee, "Digital Protein Analysis: Technologies for Protein Diagnostics and Proteomics through Single-Molecule Detection" (2011). *All Theses and Dissertations (ETDs)*. 346.
<https://openscholarship.wustl.edu/etd/346>

This Dissertation is brought to you for free and open access by Washington University Open Scholarship. It has been accepted for inclusion in All Theses and Dissertations (ETDs) by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences

Computational and Systems Biology

Dissertation Examination Committee:

Robi D. Mitra, Chairperson

Barak A. Cohen

Donald L. Elbert

James J. Havranek

Joshua A. Maurer

Gary D. Stormo

**DIGITAL PROTEIN ANALYSIS:
TECHNOLOGIES FOR PROTEIN DIAGNOSTICS AND PROTEOMICS THROUGH
SINGLE-MOLECULE DETECTION**

by

Lee Aaron Tessler

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2011

Saint Louis, Missouri

copyright by

Lee Aaron Tessler

2011

ABSTRACT OF THE DISSERTATION

DIGITAL PROTEIN ANALYSIS:
TECHNOLOGIES FOR PROTEIN DIAGNOSTICS AND PROTEOMICS THROUGH
SINGLE-MOLECULE DETECTION

by
Lee Aaron Tessler

Doctor of Philosophy in Computational and Systems Biology
Washington University in St. Louis, 2011

Professor Robi D. Mitra, Chairperson

Protein quantification tools are valuable in biological and biomedical research and in clinical applications because they enable measurements that elucidate physiological states, including disease states. Protein biomarker detection is likely to play a major role in patient health and aid in the personalization of medicine. However, protein detection has lagged behind other bio-analytical methods due to intrinsic properties of proteins as well as the complexity of biological fluids, such as blood. This thesis describes the design, development, and testing of several technologies for the advancement of protein detection in clinical and research settings. A common thread through these technologies is the isolation of millions of protein molecules on a solid surface, the interrogation of those molecules by fluorescently-labeled antibodies, and the identification and quantification of those molecules by single molecule resolution imaging. By systematically investigating the parameters that have prevented single-molecule quantification of surface-immobilized proteins, we achieved technologies that can

digitally quantify biomarkers in single-plex and multi-plex formats and be used to detect protein-protein interactions. Additionally we characterized a surface coating method that will be of benefit to a wide array of biophysical studies. Finally, we conducted proof-of-principle experiments and computer simulations for the development of a high-throughput proteomics technology that relies on only a small set of probes (8 to 50) to analyze each of the approximately 23,000 human proteins.

ACKNOWLEDGEMENTS

I thank my colleagues in the Center for Genome Sciences and Systems Biology whose collegiality, support, and diverse expertise were invaluable to this work. This includes the members of the Michael R. Brent Lab, the Barak A. Cohen Lab, the Gautam Dantas Lab, the Todd E. Druley Lab, the Justin C. Fay Lab, the Jeffrey I. Gordon Lab, the Gary D. Stormo Lab, the Ting Wang Lab, and formerly the Mark Johnston Lab and the Sean R. Eddy Lab.

I thank the current and past members of the Mitra Lab who made the challenge of a Ph.D. as enjoyable as possible, including German LeParc, K-T Varley, Yue Yun, Michael Brooks, David Mayhew, Francesco Vallania, Maxim Schillebeeckx, Todd Druley, Gabriel Bien-Willner, Xuhua Chen, Jessica Hoisington-Lopez, and Kay Tweedy. Each of them is kind, decent, caring, and highly intelligent. It has been a constant source of pleasure to work alongside them, and they will be sorely missed.

I thank my friends for keeping my life in balance in and out of the lab. I thank my girlfriend Jessica Hoy for her continual advice and support; she has made me a better scientist and a better person. I thank my family for being a source of unconditional support: Veronica keeps me grounded and Michael Henry keeps me looking to the stars. I thank my Mom and Dad, Alexander and Barbara, for their love and inspiration. I am lucky to have had a teacher and a scientist show me the important things in life.

I thank the members of my thesis Research Advisory Committee, Gary Stormo (Chair of meetings), Barak Cohen, Don Elbert, Jim Havranek, and Josh Maurer. Their

advice, criticisms, and suggestions helped form this work into a successful thesis. I am grateful for the time, effort, and genuine interest they invested in me.

I thank my advisor Rob Mitra for his thoughtful guidance, his inspiring degree of ingenuity, and his monastic level of perseverance. I appreciate Rob for the inherent respect he shows to others. From my early days in the lab, he treated me with the esteem that made me feel like a colleague, and that was an encouraging and positive force in my life. I also thank Rob for sponsoring me and my work beyond the campus community. Rob has invited me to several meetings with scientists and businesspeople and encouraged my activity in discussions. Throughout my time in the lab he has continually encouraged me to act on my own ideas even if it entailed investing time and lab funds. Finally, Rob has gone beyond the typical ways that advisors help students acquire skills. Rob made the extremely farsighted decision to sponsor me on a two-week training visit at Helicos Biosciences in Cambridge, MA. That training visit in 2007 enabled much of my thesis. In summary, I thank Rob for giving me the tools to become a successful scholar and professional.

TABLE OF CONTENTS

ABSTRACT OF THE DISSERTATION	ii
ACKNOWLEDGEMENTS	iv
LIST OF FIGURES	xi
ABBREVIATIONS	xiii
CHAPTER ONE: INTRODUCTION	1
CHAPTER TWO: PROTEIN QUANTIFICATION IN COMPLEX MIXTURES BY SOLID PHASE SINGLE-MOLECULE COUNTING	15
ABSTRACT	16
INTRODUCTION	16
RESULTS AND DISCUSSION	19
<i>Iterative Thresholding Improves Detection of Labeled Antibody Molecules</i>	<i>20</i>
<i>Study of Surface Nonspecific Adsorption</i>	<i>21</i>
<i>Robust Immobilization of Protein Ligands on BSA-Coated Glass</i>	<i>22</i>
<i>Efficient Detection of Single Protein Molecules by Antibody Binding</i>	<i>24</i>
<i>Surface Dissociation of Antibodies</i>	<i>25</i>
<i>Dual-round Protein Binding</i>	<i>27</i>
<i>Quantification of Single Molecules by Antibody Binding</i>	<i>28</i>
<i>Accurate Quantification of an Endogenous Serum Protein</i>	<i>30</i>
CONCLUSION	30
MATERIALS AND METHODS	32
<i>Imaging</i>	<i>32</i>
<i>Fluidics</i>	<i>34</i>
<i>Target Proteins</i>	<i>34</i>

<i>Serum Samples</i>	34
<i>Antibodies</i>	35
<i>Preparation of Surfaces</i>	35
<i>Measuring Nonspecific Adsorption</i>	36
<i>Immobilizing Protein Samples</i>	36
<i>Antibody Binding and Oxygen Scavenging</i>	37
ACKNOWLEDGEMENTS	37
FIGURES	38

CHAPTER THREE: NANOGEL COATINGS FOR IMPROVED SINGLE-MOLECULE DETECTION SUBSTRATES

ABSTRACT	45
INTRODUCTION	45
RESULTS	47
<i>Nanogel coatings display lower protein adsorption than BSA or PEG coatings</i>	47
<i>Nanogel coatings resist adsorption by a variety of biomolecules</i>	48
<i>Nanogel-coated surfaces are resilient to detergent</i>	48
<i>Nanogel coatings are compatible with digital antibody binding experiments</i>	49
<i>Investigation into the role of albumin in nanogel capping and cross-linking</i> ...	50
<i>Investigation into alternative coupling chemistries</i>	51
DISCUSSION	52
MATERIALS AND METHODS	53
<i>Synthesizing solutions of nanogels</i>	53
<i>Coating substrates with nanogels</i>	54
<i>Covalently coupled BSA-coated surfaces</i>	55
<i>Multi-arm PEG monolayer-coated surfaces</i>	55

<i>SM adsorption measurements</i>	55
<i>Measuring detergent resistance</i>	56
<i>Digital immunoassays</i>	57
ACKNOWLEDGEMENTS	57
FIGURES	58

CHAPTER FOUR: SENSITIVE SINGLE-MOLECULE PROTEIN QUANTIFICATION AND PROTEIN COMPLEX DETECTION IN A MICROARRAY FORMAT..... 63.

ABSTRACT	64
TECHNICAL BRIEF	64
ACKNOWLEDGEMENTS	69
FIGURES	71

CHAPTER FIVE: PROTEOMICS BY LINEAR MOTIF ANALYSIS AND PARALLEL PEPTIDE SEQUENCING..... 74.

INTRODUCTION	75
DIPEPTIDE MOTIF ANALYSIS.....	76
<i>Overview of the Method</i>	76
<i>Theoretical Performance</i>	77
<i>Generation of Dipeptide Motif Antibodies</i>	78
<i>Purification of Dipeptide Motif Antibodies</i>	79
<i>Analysis of Antibody Cross-Reactivity</i>	80
PARALLEL PEPTIDE SEQUENCING.....	81
<i>Overview of the Method</i>	81
<i>Theoretical Performance</i>	82
CONCLUSIONS	83
MATERIALS AND METHODS.....	85

<i>Producing Polyclonal Antisera Against Dipeptide Epitopes</i>	85
<i>Affinity Purification</i>	85
<i>Immunodepletion</i>	86
ACKNOWLEDGEMENTS	86
FIGURES	87
CHAPTER SIX: CONCLUSIONS AND FUTURE IMPLICATIONS	98
REFERENCES	100
APPENDIX ONE: SUPPORTING INFORMATION	109
SUPPLEMENT TO “PROTEIN QUANTIFICATION IN COMPLEX MIXTURES BY SOLID-PHASE SINGLE-MOLECULE COUNTING”	109
<i>Analyzing Images by Iterative Thresholding</i>	109
<i>Measuring Protein Detection Efficiency</i>	110
<i>Testing for Random Correlations and Specificity of Binding</i>	111
<i>Estimating Nonspecific Binding Based on the Observed Antibody Density</i>	112
<i>Quantifying Total IgG in Rabbit Serum by ELISA</i>	112
<i>Supplemental Figures</i>	113
SUPPLEMENT TO “NANO GEL SURFACE COATINGS FOR IMPROVED SINGLE- MOLECULE DETECTION SUBSTRATES”	119
<i>Design of SM Antibody Adsorption experiments</i>	119
<i>SM Imaging</i>	120
<i>Image Processing</i>	121
<i>Correlation Analysis</i>	122
<i>Cell Adhesion Assay and Time Course</i>	124
<i>Atomic Force Microscopy</i>	126

SUPPLEMENT TO “SENSITIVE SINGLE-MOLECULE PROTEIN QUANTIFICATION AND PROTEIN COMPLEX DETECTION IN A MICROARRAY FORMAT” 127

Coating Glass with Nanogel..... 127

Capture and Detection..... 127

Cell Culture and Protein Extraction..... 128

Imaging..... 129

Supporting Figures..... 130

APPENDIX TWO: DEEP SEQUENCING DETECTS GENETIC POLYMORPHISMS ASSOCIATED WITH RESPIRATORY DISTRESS SYNDROME IN NEWBORNS131

LIST OF FIGURES

Figure 1: Disease outcomes with early detection	10
Figure 2: Protein composition of serum.....	11
Figure 3: Concentration range of the serum proteome	12
Figure 4: Total internal reflection fluorescence (TIRF) microscopy.....	13
Figure 5: Literature of fluorescence single-molecule quantification studies for proteins and DNA	14
Figure 6: Schematic of immunoassay surface architecture, adsorption study, and image processing.....	38
Figure 7: Determination of antibody detection efficiency (dual-color experiment).....	39
Figure 8: Antibody detection efficiency as a function of concentration.....	40
Figure 9: Protein rebinding.....	41
Figure 10: Single-molecule protein quantification	42
Figure 11: Quantification of endogenous IgG protein in serum.....	43
Figure 12: Schematic of nanogel, PEG monolayer, and BSA coatings. TIRF measurements of antibody adsorption	58
Figure 13: TIRF measurements for adsorption of a DNA, protein and enzyme.....	59
Figure 14: Nanogel coating resistance to a surfactant exposure.....	60
Figure 15: Digital antibody binding on nanogel coatings.....	61
Figure 16: Investigation into capping steps and BSA cross-linker. Atomic force microscopy of nanogel coating topology.....	62
Figure 17: Scheme of the single-molecule antibody microarray method	71
Figure 18: Digital versus analog quantification for a single-molecule microarray	72

Figure 19: Measurement of protein levels from a colorectal cancer cell line with and without DNA damage, and detection of a protein complex.....	73
Figure 20: Schematic of Dipeptide Motif Analysis	87
Figure 21: Schematic of Parallel Peptide Sequencing	88
Figure 22: Presence of dipeptides throughout the proteome	89
Figure 23: Dipeptide antibodies needed to identify all human proteins	90
Figure 24: Distribution of dipeptides within the proteome is fairly uniform	91
Figure 25: Dipeptide Motif Analysis is robust to noise	92
Figure 26: Immunodepletion increases specificity after immunization.....	93
Figure 27: Five examples out of sixteen of ELISAs performed on a dipeptide library ..	94
Figure 28: Heatmap of dipeptide antibody specificity.....	95
Figure 29: Specificity of the polyclonal antibodies is tunable.....	96
Figure 30: Cycles of sequencing required to map peptides to the proteome	97
Figure S1: Demonstrating single-antibody detection	113
Figure S2: Attachment efficiency	114
Figure S3: Negative control for binding	115
Figure S4: Dissociation of surface-bound antibody:target complexes is insignificant over 48 hours.....	116
Figure S5: Efficient multiplexing	117
Figure S6: Cell adhesion assay and time course.....	125
Figure S7: Standard curves for p53 and MDM2 single-molecule microarray.....	130

ABBREVIATIONS

5-FU – 5-fluorouracil

AFM – Atomic force microscopy

BSA – Bovine serum albumin

CM – Carboxymethyl

DLS – Dynamic light scattering

DNA – Deoxyribonucleic acid

EDC – 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride

ELISA – Enzyme-linked immunosorbent assay

FN – False negative

FP – False positive

IT – Iterative thresholding

LOD – Limit of detection

LPA – linear polyacrylamide

MPTS – Mercaptopropyltrimethoxysilane

mRNA – Messenger RNA

NHS – N-hydroxysuccinimide

OWLS – Optical waveguide lightmode spectroscopy

PBS – Phosphate buffered saline

PCR – Polymerase chain reaction

PEG – Poly(ethylene glycol)

qPCR – Quantitative polymerase chain reaction

RDS – Respiratory distress syndrome

RNA – Ribonucleic acid

SDS – Sodium dodecyl sulfate

SM – Single-molecule

SMD – Single-molecule detection

SNP – Single nucleotide polymorphism

SPB – Sodium phosphate buffer

SRM – Selective reaction monitoring

SVT – Single-value thresholding

TIRF – Total internal reflection fluorescence

CHAPTER ONE: INTRODUCTION

Diagnostics – the identification and characterization of disease in individuals – is of critical importance to patient health. Despite the diverse ways that diseases form, be it by microbial infection, congenital mal-development, environmental exposure, autoimmunity, genetic background, or combinations of the above factors, all treatment structures are inextricably linked with diagnosis[1].

Evidence suggests that improving diagnostics in terms of accuracy and early detection can be sufficient to reduce patient mortality and morbidity. For example, cervical cancer deaths have been diminished by 74% since the introduction of the Pap Test which can detect the cancer at an early stage[2]. Notably, the Pap Test was a tremendous success against cervical cancer despite the fact that it was introduced in 1928 when scientists had little understanding of the molecular basis of the cancer. Disease detection can save lives even from diseases we know little about[3]. Early detection has also proven valuable in the fights against heart disease[4], diabetes[5], and other types of cancer[6-10], so it is likely to be valuable in many diseases.

Not only can early diagnosis greatly improve patient outcomes, but the lack of early diagnosis can be severely detrimental. Adenocarcinomas are cancers of glandular tissue epithelia and include colon, breast, and pancreatic cancers. Adenocarcinomas are curable; with early detection, the five-year survival rates of colon and breast cancers is over 90% and 95%, respectively (Figure 1)[11]. By contrast, with late detection, those survival rates are decreased to 8% and 16% respectively. Interestingly, pancreatic cancer,

which is an adenocarcinoma for which no early stage test exists and which does not cause symptoms until late stages, is one of the most fatal cancers with a less than 5% five-year survival rate[12]. Taken together, high patient fatality rates from adenocarcinomas are less due to our inability to treat the disease than to the inability to detect the disease until late stages.

The reason that early diagnosis may be so helpful is the nature of disease progression. For example, cancer evolves from a small group of cells, becomes more mature over time, and eventually invades different parts of the body. Detecting a tumor at an early stage implies that the tumor is likely to be relatively small and localized making surgical interventions more effective. Early detection also allows more time for various treatment strategies to be attempted. Thus, the ability to make a diagnosis at an early stage increases the effectiveness of interventions and increases the number of possible interventions[13].

How then, can biomedical science accelerate the development of new and better diagnostic tests? Histological, microbiological and immunohistochemical tests are used to analyze the characteristics of cells and tissues for disease detection. Molecular and biochemical tests allow for measurements of nucleic acids, proteins, or metabolites that are circulating in biological fluids like blood. When these measurements statistically correlate with disease-state across a wide body of individuals, they are called biomarkers. Molecular and biochemical biomarkers are especially promising because they are inherently quantitative and, as opposed to histological methods, enable a view of biological samples at the resolution at which underlying biochemical mechanisms occur.

The collection of human proteins is the proteome, and the analysis of the proteome holds great promise for the discovery of diagnostic biomarkers for several reasons. Firstly, proteins are the major effectors of the physiology of an organism. The Central Dogma of Molecular Biology generally holds true in that DNA and RNA act to produce the appropriate collection of proteins which then carry out cellular functions. The presence of certain proteins indicates cells in the body are carrying out or are poised to carry out a particular physiological function. Secondly, protein expression can change based on environmental and temporal stimulæ, so there is a reasonable likelihood that disease progression may be able to be monitored through routine screenings. Finally, the proteome is diverse: vastly more diverse than the mRNA which codes it. The abundance and activation state of each protein depends not only on transcriptional regulation but splice-form regulation, translational regulation, and post-translational modification[14]. So, protein biomarkers have a greater chance of correlating with physiological states than genomic or transcriptome biomarkers[13].

With the potential so high for the discovery and application of protein biomarkers, what has prevented their realization? The main challenge lies in the measurement of protein abundance, and that is mostly due to the enormous concentration ranges at which proteins exist in biological samples[15]. After removing cells and clotting factors from blood, the remaining serum provides enormous potential for finding biomarkers since it contains blood-specific proteins as well as leakage products from all of the body's tissues (Figure 2)[16]. However, proteins in serum exist over a range of ten orders of magnitude (Figure 3)[17], so when analyzing a serum sample, a protein of interest may be present at one copy out of ten-billion. This "copy number problem"

dominates the proteomic technology landscape to such an extent that, whereas, during the Human Genome Project, the yardstick of technological advancement was “base pairs sequenced per unit of time,” the yardstick in proteomics is “protein copies per unit volume.”[18]

Nucleic acid quantification technologies have been able to get around the copy number problem (e.g. qPCR can detect single DNA molecules via amplification), so why have similar methods for quantifying proteins lagged behind? For one, there is no protein analog to PCR, and that prevents low abundance proteins from being amplified to bulk quantities. Additionally, there is no efficient way to create a molecule that partners to, or binds to, a protein with high affinity, as there are for nucleic acids (i.e base pair complementarity). Finally, nucleic acid methods benefit from the fact that all DNAs and RNAs have similar biochemistry: they are all negatively charged, due to a phosphodiester backbone. By contrast, proteins have diverse biochemistry in several ways. Amino acids have vastly different chemical properties, ranging from acidic to basic and from hydrophobic to hydrophilic. Stretches of similar amino acids can create proteins with very different bulk properties. Also, proteins are often decorated by post-translational modifications such as phosphotyrosines, lysines, and polysachharides, which can greatly affect their charge and hydrophobicity. Thus proteins have proven more difficult to analyze than nucleic acids, and better molecular tools for detecting proteins are necessary for advancing disease detection capabilities.

Proteomics and protein diagnostics are currently carried out by two major classes of technologies, the immunoassay and mass spectrometry[19]. Each of these classes has their classical applications, which can be found in research laboratories and in clinical

diagnostic facilities. Additionally, both the immunoassay and mass spectrometry fields are rife with novel technological variations that enable increased performance in areas like sensitivity, reproducibility, and multiplexing.

Mass spectrometry operates by the ionization of proteins, the fragmentation of proteins into peptides, the determination of the mass of the peptides to a high degree of accuracy, and the comparison of that mass to a database of known peptide masses[20]. This method is often preceded by separating proteins by chromatography to resolve different proteins within a sample. A notable benefit of mass spectrometry is the ability to analyze many proteins with three to four orders of magnitude dynamic range. Additionally, novel mass spectrometric methods have appeared over the last decade that improve sensitivity and extend dynamic range even further. One promising development is Selective Reaction Monitoring (SRM) mass spectrometry which operates by training the instrument on peptides that the experimenter expects to observe in the sample and extends the dynamic range to five orders of magnitude[21]. Despite the promise of SRM, the general utility of mass spectrometry is hampered by several issues[22]. Its main drawback is its inadequately high limit of detection (LOD) which lies at 10^{-9} M. Clinically-relevant biomarkers are present in the 10^{-12} - 10^{-10} M range in serum. Also, despite the fifty year legacy of peptide mass spectrometry, it continues to be highly specialized, needing expert oversight in experimental planning, implementation, optimization, and data analysis. These are some of the shortcomings that have prevented mass spectrometry from becoming widely used to screen populations in clinical settings, where automatability and reproducibility are highly valued.

The immunoassay, or antibody-based detection, is the most widespread way for research and diagnostic laboratories to detect proteins. It relies on the biological phenomena that antibodies – the molecules that confer immunity for organisms – behave by binding non-covalently to their target via physical contact points[23]. Antibodies produced by organisms like rabbits can be isolated from the blood of the rabbits within a laboratory setting, purified, and used for *in vitro* studies. The most common research methods involving the use of antibodies are the Western blot and the enzyme-linked immunosorbent assay (ELISA). The principle behind both of these is that the antibody is labeled by a reporter molecule (such as an enzyme that produces a change in the optical absorbance of a liquid, a fluorophore, or a magnetic particle) and then the labeled antibody binds to its target if the target is present. After a wash step to remove un-bound antibody, the amount of the remaining antibody, which is bound to its target, is read by an appropriate detector such as an absorbance spectrometer, fluorometer, or electromagnetic sensor. The signal produced by the remaining antibody, (subtracted from a target-less control to account for non-specific binding of the antibody) defines an analytical measurement of the abundance of the target.

The Immunoassay, and ELISA in particular, is widely use in diagnostics due to its high sensitivity (LOD down to 10^{-15} M), practical dynamic range (three to four orders of magnitude), and automatability. They have been developed into a variety of platforms that have been shown to improve performance in terms of multiplexing and sensitivity. One solution to multiplexing the immunoassay has been the development of antibody microarrays[24]. The main advantage of this technology is that is increases the number of targets that can be analyzed given a volume of sample. Other technological innovations

have included proximity ligation assays[25, 26], bead-based arrays[27, 28], and hybrid antibody-DNA detection methods[29]. Interestingly however, no immunoassay had been shown that employed single-molecule detection, prior to the initiation of this thesis.

Single-molecule detection describes analytical techniques that are able to identify the presence of an individual molecule, and most typically, fluorescent molecules. Single-molecule techniques can be achieved by a variety of optical configurations, and all of these require efficient elimination of background fluorescence [30]. These optical techniques include narrow-field epifluorescence[31], fluorescence correlation and cross-correlation spectroscopy[32] via laser confocal microscopy[33], and various other methods that restrict the fluorescence excitation volume[34, 35]. By one method, total internal reflection fluorescence (TIRF) microscopy, the excitation volume is confined to a one-hundred nanometer section of the specimen, enabling high signal-to-noise for molecules located on the imaging surface (Figure 4) [36, 37].

Single-molecule detection holds great potential for biomarker analysis since with it comes the ability quantify molecules with digital resolution, i.e. by counting each molecule. Single-molecule detection and quantification has proven to be a boon in the DNA sequencing world. Since the development of Polony (polymerase colony) technology[38], which isolates individual molecules and performs separate sequencing reactions in parallel, next-generation sequencing technologies have exploded[39]. (Until 2009, the number of sequenced human genomes was less than ten. By the end of 2010 that number was approximately 3,000. By the end of 2011, that number is estimated to be approximately 30,000[40].) The single-molecule nature of these technologies has been proven to be more quantitative, accurate, and precise than previous analog methods like

the microarray[41, 42]. One of the reasons for this is that dynamic range becomes trivialized to the task of sampling more molecules (e.g. by expanding the imaging area). Theoretically speaking, to sample an event present at a one-in-a-million, one needs only to image around one million molecules at single molecule resolution – a task attainable on 1 cm² of surface. Despite this potential, a deep survey of the literature at the outset of this thesis and until 2009 shows that no single-molecule method had been published that examines immunoassay protein detection on a solid surface (Figure 5).

With a strong impetus that single-molecule detection could greatly benefit the world of protein detection, I embarked on this thesis to develop single-molecule tools for protein diagnostics and proteomics. The aim of Chapter Two of this thesis is to develop a single-molecule immunoassay on a solid surface, and in the process, uncover the technological barriers that exist to single-molecule diagnostics. Throughout the course of that aim I discovered that low-background surface chemistry was one of the major barriers to the adoption of single-molecule methods into protein detection. This paved the way for Chapter Three, in which I examine a recently published biomaterial called poly(ethylene glycol)-bovine serum albumin (PEG-BSA) nanogels[43] and study its use in single-molecule detection applications. Initially directed as a study to characterize the nanogels for protein diagnostic applications, I was satisfied to find the nanogel biomaterial to have benefits to general biophysical studies involving single-molecule imaging. Having established the principle techniques in single-molecule imaging, surface chemistry, and surface architecture, I sought to fully integrate single-molecule detection into the immunoassay toolkit. I chose the antibody microarray as the technology that can most readily benefit from single-molecule detection, and in Chapter Four, I provide a

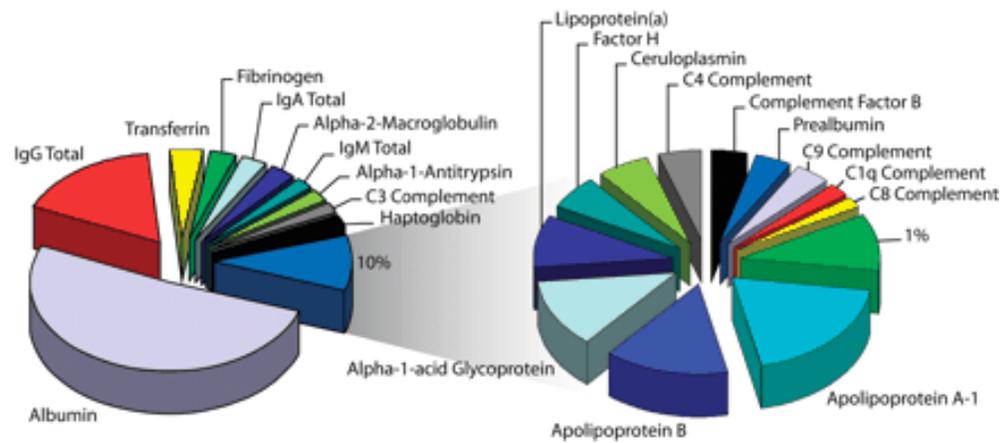
proof-of-principle for the use of single-molecule imaging and nanogel surface chemistry in the antibody microarray arena. Finally, in Chapter Five, I introduce a novel proteomics technology that relies neither on mass spectrometry nor on a large collection of antibodies. With this technology, a small number of antibodies (8 to 50) could be used to obtain sequence information from denatured proteins or peptides in a massively parallel fashion. These antibodies, which detect peptide sequence information, may be bound and washed away in sequence with one another and detected by single-molecule imaging to identify and quantify the collection of over 23,000 human proteins in a single experiment. In addition to introducing the method and its two variations, I provide computational and experimental data that support the development of the technology.

Figure 1

Cancers	5-year survival rates	
	Late detection	Early detection
Lung	< 10%	48%
Colorectal	8%	90%
Breast	16%	>95%
Prostate	33%	>95%

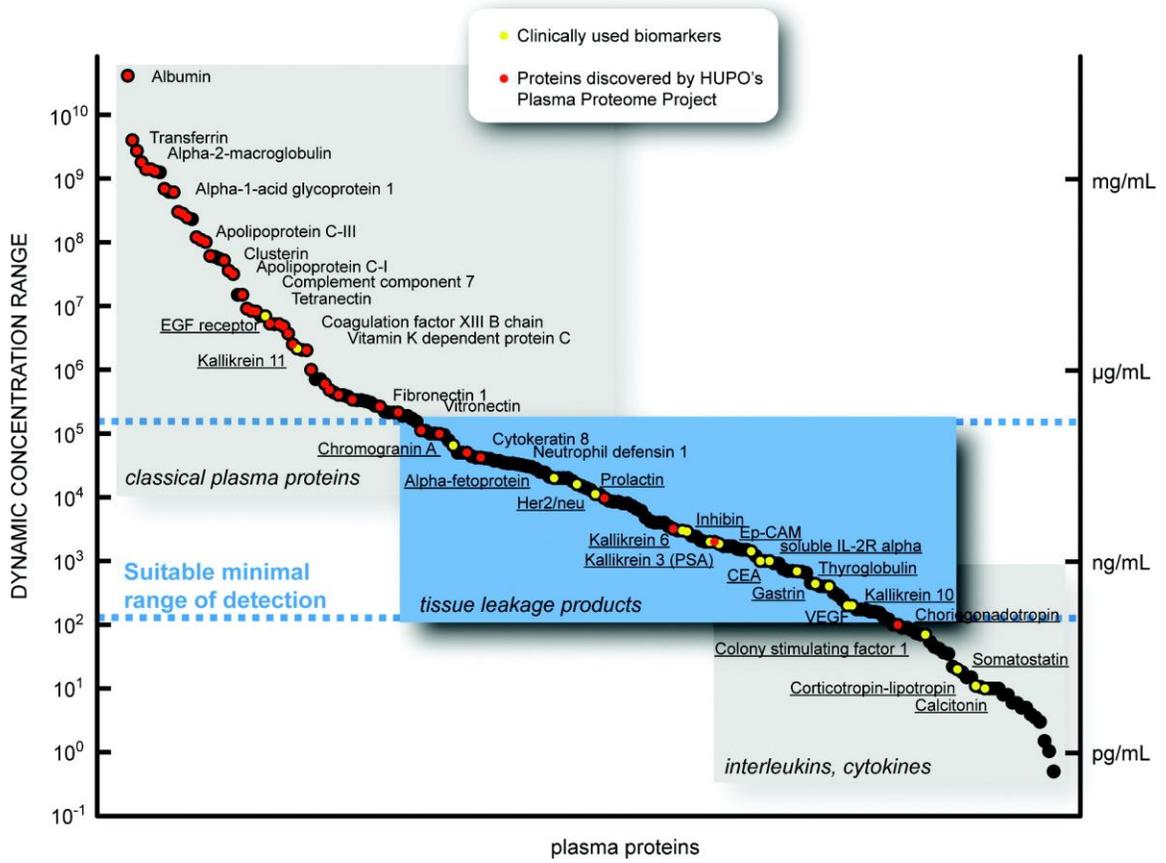
Cancer survival rates from adenocarcinomas with early and late detection[11]

Figure 2



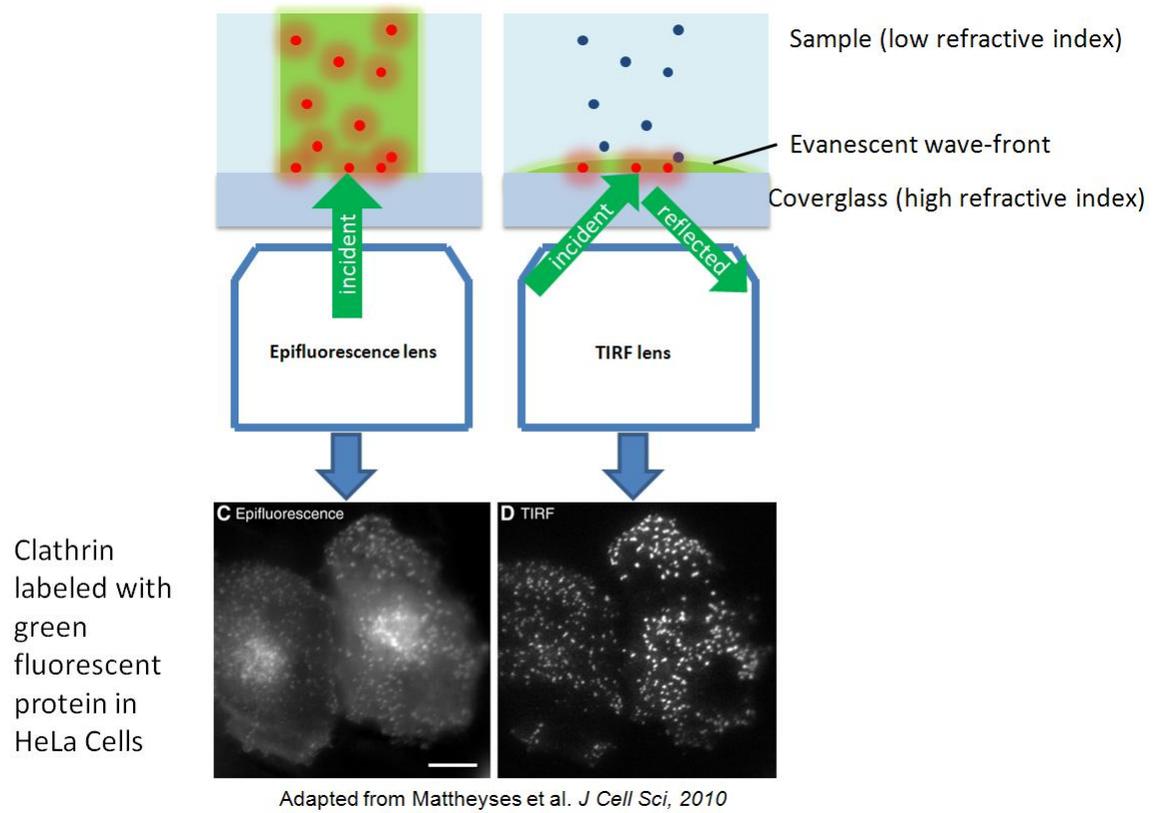
The relative contribution of proteins within plasma[16]

Figure 3



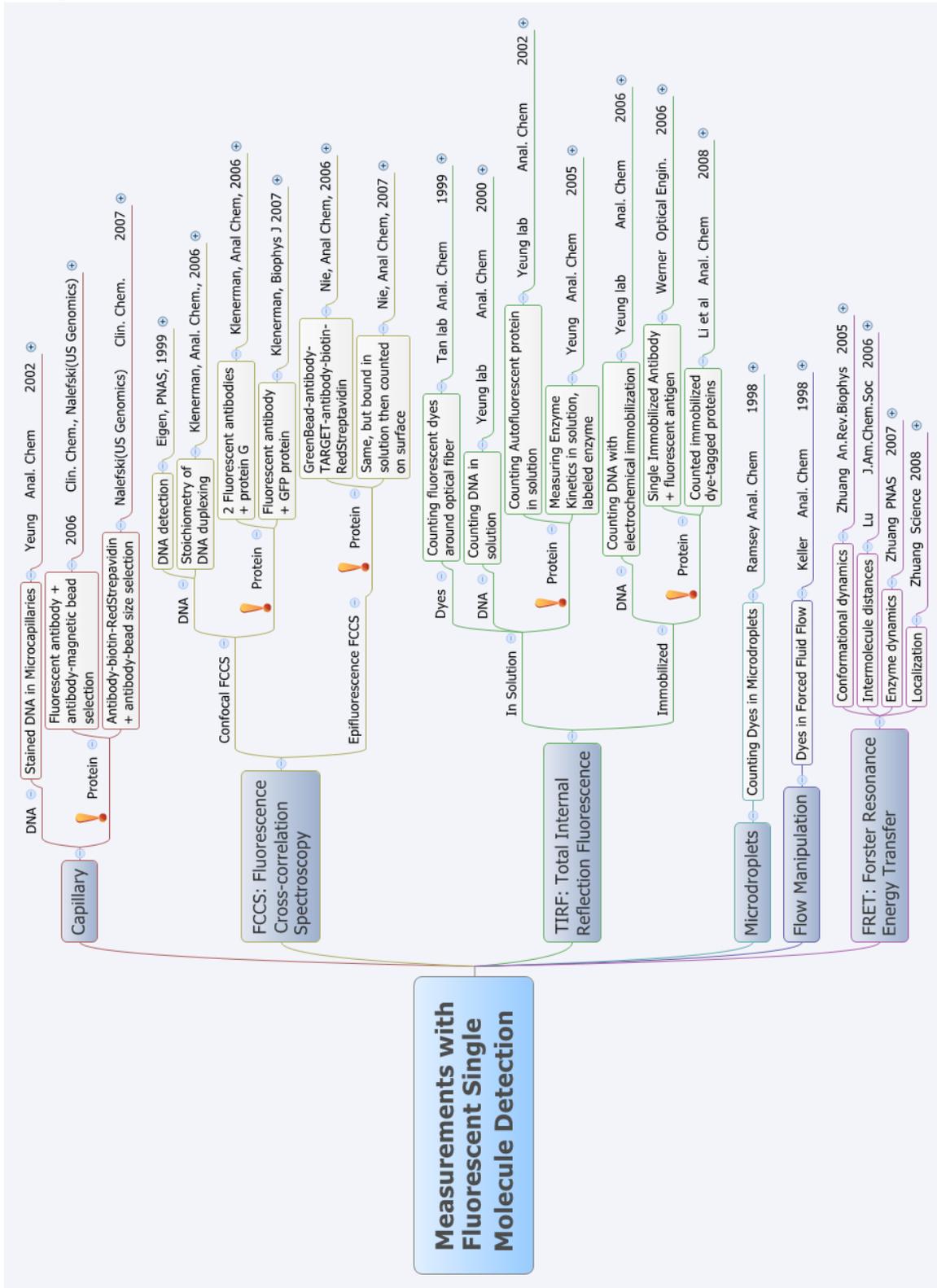
Dynamic range of plasma proteins spans ten orders of magnitude.[17]

Figure 4



Total internal reflection fluorescence (TIRF) microscopy[44]

Figure 5



A deep literature search of the single-molecule fluorescence literature in 2009

CHAPTER TWO: PROTEIN QUANTIFICATION IN COMPLEX MIXTURES BY SOLID PHASE SINGLE- MOLECULE COUNTING

This chapter embodies a manuscript that was published in the journal *Analytical Chemistry* on July 14, 2009. This chapter describes a novel method that Robi D. Mitra and I developed for quantifying proteins on a solid surface using an immunoassay that has single-molecule sensitivity. In this chapter, I overcome the obstacles that prevented the use of single-molecule imaging in an immunoassay and demonstrate the utility of the method in research and clinical applications. This study was conducted under a sponsored research agreement with Helicos Biosciences, Inc. Jeffrey Reifenger trained me in Cambridge, MA to use and set up a total internal reflection fluorescence (TIRF) microscope. The experiments in this study were designed by Robi D. Mitra and me, and the experiments, software development, and TIRF setup were performed by me.

ABSTRACT

Here we present a procedure for quantifying single protein molecules affixed to a surface by counting bound antibodies. We systematically investigate many of the parameters that have prevented the robust single-molecule detection of surface-immobilized proteins. We find that a chemically adsorbed bovine serum albumin surface facilitates the efficient detection of single target molecules with fluorescent antibodies, and we show that these antibodies bind for lengths of time sufficient for imaging billions of individual protein molecules. This surface displays a low level of nonspecific protein adsorption, so that bound antibodies can be directly counted without employing two-color coincidence detection. We accurately quantify protein abundance by counting bound antibody molecules and perform this robustly in real-world serum samples. The number of antibody molecules we quantify relates linearly to the number of immobilized protein molecules ($R^2 = 0.98$), and our precision (1-5% CV) facilitates the reliable detection of small changes in abundance (7%). Thus, our procedure allows for single, surface-immobilized protein molecules to be detected with high sensitivity and accurately quantified by counting bound antibody molecules. Promisingly, we can probe flow cells multiple times with antibodies, suggesting that in the future it should be possible to perform multiplexed single-molecule immunoassays.

INTRODUCTION

Our ability to detect and quantify proteins has lagged behind our ability to analyze nucleic acids. Closing this gap by developing more sensitive and quantitative protein analysis methods would greatly aid efforts to understand cellular processes[45, 46] and to

search for protein biomarkers that reveal disease state.[15, 47] The application of single-molecule detection (SMD) methods to proteins holds great promise in this regard for five reasons: 1) Recent advances have made SMD methods inexpensive, robust, and reliable.[48, 49] 2) SMD methods can enable the detection of low-abundance proteins[50-52], which is especially important because the poor sensitivities of current proteomic methods are limiting progress in the area of biomarker discovery.[53, 54] 3) SMD methods can enable protein quantification by employing single-molecule counting, which can be significantly more accurate than bulk methods.[41, 42] 4) SMD methods can enable analysis of protein-protein interactions by detecting single-molecule co-localization.[55] 5) SMD methods for proteins affixed to a surface could enable highly multiplexed immunoassays. For example, by creating ~ 20 overlapping pools of labeled antibodies using a logarithmic pooling strategy like the one used to decode bead-based random microarrays[56], a single assay could detect the protein targets of all 6,000 non-redundant human proteome antibodies[57] with only ~ 20 binding rounds.

There are several obstacles that have hampered the development of single-molecule immunoassays. One is the lack of a good surface for the SMD of surface-immobilized proteins. An ideal surface would be resistant to nonspecific antibody adsorption, while still allowing for the specific binding of antibodies to their target molecules. Efforts have been made to develop better surfaces[43, 58-68], however, the nonspecific adsorption on these surfaces has not been characterized with single-molecule resolution, with a few exceptions.[59, 60] To work around high background surfaces, researchers have introduced innovative detection schemes, often relying on two-color coincidence detection.[69, 70] This helps to reduce detection noise but is suboptimal for a

proteomic method, since 1) pairs of protein-specific capture reagents with non-overlapping epitopes may not be obtainable for all proteins, 2) generating pairs of reagents increases cost, and 3) determining the optimal binding conditions for dual antibody sandwich immunoassays can require more optimization than for single-capture antibody immunoassays.[71] A second obstacle to single-molecule immunoassays is the dissociation of antibodies from their individual targets during imaging. Many antibodies rapidly dissociate from their ligands in solution, but surface dissociation is often slower. It is not known whether the surface dissociation rates of antibodies will enable the sensitive detection of single ligand molecules. Finally, a single-molecule immunoassay must be able to sample large number of molecules in each experiment to ensure accurate protein quantification and to maximize the dynamic range.

Here we demonstrate a method for quantifying protein molecules on a surface by counting bound antibodies. To achieve this, we first optimized an image acquisition and processing method for SMD of fluorescently labeled antibodies on the surface of a flow cell. Then we systematically evaluated 12 surface chemistries for single-protein detection. For each surface, we quantified the nonspecific adsorption of single antibody molecules and characterized the efficiency of target protein immobilization. We found that a chemically adsorbed bovine serum albumin (BSA) surface had the lowest nonspecific binding and still allowed for efficient protein sample attachment. Using this surface incorporated into a flow cell, we measured the fraction of immobilized proteins that could be detected by direct antibody binding, and found that a high fraction of targets (at least 70%) were bound specifically. We directly measured the surface dissociation of antibodies from their ligands and found it to be highly suited for large scale single-

molecule quantification. We further showed that proteins were accessible to antibody binding over multiple binding rounds. Finally, we were able to quantify immobilized proteins by directly counting bound antibody molecules. A sufficiently low level of background binding was observed such that single target molecules could be detected without employing two-color coincidence detection. We found this method was both accurate and sensitive – the number of antibody molecules counted was linearly related to the number of proteins ($R^2 = 0.98$), and as few as 55 ligand molecules per 1,000 μm^2 image (1.4 pg cm^{-2}) could be detected over background. Our detection method showed robustness in the background of a complex biological fluid, and we demonstrated the accurate quantification of an endogenous protein within blood serum samples. Thus, we have resolved many of the issues that have limited the feasibility of solid phase single-molecule protein analysis and have demonstrated reliable protein quantification in biological samples by single molecule counting on a solid surface.

RESULTS AND DISCUSSION

There are a number of formats and methods by which the SMD of biomolecules can be achieved.[30] We chose total internal reflection fluorescence (TIRF) microscopy as the basis for our single-molecule immunoassays. We attach a small amount of protein sample to the surface of a flow cell, probe with fluorescent antibodies, remove unbound antibodies and directly image the bound antibodies (Figure 6A).

Iterative Thresholding Improves Detection of Labeled Antibody Molecules

We first sought to verify that our detection system could achieve single-molecule resolution of fluorescent antibody molecules affixed to glass. We mixed Cy3-labeled antibodies with identical antibodies labeled with Cy5, diluted the mixture, and reacted it to an epoxide-coated glass slide. We performed TIRF imaging with Cy3 and Cy5 channels, determined the positions of the Cy3 and Cy5 antibodies using software, and overlaid the positions (Figure S1). The Cy3-labeled molecules did not co-localize with Cy5-labeled molecules more than would be expected by chance ($p = 0.78$, Fisher's Exact Test), demonstrating that the fluorescence objects detected were not clusters of antibodies (which would have been detected as co-localized molecules) but single antibody molecules.

In these initial experiments, we observed considerably fewer fluorescent antibodies at the edges of our field of view relative to the center of the image. This is due to the non-uniform laser illumination intrinsic to our Nikon optical design (Figure 6B). Since this non-uniform illumination greatly reduces the number of antibody molecules that can be analyzed in a single field of view using standard, single value thresholding, we developed an automated and unbiased image processing technique, "iterative thresholding" (see Supporting Methods). The algorithm uses local thresholds to compensate for the lower intensities at the edges and is able to accurately identify the locations of fluorescent antibodies independent of their position within the field of view. We tested the performance of this technique by comparing images processed with our iterative thresholding algorithm with the same images analyzed by single value thresholding. Our iterative thresholding algorithm (Figure 6D) identified the positions of

14-fold more antibodies per field of view ($1,408\% \pm 420\%$), on average, than the standard method (Figure 6C) while introducing few false positives with respect to the raw data (sensitivity = $99.96\% \pm 0.07\%$, specificity = $98.47\% \pm 0.76\%$). Thus iterative thresholding substantially increased the efficiency of fluorescent antibody analysis using objective TIRF and provided a foundation for our protein quantification method.

Study of Surface Nonspecific Adsorption

Minimizing the nonspecific adsorption of antibodies to surfaces is critical for the development of single-molecule immunoassays because it causes false positive events, decreasing the accuracy and sensitivity of the assays. To find the best surface for single-molecule immunoassays, we systematically searched the literature to identify surfaces that were shown to have minimal interactions with antibodies. We chose surface chemistries previously used for SMD[60] and for biosensors[58, 62, 63, 66], as well as several that we speculated would exhibit low levels of non-specific protein adsorption. Some protocols were followed directly from the literature while others, such as the chemically adsorbed BSA protocols, were modified (see Materials and Methods).

We quantified the nonspecific adsorption of antibodies for 12 different surface chemistries. We loaded a glass slide into a flow cell, treated it according to a particular surface protocol, exposed it to Cy5-labeled antibody, washed away unbound antibodies, and quantified the number of adsorbed antibody molecules by single-molecule counting (Figure 6E). Since no ligand was present on the surface of the slide, each surface-bound antibody represented a nonspecific adsorption event. We observed 122 to 2,600 antibodies per $1,000 \mu\text{m}^2$, (3 to 68 pg cm^{-2}). We found that a chemically adsorbed BSA

surface (first developed by Heyes et al[60] and modified here to allow adsorption to the glass via epoxide crosslinking and capping) showed the least amount of nonspecific binding. Dextran, aminodextran, and the smaller molecular weight linear polyacrylamide (LPA) surfaces showed moderate adsorption. Among these three protocols, polymers of lower molecular weight (1,500 and 5,000 g/mol) performed better than those of higher molecular weight (10,000 and 500,000g/mol). Large LPA, CM-dextran, glucose, IgG, amino-PEG and PEG performed the worst.

These findings were consistent with those reported by Heyes et al[60], who found that chemical immobilization of BSA onto a glass surface provided great reduction in nonspecific adsorption of streptavidin molecules. By atomic force microscopy, they showed that this surface was highly homogeneous, supporting the hypothesis that the 75 kDa BSA protein creates a neutral, hydrophilic layer that sterically hinders proteins from nonspecifically adsorbing to the sticky silicon dioxide below. Based on the performance of our adapted BSA surface, we selected the chemically adsorbed BSA surface for further characterization.

Robust Immobilization of Protein Ligands on BSA-Coated Glass

It is important that a single-molecule immunoassay surface allows for the robust anchoring of ligand molecules. However, it was not clear whether the low background BSA surface discussed above could provide enough functional groups for the attachment of a protein sample. Therefore, we tested how efficiently proteins would anchor to the BSA surface using the heterobifunctional crosslinking system 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) and N-hydroxysuccinimide

(NHS). We prepared a BSA surface in a flow cell and activated the free carboxyl groups on the BSA molecules with EDC and NHS. We washed the flow cell to remove unbound crosslinker and then exposed the flow cell to Cy3-labeled protein to immobilize the proteins via their primary amines. The flow cell was washed again to remove unbound protein molecules, un-reacted crosslinking sites were quenched, and the flow cell was imaged.

Crosslinking proteins to the BSA surface allowed for a ten-fold increase in the number of protein molecules affixed to the surface compared to the surface without EDC/NHS activation ($950\% \pm 52\%$). Also, the proteins were able to be attached at over 1,000 molecules per field of view – a density that allows for high-throughput single-molecule sampling (Figure S2). Thus, we concluded that the EDC/NHS system was able to effectively activate the BSA surface and attach a protein sample. The chemically adsorbed BSA surface with EDC/NHS sample immobilization provided the surface chemistry for all subsequent experiments (Figure 6A).

We enable protein sample attachment by generating peptide bonds between the solvent-accessible carboxyl groups of the BSA and the primary amine groups of the target proteins. This contrasts the approach of some single-molecule studies which have relied on biotin streptavidin linkage.[60, 72] Our method does not rely on pre-labeling samples by biotinylation, instead taking advantage of endogenous lysine residues present on most proteins. Therefore our approach may provide a more universal way of attaching heterogeneous biological samples.

Efficient Detection of Single Protein Molecules by Antibody Binding

The accuracy of a single-molecule immunoassay depends on the accessibility of target molecules to antibodies; inaccessible ligands will not be detected or counted. There are several mechanisms that can prevent an antibody from binding a ligand immobilized on a solid substrate. Steric, electrodynamic, and thermodynamic variables can hinder binding when repulsive forces of the surface overcome the attractive forces of the antibody-protein complex. Kinetics can also hinder binding if a free energy barrier is sufficiently high to prevent docking on relevant time scales.[73] We sought to determine to what degree these variables affect the accessibility of target molecules to antibodies in our system.

To analyze the binding of target molecules by antibodies, we performed a dual-color, single-molecule protein accessibility assay (Figure 7). Here, the target proteins were labeled with Cy3 and the antibodies were labeled with Cy5. We prepared a BSA surface within a flow cell, immobilized the target proteins, capped the reactive crosslinking sites, and acquired a pre-antibody image. Then, we probed with antibodies, washed away unbound antibodies, and imaged. We compared the positions of the antibodies with the positions of the proteins imaged beforehand by overlaying their locations. To verify that the co-localization of proteins and antibodies was a result of specific binding, we measured the correlation between protein and antibody positions and tested that correlation for randomness (see Supporting Methods). The correlogram in Figure 7 indicates that antibody binding was specific and not due to chance correlation. (To confirm the specificity of binding we also performed the protein accessibility assay

using a nonspecific target protein with which the antibodies should have had no affinity and observed a correlogram showing no significant correlations (Figure S3).)

To quantify ligand accessibility, we measured the fraction of proteins that were co-localized with antibodies. We then performed this protein accessibility assay for different antibody concentrations. The total fraction of proteins bound by antibodies is shown by the dashed line in Figure 8. To better determine the amount of specific binding, we estimated nonspecific binding based on the observed antibody density and subtracted that from the total binding (see Supporting Methods). The dotted line shows the estimated fraction of proteins that overlapped with antibodies as a result of nonspecific binding, while the solid line shows the fraction of specifically bound ligand molecules.

The accessibility curve follows the behavior of fractional occupancy that is expected from binding theory. When 1 $\mu\text{g/ml}$ antibody is used, $\sim 70\%$ of the target molecules were specifically bound by antibodies. From these results, we conclude that single protein molecules can be efficiently detected by counting bound antibody molecules.

Surface Dissociation of Antibodies

We sought next to determine if the ligand molecules that we failed to detect in the protein accessibility experiments described above were not detected because they were never bound by antibodies, or if they were initially bound by antibodies but the complexes dissociated before imaging. Antibody-ligand interactions are known to have dissociation half-lives in solution ranging from minutes to several hours. However, the surface dissociation rate may be slower due to surface-antibody interactions that stabilize

the complex. Therefore, we designed an experiment to measure the surface dissociation rate of antibodies bound to single ligand molecules.

To measure the surface dissociation rate of the antibodies, we allowed antibodies to bind to target proteins that were immobilized on the surface of the flow cell, as previously described. We imaged the surface to determine the starting number of antibody:ligand complexes and then began a continual wash to remove unbound antibodies from the flow cell. We imaged the surface every 8 hours over a 48 hour period. At each time-point we quantified the number of antibody:ligand complexes that were lost relative to the starting time point, and from this we measured the surface dissociation of the antibodies.

Nearly all (> 90%) of the co-localized pairs of proteins and antibodies remained intact for 48 hours at room temperature (Figure S4). Furthermore, antibody dissociation did not follow exponential decay over this time period. Together, these results suggest a strong antibody-surface interaction. The high stability of bound antibodies also explains how we were able to detect single ligand molecules with high efficiency (i.e. Figure 8) even though we thoroughly washed the flow cell.

The half life of a typical antibody:ligand complex can be as short as several minutes in solution. Such rapid dissociation would pose a serious barrier to the development of a solid phase, single-molecule immunoassay because antibodies would be washed off of the surface of the flow cell before they could be detected. Fortunately, surface interactions appear to stabilize antibody:ligand interactions. Using the observed surface dissociation rate we calculated the dynamic range that can theoretically be achieved. If ligand molecules are immobilized at a density of 1,000 target molecules

per image and 10 images are acquired per second (a rate possible with the current generation of charge-coupled device cameras), then one can acquire images of $1,000 \times 0.9 \times 10 \times 60 \times 60 \times 48 = 1.5$ billion target molecules while retaining 90% of the antibodies on the surface. Thus, the observed surface dissociation rate will support a dynamic range of 9 orders of magnitude. This suggests that it should be possible to develop single-molecule immunoassays with a high dynamic range.

Dual-round protein binding

Ligand rebinding in successive binding rounds could be used to increase detection specificity or to enable efficient sample multiplexing.[56] However, as our surface dissociation experiments illustrated, it was difficult to remove bound antibodies from the surface. This was true even after washing with various antibody eluting reagents (data not shown). Therefore, we wanted to explore rebinding ligands by “erasing” antibodies from the surface via photobleaching. We believed rebinding after photobleaching might be possible because the antibodies we used was polyclonal and could theoretically bind multiple epitopes on a single ligand.

One hurdle to performing multiple binding rounds with an intermediate photobleaching step is that the antibodies that bind in the first round could competitively inhibit the binding of antibodies in subsequent rounds. To test whether competitive binding would be a major phenomenon, we probed Cy3-labeled ligand molecules with Cy5-labeled antibodies as described above and acquired the positions of the bound antibodies. We then photobleached the antibodies with 640nm light before performing a second round of binding with the same antibody. (Target molecules were not bleached.)

If antibodies competitively inhibited the second round of binding, then we should not have observed any ligand molecules that were bound in both rounds.

We observed 2,829 Cy3-labeled ligand molecules. Of these, 1,497 proteins were bound in round 1, 1,146 were bound in round 2, and 526 (18.6%) were bound twice. Assuming independent binding in round 1 and round 2, we would expect 21% of the ligands to be bound twice. Thus, approximately 87% of the proteins bound in round 1 were available for binding in round 2 (Figure 9). This result supports the feasibility of performing multiple rounds of single-molecule protein detection.

Quantification of Single Molecules by Antibody Binding

We next set out to perform a quantitative immunoassay, counting single, immobilized protein molecules by detecting bound antibodies. We affixed varying amounts of Cy3-labeled protein onto the surface and quantified the number of immobilized target molecules by imaging. Then we probed the surface with Cy5-labeled antibodies and counted the total number of bound antibodies after washing.

The solid line in Figure 10 illustrates the relationship between number of antibody molecules and number of protein molecules affixed to the surface. In the range of 55 to 1,676 target molecules per 1,000 μm^2 image, we observed a linear relationship between the number target molecules and antibodies. The lower limit of detection (LOD) of 55 molecules per 1,000 μm^2 image (1.4 pg cm^{-2}), was achieved by acquiring only 5 images. It should be possible to detect lower quantities of surface-bound proteins by acquiring greater numbers of images.[51] Given our sample immobilization efficiency and this LOD, we were able to detect proteins in solution down to 100 pM. In this proof-of-

principle study, we did not attempt to maximize the attachment efficiency, but doing so should increase the detection sensitivity.[51]

The standard curve displays high correlation ($R^2 = 0.98$), and we obtain precision between 1% and 5% CV. By acquiring only 5 images we can robustly detect abundance changes down to 7% – we generated 99% confidence intervals around each data point, and the widest interval was a 7% deviation. This result demonstrates the utility of digital quantification.

We also quantified the Cy3-labeled target protein in the presence of serum. Here, we spiked Cy3-labeled target protein at varying concentrations into neat rabbit serum. The complex mixture, including target and non-target proteins, was immobilized to the BSA surface. We probed with fluorescently-labeled antibody and quantified the number of target proteins versus the number of antibodies on the surface.

Similar results to the purified protein detection curve were obtained, demonstrating the robustness of the method in the presence of a complex biological fluid (Figure 10, dashed line). The LOD in serum was 390 molecules per $1,000 \mu\text{m}^2$ (10 pg cm^{-2}) corresponding to a target starting concentration of $1 \mu\text{g/ml}$. By comparison, the total concentration of the serum was 74 mg/ml (by dry weight). Therefore, despite the overabundance of serum proteins, the serum introduced almost no background. This indicates that single-antibody, direct binding can be used to make specific detection measurements in a highly complex biological fluid.

Accurate Quantification of an Endogenous Serum Protein

We next applied our method to quantify endogenous protein in a biological sample. We quantified the amount of total IgG in blood of a rabbit at various time points after immunization. Serum samples were diluted in PBS, immobilized to flow cell surfaces, and probed with anti-rabbit IgG Cy5-antibody. Then we quantified the antibodies remaining on the flow cell surface after washing.

We detected a 70.0% increase ($\pm 8.1\%$) in total IgG between pre-immunization and week 4, as well as a subtle 11.7% increase ($\pm 4.4\%$) between weeks 4 and 5 (Figure 11). Our single-molecule quantitation measurements matched bulk measurements obtained by ELISA, deviating from the gold-standard by at most 4.2% (see Supporting Methods). This demonstrates the accuracy of single-molecule quantitation in complex, real-world samples.

CONCLUSION

Recent advances in SMD have the potential to usher in a new generation of proteomics tools. Toward the goal of uniting the field of protein detection with single-molecule counting, we present a proof-of-principle in which we quantify the abundance of individual proteins on a solid surface by counting bound antibodies. Further, we demonstrate quantitation of an endogenous protein in real-world serum samples while eliminating the need for two-color coincidence detection. We optimized key parameters – image acquisition and processing, nonspecific antibody adsorption, sample immobilization, sample accessibility, and surface dissociation – in a systematic way to enable a quantitative immunoassay. Because these parameters are interconnected, we

found that it was important to optimize them simultaneously, which allowed us to quantify small changes (7%) in abundance of target proteins by single-molecule counting.

An important future goal is to perform multiple rounds of antibody binding on a solid surface to allow for efficient multiplexing.[56] This would be achieved by encoding each binding pool with a predetermined mixture of antibodies, so that n protein targets could be quantified in $\sim \log_2 n$ binding rounds (Figure S5). The ability to perform multiple rounds of binding would also enable error-checking, since antibodies would get a second pass at detecting a particular target. Additionally, analysis of protein-protein interactions would follow easily from such an approach, since interacting proteins will be present at the same positions on the flow cell.

Towards this goal, we have demonstrated the serial detection of proteins by two rounds of antibody binding. We used a photobleaching step after the first round of binding to erase surface-associated fluorescence prior to the second hybridization. We used photobleaching because the rate at which specifically bound antibodies dissociated from the surface was low enough that we found it difficult to completely remove them from the flow cell in a reasonable amount of time. Using this approach, we found that the majority (87%), of proteins that we expected to be bound in two binding rounds were in fact bound twice, indicating that competitive binding by the bleached, surface-bound antibodies was minimal. This lends support to the feasibility of multiple rounds of antibody binding and detection, with each round separated by a photobleaching step. (Alternatively, one could use a cleavable linker between the antibody and fluorophore, which would enable dye removal by exposure to a reducing agent or to light.[74])

Washing with surfactants and denaturants may allow us to better remove bound antibodies from their targets. For example, recent work demonstrated the efficient stripping of antibodies from western blots without disrupting protein attachment.[75] To develop such a protocol in a single-molecule setting will require a low-background surface that is also surfactant-compatible (the surfaces described here are not). We are now investigating low-background surfactant-compatible surfaces that utilize multi-arm PEG microgels.[76]

Some obstacles still must be overcome before it is feasible to develop a multiplexed single-molecule immunoassay. For example, since each antibody-ligand pair has variable affinities, they each may need to be characterized beforehand in order to ensure that the concentration of antibody used in the immunoassay is high enough to ensure maximal binding to its immobilized ligand. However, as antibody production and characterization becomes more standardized, it will become possible to obtain large numbers of well-characterized antibodies. For example, the Human Antibody Initiative has already generated and curated antibodies against over 6,000 human proteins, and they aim to expand the collection to the entire human proteome within the decade.[57] Solid phase single-molecule immunoassays could provide a way to leverage such antibody collections toward high-throughput proteomic applications.

MATERIALS AND METHODS

Imaging

All experiments were performed on a Nikon TE-2000 inverted microscope fitted with a total internal reflection fluorescence (TIRF) illuminator (Nikon, Melville, NY).

Two lasers, 532 nm/75 mW and 640 nm/40 mW were used for fluorescence excitation (Compass 215M, Cube-40C, Coherent, Santa Clara, CA). Illumination of the sample was controlled through a computer animated shutter (Prior Scientific, Rockland, MA). The 532 nm laser beam was attenuated by a ND 2 neutral density filter (Nikon, Melville, NY). The two beams were coupled into one end of an optical fiber cable using a dichroic mirror (Z532BCM, Chroma, Brattleboro, VT), with the other end of the cable attached to the TIRF illuminator. Before reaching the objective, each laser beam passed through a band pass filter: HQ545/30 for the green laser and D635/30 for the red laser (Chroma, Brattleboro, VT). Objective type total internal reflection was achieved through a 60x TIRF oil objective with index of refraction 1.49 (Nikon, Melville, NY). The chemistry of the assay was performed in a flow cell (see Fluidics) mounted onto the microscope stage. When the lasers are experiencing TIR, an evanescent wave decays exponentially at the glass-water interface into the flow cell to a distance of about 300 nm. TIRF allows for the excitation of only surface-bound fluorophore-labeled antibodies and therefore reduces the overall fluorescence background. The emitted photons from the labeled antibodies were collected by the objective and passed through a dichroic mirror (custom Cy3/Cy5, Semrock, Rochester, NY) and an emission filter for either the green channel (HQ610/75, Chroma, Brattleboro, VT) or the red channel (LP02-647RU-25, Semrock, Rochester, NY). Light was then detected by a charge coupled device (CoolSnap ED, Roper Scientific, Tucson, AZ) which imaged a 140 μm by 100 μm (1,400 px x 1,000 px) region of the surface.

Immediately prior to image acquisition, the flow cell was washed with 600 μl PBS and loaded with 600 μl oxygen scavenger and blink-reduction system[77] to prevent

dyes from photobleaching and blinking. Then images were acquired in the red and green fluorescence channels at five different positions across length of the flow cell, with 0.5 second exposure. Custom software written in Metamorph (Molecular Devices, Sunnyvale, CA) and Matlab (Mathworks, Natick, MA) was used to analyze the locations and intensities of the fluorescent molecules (see Supporting Methods).

Fluidics

The analysis substrate was a 40mm diameter, #1.5 glass slide (Erie Scientific, Waltham, MA). The substrate was epoxide-derivatized by the vendor unless otherwise specified in Surface Preparations. The slide was loaded into a flow cell (FSC2, Bioptechs, Butler, PA) fitted with perfusion ports to allow for reagents to be passed over the surface. Reagents were flowed through by a custom-made negative pressure vacuum pump.

Target Proteins

The target proteins were polyclonal goat IgG molecules labeled with an average of 8 Cy3 dyes per molecule. The nonspecific target proteins used as a negative control in the target binding accessibility assay were polyclonal rabbit IgG molecules labeled with Cy3. Proteins were obtained from Abcam (Cambridge, MA).

Serum Samples

The serum sample used for the spike-in quantification experiment was obtained from rabbit. The serum samples used for the endogenous protein quantification experiment were from pre-immunized, week 4, and week 5 rabbits in an antibody

production protocol (for an unrelated study) during which rabbits were immunized with antigen and adjuvant. All serum samples were obtained from 21st Century Biochemicals (Marlboro, MA).

Antibodies

The antibodies used in all experiments with the exception of the endogenous protein quantification experiment were polyclonal anti-goat, labeled with Cy5. The antibodies used to detect endogenous rabbit IgG were polyclonal anti-rabbit, labeled with Cy5. All antibodies were obtained from Abcam (Cambridge, MA).

Preparation of Surfaces

Twelve surfaces were generated by protocols taken directly from or adapted from surface blocking protocols in the literature.[43, 59-68] Nine of the surface chemistries were generated by the chemical attachment of primary amine groups of a polymer or small molecule to epoxide-derivatized glass. The epoxide-coated glass was loaded into the flow cell and washed in 600 μ L phosphate buffered saline pH 7.3 (PBS). The glass was reacted with one of the following solutions in PBS, for 1 hr at room temperature: 1% bovine serum albumin (BSA) (Fisher Scientific, Pittsburg, PA), 1% BSA / 0.1% cold water fish skin gelatin (Aurion, Netherlands), 1M glucose, 10% linear polyacrylamide (LPA) MW 1500 Da, 10% LPA MW 10 kDa, 10% LPA MW 1 MDa, 100 mg/ml amino-PEG (Sigma-Aldrich, St. Louis, MO), 200 μ g/ml rabbit IgG (Abcam, Cambridge, MA), and 100 mg/ml aminodextran MW 500 kDa (CarboMer, San Diego, CA). These surfaces were then capped with 1 M Tris pH 8.0 for 20 minutes. Two of the surfaces were

generated by the noncovalent adsorption of a polymer to the glass. Here the epoxide-coated glass was first capped with ethanolamine-HCl pH 8.0 for 20 minutes and then treated with one of the following solutions in PBS for 1 hr at room temperature: 100 mg/ml dextran and 1% PEG MW 5 kDa (Sigma-Aldrich, St. Louis, MO). The carboxymethyl (CM) dextran surface was generated as previously described.[62]

Measuring Nonspecific Adsorption

A flow cell containing the surface to be tested was loaded with 600 μ l, 100 ng/ml Cy5 antibody. The surface was exposed to the antibody in the dark for 25 minutes at room temperature. Then, unbound antibodies were removed with a 600 μ l PBS wash, and the flow cell was imaged as described above.

Immobilizing Protein Samples

A chemically adsorbed BSA surface was formed as described above, and the surface was activated with 0.2 M 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) and 0.05 M N-hydroxysuccinimide (NHS) (Pierce, Rockford, IL) in sodium phosphate buffer pH 5.8 (SPB) for 10 minutes. Free EDC and NHS was washed away with 600 μ l SPB.

The attachment of the protein sample of interest to the activated surface was as follows. For immobilization of purified target protein, 100 ng/ml (unless otherwise specified) of target protein in PBS was loaded into the flow cell. To generate a standard curve of detection, dilutions of target protein in PBS were loaded into the flow cell. To generate a standard curve of detection for target protein in the presence of serum,

dilutions of target protein were spiked-in to whole rabbit serum, and the spiked-in serum was loaded into the flow cell. To detect endogenous IgG in serum, whole rabbit serum was diluted 1:105 in PBS and loaded into the flow cell.

Proteins samples that were loaded into the flow cell were allowed to react with the surface for 10 minutes at room temperature, in the dark. Then, unbound proteins were removed with a 600 μ l PBS wash, and unreacted EDC-NHS sites on the BSA surface were quenched with 1M Tris pH 8.0 for 20 minutes.

Antibody Binding and Oxygen Scavenging

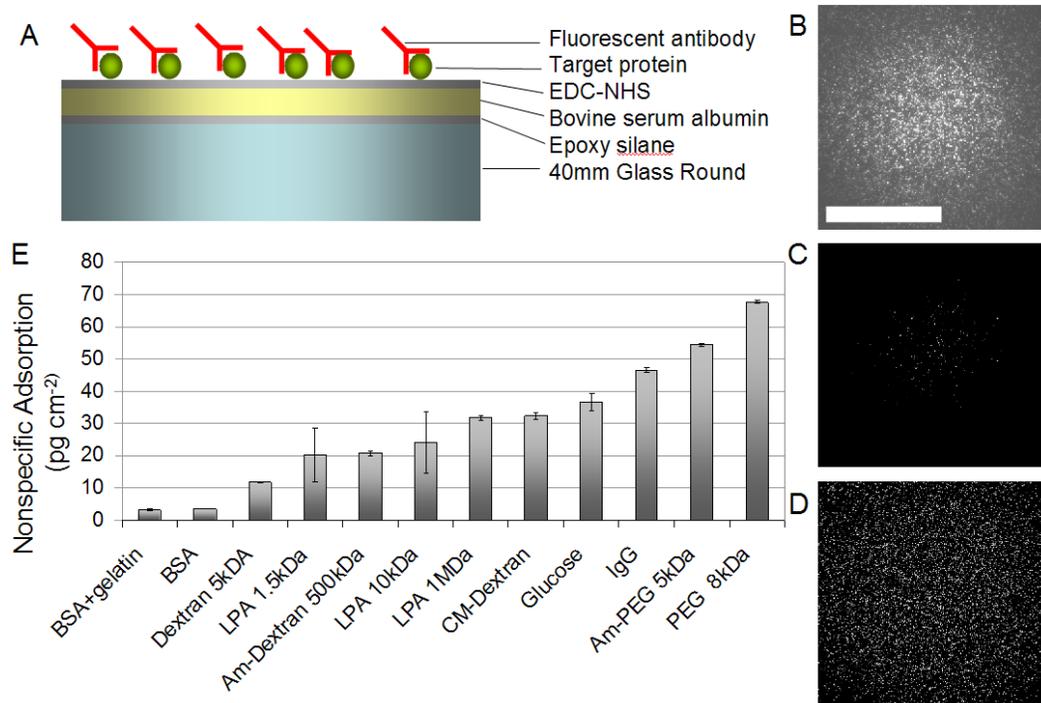
After a protein sample was immobilized onto the flow cell surface (as described above), the Cy5-labeled antibody was loaded into the flow cell at 100 ng/ml (unless otherwise noted) in PBS and incubated for 2 hrs in the dark at room temperature. Unbound antibodies were washed away, and the flow cell was imaged.

ACKNOWLEDGEMENTS

We thank Stan Lapidus and Patrice Milos for advice and helpful discussions. We thank Todd Druley, Jay Gertz, Katherine Varley, Haoyi Wang, and other members of the Mitra lab for helpful discussions and critical readings of the manuscript. This work was supported by Helicos Biosciences.

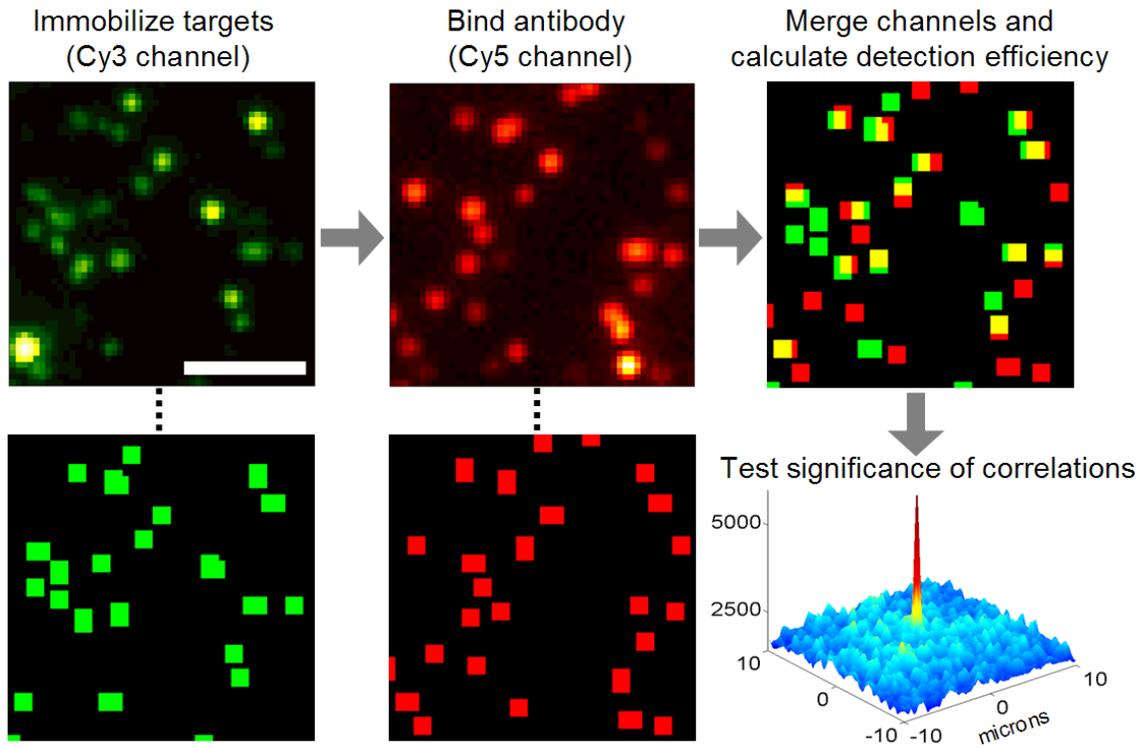
FIGURES

Figure 6



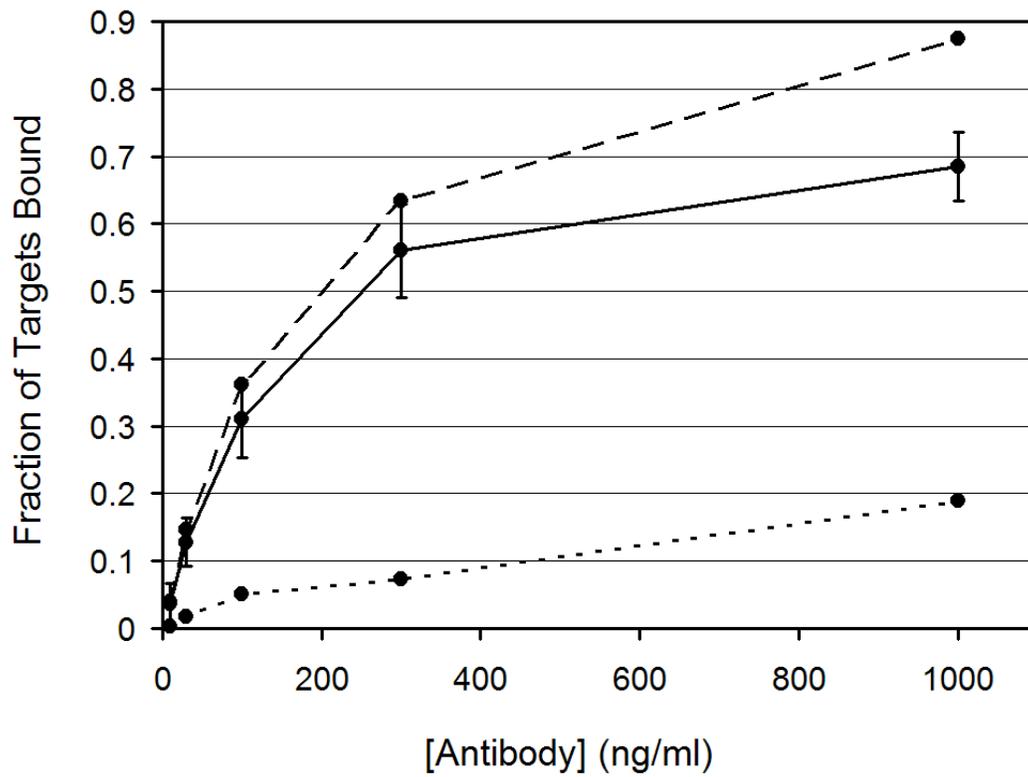
(A) Illustration of the single-molecule immunoassay. A chemically adsorbed BSA surface was prepared by reacting BSA with an epoxide-coated glass slide within a flow cell. Unreacted epoxides were quenched and the BSA was activated for sample immobilization by EDC/NHS. The protein sample (circles) was immobilized to the BSA surface, and unreacted sites were passivated. The flow cell was probed with fluorescently labeled antibody and imaged. (B) The raw TIRF image of Cy5-labeled antibodies (scale bar = 50 μm) illustrates the nonuniform TIRF illumination. (C) Image processing by standard, single value thresholding allowed only a small portion of the raw image (the brightest spots) to be used for molecule identification. (D) Image processing by iterative thresholding allowed for most of the raw image (regardless of intensity) to be used to molecule identification. (E) Nonspecific adsorption of antibodies onto twelve surface protocols. Molecules were counted in $5 \times 1,000 \mu\text{m}^2$ images, and units were converted to picograms per cm^2 assuming a 155kDa molecular weight. The chemically adsorbed BSA surfaces suppressed nonspecific adsorption the most.

Figure 7



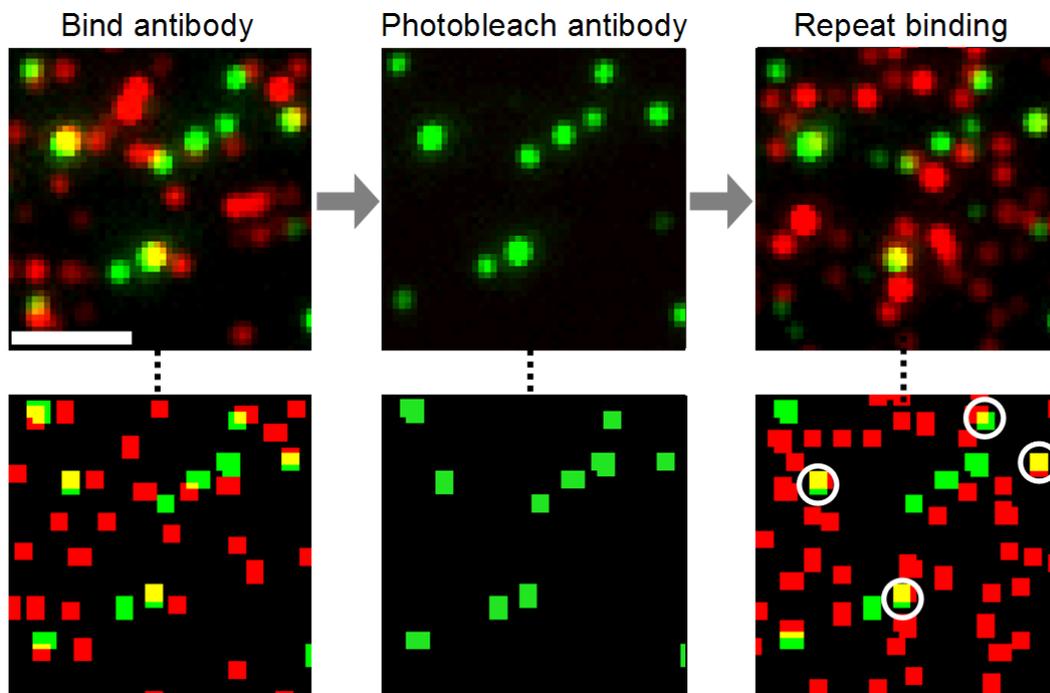
Determination of protein accessibility (detection efficiency). The image series illustrates target immobilization, antibody binding, and correlation detection. Each frame is an image of the same position in the flow cell (scale bar = 2 μm) and shows ~ 21 of the $\sim 10^3$ targets analyzed in each binding experiment. (i) After target protein immobilization, images of the Cy3-labeled proteins were acquired (top), and the positions of the proteins were determined by software (bottom). (ii) The surface was probed with antibody, images of bound Cy5-labeled antibodies were acquired (top), and positions of the antibodies were determined by software (bottom). (iii) Positions of the targets (green) and antibodies (red) were overlaid. Yellow pixels represent the co-localized molecules, indicating antibody-bound proteins. (iv) The correlogram analysis of this flow cell indicated that protein and antibody co-localization was nonrandom (i.e. antibodies were specifically binding to targets).

Figure 8



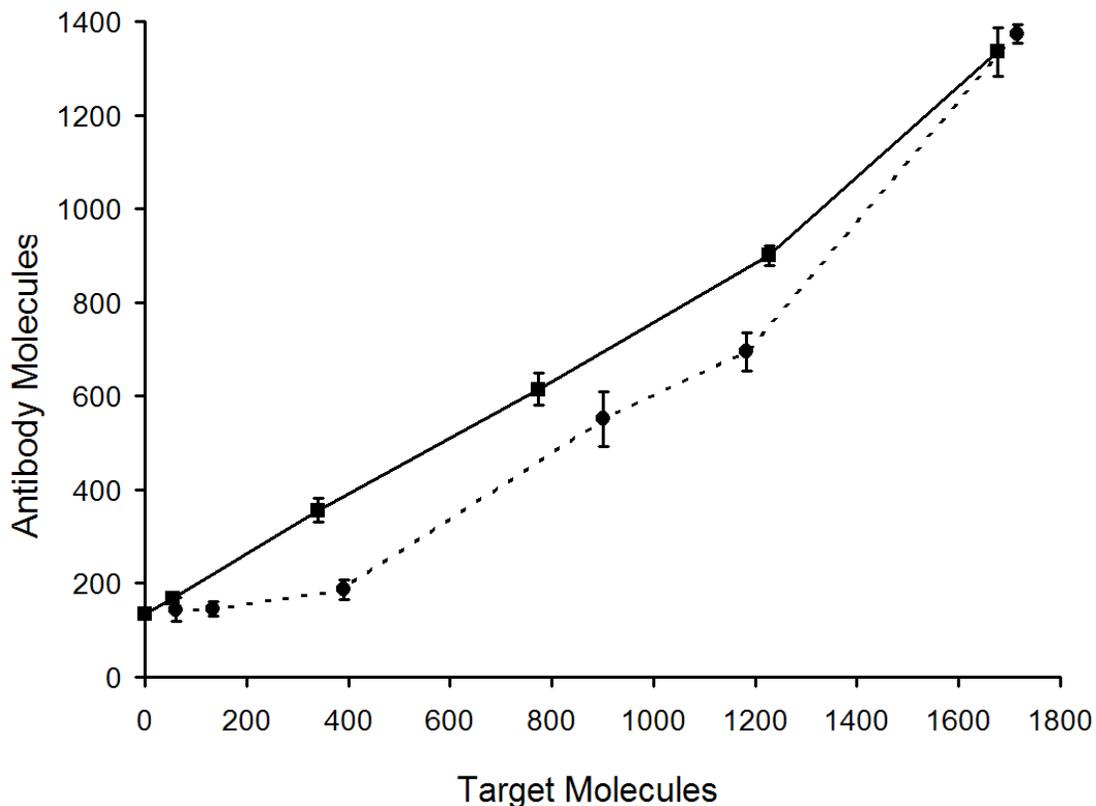
Protein accessibility (detection efficiency) as a function of antibody concentration. Specific binding (solid) was calculated by subtracting the nonspecific binding (dotted) from total binding (dashed). We could specifically bind as much as ~ 70% of the target molecules, enabling efficient protein detection.

Figure 9



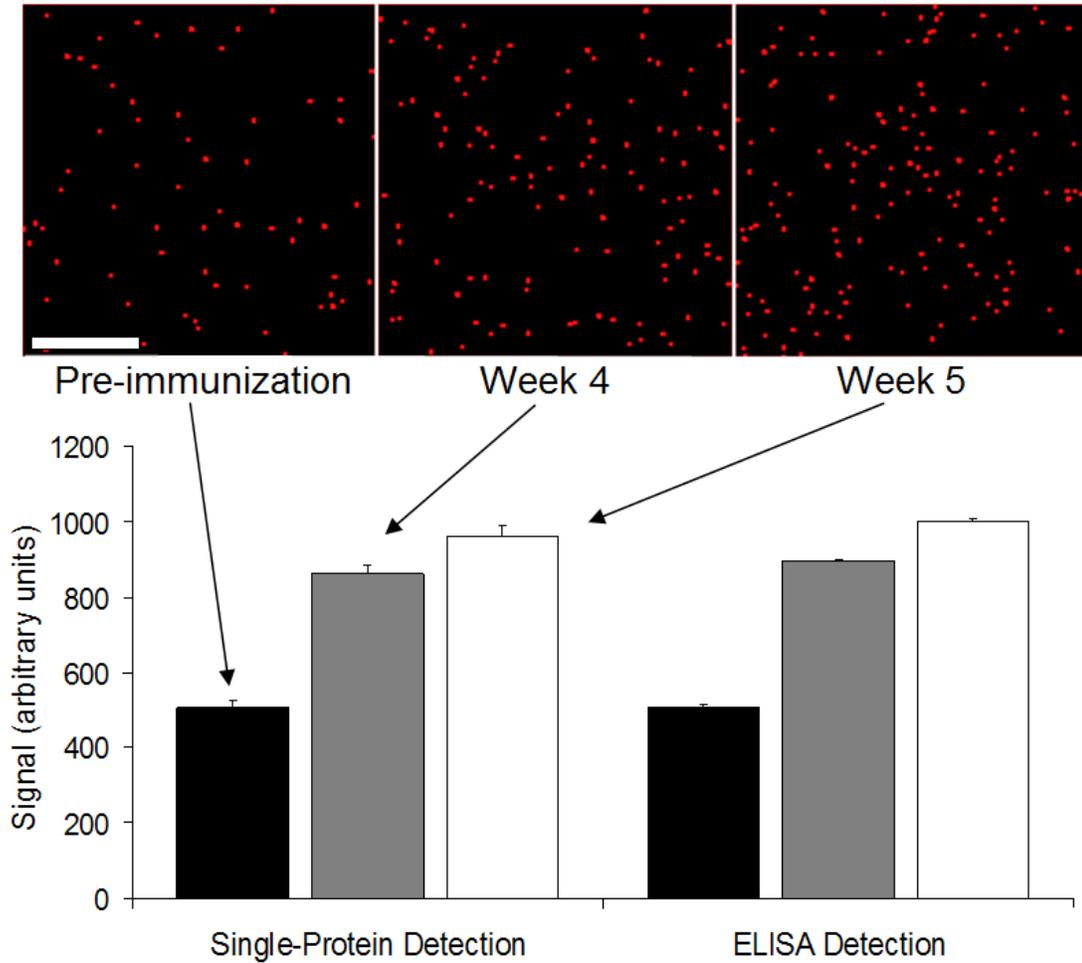
Protein rebinding. The image series illustrates two binding rounds separated by a photobleaching step (scale bar = 2 μm). Top: Raw images of Cy3 (green) and Cy5 (red) channels. Yellow represents merged channels. Bottom: Analyzed positions of protein molecules (green), antibody molecules (red) and co-localized antibody:protein molecules (yellow). Circles indicate proteins that were bound in both rounds. (i) First round of antibody binding detects immobilized proteins. (ii) Antibodies are photobleached. (iii) Second round of antibody binding detects many of the same proteins. We found that our surface allows for two successful rounds of binding.

Figure 10



Single-molecule protein quantification. Solid line: We demonstrate a linear relationship between the number of antibody molecules and the number of protein molecules on the surface when detecting a purified protein sample. Linear fit $R^2 = 0.988$; coefficient of variation = 1 - 7%; lower limit of detection 55 molecules per $1,000 \mu\text{m}^2$ image (1.4 pg cm^{-2}). Dashed line: We achieve accurate quantification in a complex protein sample. Detection of target protein spiked into undiluted rabbit serum produces a quantification curve that deviates only slightly from quantification of the purified sample. We did not observe any increase in background when detecting in serum.

Figure 11



Quantification of endogenous IgG in serum. Using single-molecule protein quantification, we accurately measured the total IgG levels of a rabbit at various time points after immunization. Top: Single molecule counting at three time points (scale bar = 5 μm). We detected a 70.0% increase ($\pm 8.1\%$) in IgG levels between pre-immunization and week 4, and a 11.7% increase ($\pm 4.4\%$) between weeks 4 and 5. Bottom: Bar graph representations of the above single-molecule counting data and of ELISA validation data (black = pre-immunization; gray = week 4; white = week 5). Deviation of single-protein counting measurements from ELISAs were at most 4.2%. The single-molecule counting data and ELISA data were normalized to the pre-immunization time point.

CHAPTER THREE: NANOGEL COATINGS FOR IMPROVED SINGLE-MOLECULE DETECTION SUBSTRATES

This chapter embodies a manuscript published in *Journal of the Royal Society, Interface* on Feb 12, 2011. This chapter demonstrates that nanogel surface coatings can improve the performance of single-molecule immunoassays as well as provide the best surface coating for single-molecule detection studies, such as those that involve *in vitro* binding and single-molecule biophysics. This work was a collaboration with the Elbert Lab and the Jun Lab and is a testament to the atmosphere of collaboration that our university harbors. The experiments were designed by Donald L. Elbert, Robi D. Mitra, and me. Casey D. Donahoe synthesized the nanogel solutions and performed cell adhesion time course assay. Young-Shin Jun and Daniel L. Garcia conducted the atomic force microscopy studies. I created the nanogel-coated surfaces and carried out the single molecule adsorption assays. I also performed the antibody binding assay, imaging and analysis.

ABSTRACT

Surfaces that resist protein adsorption are important for many bioanalytical applications. Bovine serum albumin (BSA) coatings and multi-arm poly(ethylene glycol) (PEG) coatings display low levels of non-specific protein adsorption and have enabled highly quantitative single-molecule protein studies. Recently, a method was developed for coating glass with PEG-BSA nanogels, a promising hybrid of these two low-background coatings. We characterized the nanogel coating to determine its suitability for single-molecule protein experiments. Single-molecule adsorption counting revealed that nanogel-coated surfaces exhibit lower protein adsorption than covalently-coupled BSA surfaces and monolayers of multi-arm PEG, so this surface displays one of the lowest degrees of protein adsorption yet observed. Additionally, the nanogel coating was resistant to DNA adsorption, underscoring the utility of the coating across a variety of single-molecule experiments. The nanogel coating was found to be compatible with surfactants, whereas the BSA coating was not. Finally, applying the coating to a real-world study, we found that single ligand molecules could be tethered to this surface and detected with high sensitivity and specificity by a digital immunoassay. These results suggest that PEG-BSA nanogel coatings will be highly useful for the single-molecule analysis of proteins.

INTRODUCTION

Single-molecule (SM) fluorescence microscopy studies hold great promise for elucidating biological systems[78], but the non-specific surface adsorption of fluorescently labeled proteins [60, 68], antibodies [76], and bioconjugated nanoparticles[69] is often a significant source of experimental noise. Recently, low-background surface coatings have been

developed that reduce protein adsorption to SM levels – levels at which a digital signal from individual target molecules can be reliably quantified above the background of non-specifically adsorbed molecules. For example, Tessler et al surveyed 12 different surface coatings and found that the best performing of these, a covalently-coupled bovine serum albumin (BSA) coating, allowed accurate protein quantification of as few as 55 molecules per $1,000 \mu\text{m}^2$ by SM antibody binding. The low-background surface enabled target protein molecules to be tethered to the surface and digitally detected with total internal reflection fluorescence (TIRF) microscopy and without the need for fluorescence resonance energy transfer or two-color coincidence detection.[76] Poly(ethylene glycol) (PEG) coatings that are highly resistant to protein adsorption have also been developed.[49] Groll et al demonstrated that a monolayer of multi-arm PEG covalently-coupled to a surface reduced protein adsorption to single-molecule levels, allowing quantitative monitoring of protein folding by TIRF.[79]

Recently, Scott et al developed a method for coating surfaces with nano-scale hydrogels (nanogels) formed by crosslinking multi-arm PEG to albumin.[43] Surfaces treated with these nanogels displayed very low levels of protein adsorption, as measured by optical waveguide lightmode spectroscopy (OWLS) and cell adhesion. The nanogel coating holds great potential for SM applications because it is a hybrid of two good SM surface coatings and is thin enough ($\cong 75$ nm) to perform total internal reflection imaging.

Here we characterized nanogel-coated surfaces for use in SM protein studies. We measured protein adsorption onto PEG-BSA nanogel coatings using a variety of proteins, fluorophores, and labeling methods, as well as with DNA. We examined the resilience of the nanogel coating to sodium dodecyl sulfate (SDS), a reagent commonly used in binding studies. To assess the utility of nanogel coatings in the context of a real-world SM experiment, we adapted the surface to perform digital measurements of antibody binding. Finally, we examined the roles of capping steps, cross-linker molecules, and substrate coupling methods via adsorption studies and atomic force microscopy (AFM). Our studies reveal that PEG-BSA nanogel surface

coatings show substantially higher performance than previously characterized SM coatings and should be of benefit to SM protein studies.

RESULTS

Nanogel coatings display lower protein adsorption than BSA or PEG coatings

We first sought to quantify antibody adsorption onto PEG-BSA nanogel-coated surfaces. We generated covalently coated BSA surfaces, multi-arm PEG monolayer-coated surfaces, and nanogel-coated surfaces within flow cells (Figure 12A), exposed them to fluorescently labelled antibody, and quantified the adsorbed molecules by TIRF imaging. For adequate sampling, images were acquired at five positions per surface, with independent surfaces analyzed in triplicate. So, thousands of molecules were sampled to obtain each reported data point. (For details see Supporting Information.)

We found that antibody adsorption onto uncoated control surfaces was too high to allow single antibodies to be reliably resolved. In contrast, the BSA-, multi-arm PEG-, and nanogel-coated surfaces were highly resistant to antibody adsorption (Figure 12B). The BSA-coated, multi-arm PEG-coated, and nanogel-coated surfaces adsorbed antibody at densities of 217, 117, and 50 molecules per 1,000 μm^2 image, respectively (Figure 12C). Thus, the nanogel-coated surfaces adsorbed over four-fold less antibody than did the BSA-coated surfaces ($p = 5.5 \times 10^{-5}$, t-test) and two-fold less antibody than did the multi-arm PEG surfaces ($p = 7.6 \times 10^{-4}$, t-test). Cell adhesion studies agreed with this finding and additionally showed that nanogel-coated surfaces can perform well for 7-9 days compared to 1-3 days for covalently coupled BSA surfaces (Figure S6).

By converting the observed molecular density on the surface to absolute mass density, we determined that the nanogel-coated surfaces adsorbed 1.28 pg cm^{-2} of antibody. Thus, the PEG-BSA nanogel coating is highly resistant to antibody adsorption, outperforming both the BSA-only coating and the PEG-only coating ($p = 4.0 \times 10^{-6}$, ANOVA). Notably, the adsorption we

measured was ~1000-fold lower than the limit of detection of standard protein adsorption measurement methods (e.g. OWLS and surface plasmon resonance).

Nanogel coatings resist adsorption of a variety of biomolecules

SM protein studies often involve labeled biomolecules other than antibodies, such as enzymes and DNA molecules. Also, studies employ proteins that are labeled by a variety of methodologies including single-dye labeling, multi-dye labeling, and fusion to fluorescent proteins. We wanted to investigate the utility of the nanogel surface in a variety of contexts. To this end, we interrogated nanogel coatings with Cy5-streptavidin (multi-labeled protein), mCherry-*E. coli* methionine aminopeptidase (singly-labeled enzyme), and Cy5-thrombin binding aptamer (singly-labeled DNA).

The protein, enzyme, and DNA were adsorbed at levels of 33, 573 and 225 molecules per 1,000 μm^2 onto the BSA-coated surface and 11, 75 and 19 molecules per 1,000 μm^2 onto the nanogel-coated surface, respectively (Figure 13A-13F). For these three biomolecules, adsorption onto nanogel coatings was significantly improved over a covalently-coupled BSA coating (respectively: $p = 0.025$, $p = 0.0002$, $p = 0.0001$, t-test) (Figure 13G). Remarkably, the improvement by the nanogel surface ranged from 3-fold to 12-fold, depending on the biomolecule. So, the relative improvement by the nanogel coating is dependent on the size and/or the biochemistry of the fluorescent biomolecule. The fact that the nanogel coating showed improved performance across a variety of biomolecules underscores the potential utility of the nanogel coating in a broad variety of SM studies.

Nanogel-coated surfaces are resilient to detergent

In solid phase *in vitro* binding studies, surfactants such as Triton X-100, Tween 20, and SDS are commonly used for performing wash steps or to regenerate ligands after protein binding. [75, 76] Therefore, we measured the resilience of PEG-BSA nanogel coatings to surfactant

exposure. This was done by quantifying protein adsorption onto coated surfaces before and after an SDS wash.

Surfaces were prepared, exposed to fluorescently labelled antibody, washed with 0.1 per cent SDS, exposed to antibody a second time, and washed with SDS a second time. Imaging was performed before and after each step to quantify the amount of attached antibody.

We found that the BSA coating was strongly affected by the SDS wash (Figure 14A, row 1) while the nanogel coating was mostly resilient (Figure 14A, row 2). The BSA coating displayed a 4.1-fold increase in antibody adsorption after SDS exposure as compared to before SDS exposure, whereas the nanogel coating displayed only a 1.7-fold increase in antibody adsorption (Figure 14B). Thus, thin, PEG-BSA nanogel coatings should be superior to BSA coatings for experiments that utilize surfactants.

Nanogel coatings are compatible with digital antibody binding experiments

Protein adsorption is an important parameter to consider when evaluating a surface for SM protein experiments, but it is not the only one. If target molecules cannot be tethered to the surface, or if the tethered molecules are not accessible for *in vitro* binding, then the method will lack sensitivity of detection. Conveniently, nanogel coatings and BSA coatings contain surface-exposed carboxyl groups (provided by the albumins), which allow for the near-universal tethering of protein analytes via the use of a crosslinker (see Materials and Methods).[76] Making use of this tethering method, we evaluated the compatibility of PEG-BSA nanogel coated surfaces with digital antibody binding experiments.

We generated nanogel-coated surfaces and tethered target protein molecules labeled with Cy3 to the surfaces. Then we bound the targets with Cy5-labeled antibody, washed away unbound antibody, and used single-molecule TIRF to detect antibody-target binding (Figure 15A). We merged and processed the Cy3 and Cy5 channels (Figure 15B-15C), and correlation

analysis showed significant levels of specific antibody binding (Figure 15D, see Supporting Information for details).

By analyzing single molecules in the two channels, we observed that the antibodies bound to their targets with high specificity: $68 \pm 1\%$ of antibodies were bound specifically to tethered, target proteins. We also found that the antibodies were able to bind to target molecules *sensitively*: a substantial fraction of targets were bound by antibody ($50.3 \pm 5.9\%$) and this fractional binding is high compared to values we obtained previously using the covalently coupled BSA surface ($31 \pm 6\%$).^[76] Notably, this is not the maximum fractional binding since more will bind with higher antibody concentrations. Also, this level of fractional binding is a lower-bound estimate, since some target molecules may be invalid (e.g. denatured proteins).

These results indicate that the nanogel coating allows targets to be immobilized and bound with high antibody occupancy and low background. Thus, nanogel coatings are highly suited for performing digital antibody binding experiments and should be generally applicable to SM *in vitro* binding studies.

Investigation into the role of albumin in nanogel capping and cross-linking

Since the final steps of both the nanogel and the multi-arm PEG coatings involve capping with BSA, we wanted to investigate to what extent the performance of the coatings depends on capping. We performed SM adsorption measurements on nanogel and multi-arm PEG coatings with and without capping steps. We found the nanogel coating behaved the same with and without the capping step, whereas the protein resistance of the multi-arm PEG coating was reliant on capping (Figure 16A). This is presumably because the uncapped nanogel, which is formed from a partially polymerized solution containing a wide size-range of PEG-BSA complexes, creates a surface that is more densely packed than that of the uncapped multi-arm PEG.

Next, since the BSA molecules in the nanogel act as a multi-functional amine crosslinker, we wanted to investigate whether a multi-arm amine-terminated PEG could be substituted for and provide the same performance as BSA. We synthesized PEG-PEG_{OA} nanogels by substituting BSA with PEG-octoamine as has been described in previous work.[80] We applied this coating and the PEG-BSA nanogel coating to glass and found that adsorption using the alternative crosslinker detracted from the performance of the coating (Figure 16B). This result points to the possibility that the PEG-BSA nanogel performance is contingent on a high molecular weight crosslinking molecule – in this case, BSA.

Investigation into alternative coupling chemistries

We next explored the use of an alternative method for coupling nanogel solutions to glass substrates with the hope of further reducing protein adsorption. In our previously described experiments, we prepared surfaces by reacting the nanogel-containing solutions with mercaptosilanated glass.[43] Coupling was achieved because the vinylsulfone groups in the PEG portion of the nanogels react with the mercaptosilanated surfaces. An alternative coupling strategy is to react the amine groups in the BSA portion of the nanogel with an epoxysilanted surface.

We prepared surfaces using this alternative coupling chemistry and measured protein adsorption. The epoxide-reacted nanogel coating adsorbed antibodies at a density of 103 molecules per 1,000 μm^2 image (Figure 16B), so this surface adsorbed approximately two times more protein than did the thiol-reacted nanogel surfaces ($p = 5.6 \times 10^{-4}$, t-test). We concluded that thiol-coupled nanogels are more resistant to protein adsorption than epoxide-coupled nanogels.

We were surprised to observe this difference in performance simply due to coupling chemistry. To try to understand this, we used AFM to characterize their morphologies (see Supporting Information). The standard, thiol-reacted nanogel surface displayed a homogeneous

background (roughness value $R_q = 0.480$ nm), suggesting that the nanogels form a continuous matrix on the surface (Figure 16C). The nanogels on the epoxide-reacted surface appeared to be generally smaller and less homogeneous (roughness value $R_q = 0.673$; Figure 16D). By contrast the multi-arm PEG coated surface formed a less-connected and non-homogeneous surface (roughness value $R_q = 0.823$; Figure 16E). (Force curve analysis performed on the thiol-reacted nanogel coating, the epoxy-reacted nanogel coating, and the multi-arm PEG coating showed detachment forces of 19.087, 41.584, and 34.225 nN respectively.) We believe the high connectivity of the thiol-reacted nanogels, reflected by roughness measurements, contributes to the lower protein adsorption we observed.

DISCUSSION

Under the SM detection regime, nanogel coatings provided significantly lower protein adsorption than BSA-coated or multi-arm PEG monolayer-coated substrates. Moreover, there was a consistent gain in performance across three protein types and DNA, labeled by dyes and fluorescent proteins. The nanogel surface was resilient to surfactant and was compatible with a SM antibody binding experiment. We did not seek to optimize the fractional binding in these immunoassays (which depends on antibody concentration), but for the concentration of antibody used here, the fraction of ligands bound was at least as high as that obtained previously in digital antibody binding experiments using BSA-coated surfaces.[76] Since the nanogel surface outperformed the covalently-coupled BSA and the multi-arm PEG SM surfaces in adsorption and resilience[76, 79], we conclude that the nanogel surface should be a useful coating for an array of SM studies.

Surfactants can be useful in *in vitro* binding experiments, such as immunoassays, for performing stringent washes and for un-binding antibodies from their epitopes in repeated-binding experiments. A caveat is that surfactants can be detrimental to SM surface coatings. We studied the effect of SDS surfactant on the nanogel coating and the BSA coating and found that

the nanogel-coated surfaces were unaffected by treatment with surfactant while the BSA-coated surfaces were substantially affected. This is surprising because the nanogels are partially composed of albumin. It can be inferred that the BSA-only coatings undergo some level of denaturation by the surfactant conditions. If this is the case, then the PEG macromolecules in the nanogels may stabilize the albumin molecules within them and enable them to withstand the denaturing conditions of the surfactant. This highlights another improved behavior that the combination of albumin and PEG in a nanogel can confer over BSA-only surfaces. It is promising, since it may enable lower-background and repeated-binding digital antibody binding studies.

SM methods play a prominent role in the understanding of many biological systems.[81-86] Looking forward, SM antibody-detection may provide strategies for parallel detection of proteins. The recent development of massively parallel SM DNA analysis technologies by Helicos Biosciences and Pacific Biosciences was facilitated by surfaces resistant to nucleic acid adsorption.[87, 88] Promisingly, the nanogel coating characterized here shows similar levels of background adsorption to that of such DNA technologies and also benefits from resilience to wash steps. Therefore, PEG-BSA nanogels provide a thin, resilient coating technique that should benefit SM protein analysis, and we hope that this coating methodology will spark an increase in SM antibody binding studies.

MATERIALS AND METHODS

Synthesizing solutions of nanogels

PEG-BSA nanogel coatings were prepared using partially crosslinked solutions of eight arm PEG and BSA. First, eight arm PEG-OH (PEG₈-OH; MW 10,000; Creative PEGWorks, Winston Salem, NC) was used to synthesize PEG-octovinylsulfone (PEG_{OVs}) in a four-step reaction, as described previously.[80, 89] PEG_{OVs} and fatty acid-free BSA (Sigma-Aldrich, St. Louis, MO) solutions were prepared at 20% (w/v) in phosphate buffered saline pH 7.4 (PBS) and

sterile filtered with 0.22 μm syringe filters (Millipore, Billerica, MA). Assuming 36 reactive amines per BSA (obtained from crystal structures), PEG_{OVS} and BSA solutions were combined at a 1:1 ratio of vinylsulfone to amine groups and reacted by incubation at 37°C, with rotation. The progress of the reactions was followed by dynamic light scattering (DLS) until a mean effect diameter (d_{PCS}) of about 100 nm was achieved – generally ~6 hours. (At this step, the DLS polydispersity index for the nanogel solution was 6.2, and the complexes in the nanogel solutions had a standard deviation of ~20 nm.) The nanogel-containing solution was then diluted 1:1 with PBS to 10% (w/v) and stored until use at -80°C. For long-term storage, nanogel solutions were stored at -140°C. To generate PEG-PEG_{OA} nanogels, the same protocol as above was used with the exception that BSA was substituted by PEG-octoamine (PEG_{OA}, MW 10,000). (For details, see Supporting Information.)

Coating substrates with nanogels

Nanogel solutions were either thiol-reacted or epoxy-reacted to the glass substrates. The glass substrates (40 mm diameter circles, #1.5) were obtained from Erie Scientific (Waltham, MA). For the epoxide-reacted coatings, the substrates were epoxysilanated by the vendor. For the thiol-reacted coatings, the coverslips were mercaptosilanated as follows. The coverslips were washed 3x in diH₂O and 3x in ethanol. Then they were etched with oxygen plasma at 100 W for 10 minutes (Femto, Diener Electronic, Royal Oak, MI). Coverslips were then washed 3x in acetone. Mercaptosilanation was achieved by treating the glass for 1 hr at 25°C with a 5% (v/v) solution of mercaptopropyltrimethoxysilane (MPTS, Sigma-Aldrich, St. Louis, MO) in acetone. Surfaces were washed 3x in acetone and cured at 100°C for 25 min.

Substrates were coated with PEG-BSA nanogel-containing solutions in either a flow cell (FSC2, Bioptechs, Butler, PA) or in a 60 mm diameter Petri dish (for cell seeding experiments). Substrates were incubated in 10% nanogel-containing solutions for 1 h at 37° C. Substrates were

washed with 600 μ L PBS and capped for 1 h with BSA 50 mg/mL in PBS, 37° C. Unreacted vinylsulfone or epoxide groups were then quenched with 1 M Tris pH 8.0, for 15 minutes at room temperature. (A time course showed that quenching went to completion.) Coverslips were then washed with 600 μ L PBS.

Covalently coupled BSA-coated surfaces

The epoxide-reacted bovine serum albumin (BSA-coated) surfaces were generated by reacting 50 mg/mL BSA, in PBS, to epoxysilanated glass coverslips within a flow cell for 1 h, at 37° C. Unbound BSA was washed away with PBS, and unreacted epoxides were quenched with 1M Tris pH 8.0 for 15 minutes at room temperature. Coverslips were then washed with 600 μ L PBS.

Multi-arm PEG monolayer-coated surfaces

The multi-arm PEG monolayers were generated by reacting 200 mg/mL PEG_{OVS}, in PBS, to mercaptosilanated glass coverslips within a flow cell for 1 h, at 37° C. Coverslips were washed with 600 μ L PBS and incubated 1 h with 50 mg/mL BSA in PBS at 37° C. Unreacted vinylsulfone groups were quenched with 1M Tris pH 8.0 for 15 minutes at room temperature. Coverslips were then washed with 600 μ L PBS.

SM adsorption measurements

Three different proteins and one DNA were used in the SM adsorption experiments. These include polyclonal goat IgG labeled with multiple Cy5 fluorophores (Abcam, Cambridge, MA), streptavidin labeled with multiple AlexaFluor 647 fluorophores (Invitrogen, Carlsbad, VA), E. coli methionine aminopeptidase fused to mCherry fluorescent protein, and DNA thrombin binding aptamer labeled with a single

Cy3 fluorophore (Integrated DNA Technologies, Coralville, IA). Each of the surfaces under investigation was prepared within a flow cell (FSC2, Bioptechs, Butler, PA). An uncoated control surface was generated by quenching an epoxysilanated glass coverslip with 1M ethanolamine-HCl pH 8.0 for 30 minutes. Flow cells were fitted with perfusion ports to allow for reagents to be passed over the surface by a custom vacuum pump. The flow cells were washed with 600 μ L PBS and loaded with 200 μ L of 1 nM fluorescent protein or DNA. The fluorescent molecules were incubated for 25 minutes in the dark at room temperature, and unbound protein or DNA was washed off with 600 μ L PBS. Images were acquired and processed as described above. Standard deviations were obtained from triplicate (for antibody) or duplicate (for all other molecules) surfaces.

Measuring detergent resistance

Each of the surfaces under investigation was prepared within a flow cell. Surfaces were exposed to 100 ng/mL Cy5-labeled antibody for 25 minutes in the dark at room temperature to assess initial levels of non-specific protein adsorption. Unbound antibody was washed out of the flow cells with 600 μ L PBS, and the flow cells were imaged. The flow cells were then exposed to 0.1% sodium dodecyl sulfate (SDS) in PBS for 5 minutes at room temperature, washed with 600 μ L PBS, and imaged. The flow cells were exposed for the second time to antibody for 25 minutes, to measure adsorption after SDS treatment. Surfaces were washed with 600 μ L PBS, and imaged. Finally, the flow cells were washed in 600 μ L 0.1% SDS in PBS for the second time, washed in 600 μ L PBS, and imaged. Images were processed as described above. Standard deviations were obtained by replicates on two separate surfaces.

Digital immunoassays

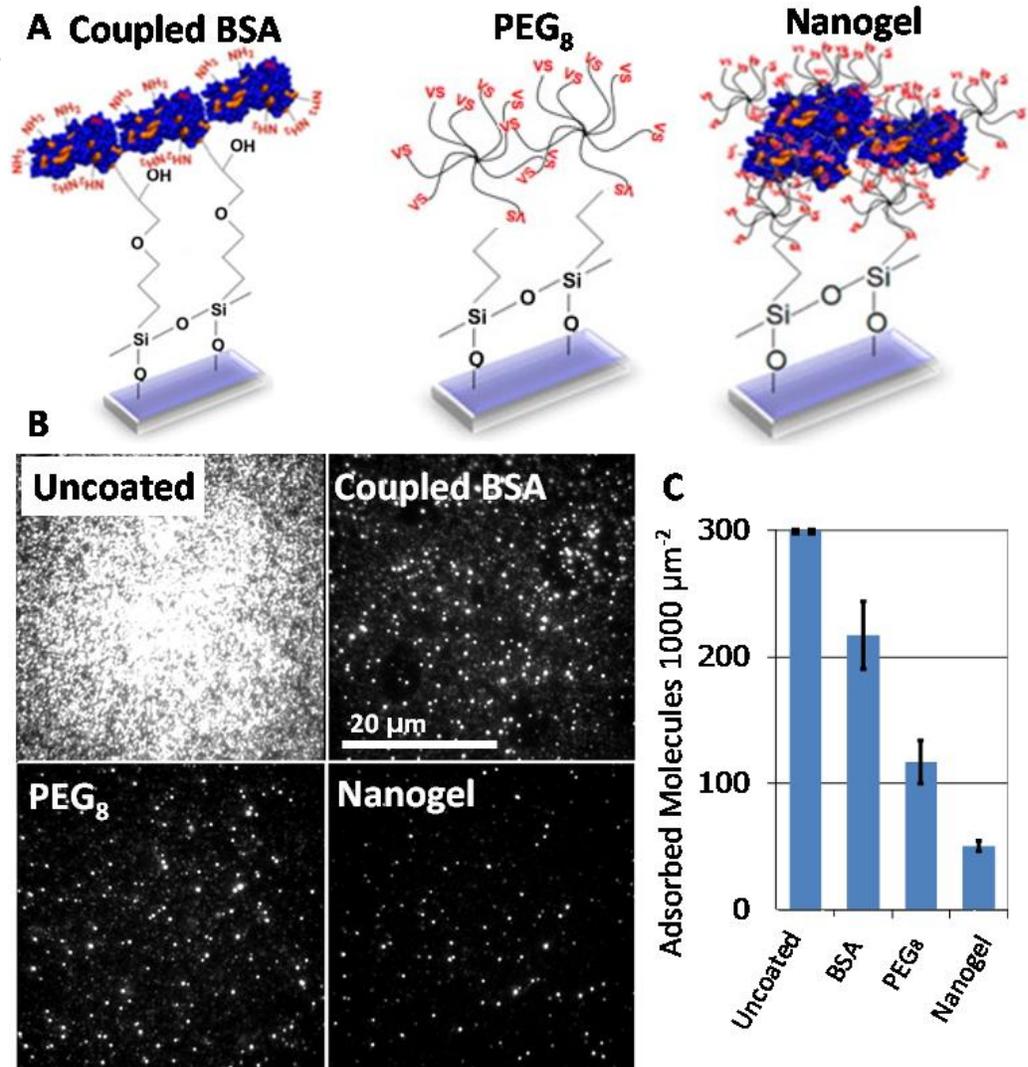
Nanogel-coated surfaces were generated in a flow cell as described above. The antibody binding experiment was performed as previously described.[76] First, the surface was activated by 0.2 M 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) and 0.05 M N-hydroxysuccinimide (NHS) (Pierce, Rockford, IL) in sodium phosphate buffer pH 5.8 (SPB) for 10 minutes. The flow cell was washed with 600 μ L SPB, and Cy3-labeled target protein (IgG obtained from goat, Abcam, Cambridge, MA) was tethered to the activated surface for 10 minutes at 100 ng/mL in PBS, in the dark. Unreacted crosslinking groups were quenched with 1 M Tris pH 8.0 for 5 minutes. Then the surface was probed with Cy5-labeled antibody (anti-Goat IgG, Abcam, Cambridge, MA) for 2 h at 100 ng/mL in PBS in the dark. The flow cell was washed with 600 μ L PBS and imaged at 540 nm and 635 nm. Images of Cy3 and Cy5 channels were merged to determine the fraction of targets that were bound by antibody and the specificity of the antibody for the targets compared to random binding. (See Supporting Information for details.)

ACKNOWLEDGEMENTS

We thank Jim Havranek and Ben Borgo for providing the mCherry-*E. coli* methionine aminopeptidase fusion protein. We thank David Mayhew, Francesco Vallania, Jessica Hoisington-Lopez, Maximiliaan Schillebeeckx, Michael Brooks, Scott Higdon, Todd Druley, Xuhua Chen, and Yue Yun for helpful discussions. This work was supported by an Innovative Molecular Analysis Technologies grant from the National Cancer Institute (1R21CA151197-01) and a Technology Development in Epigenetics grant from the National Institutes of Health (1R01DA025744-01).

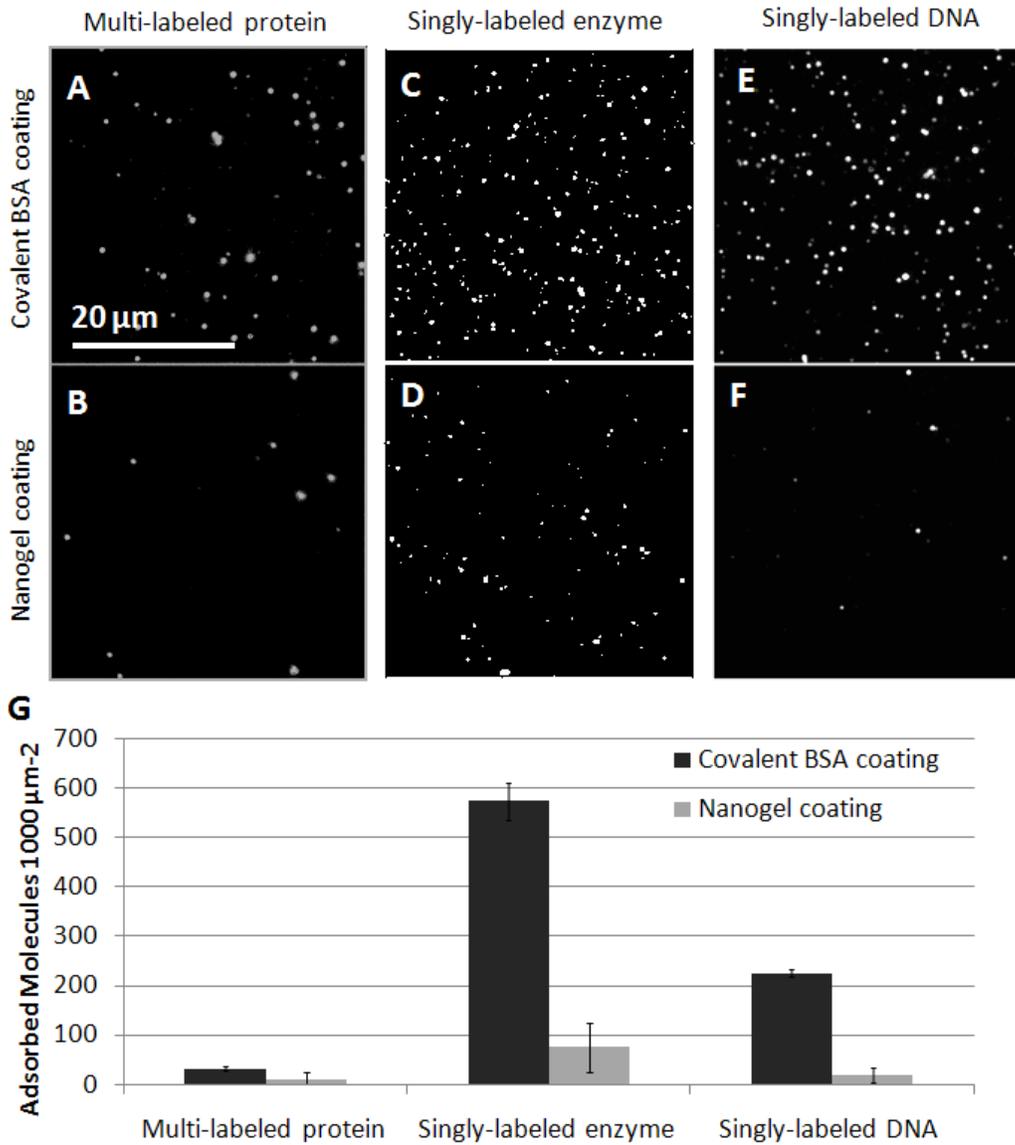
FIGURES

Figure 12



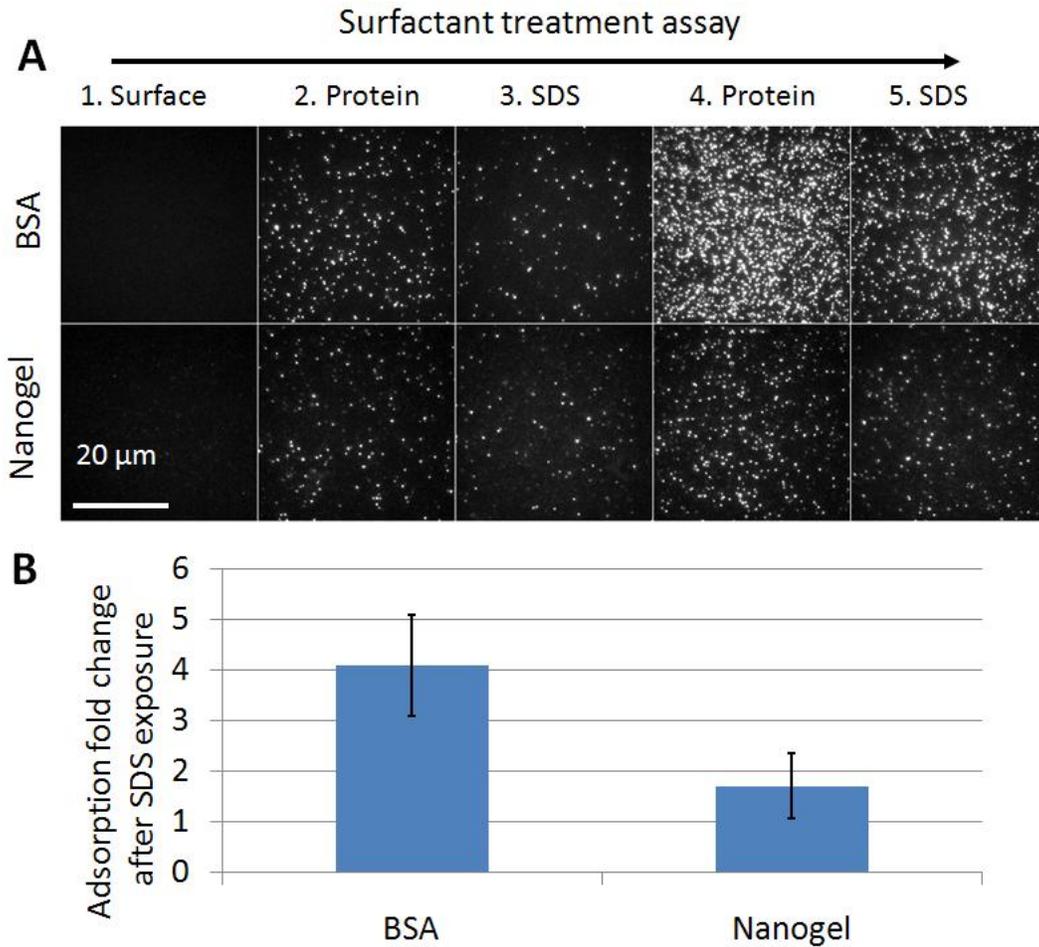
(A) Schematic of three coating methodologies: bovine serum albumin (BSA) covalently coupled to epoxysilanated glass, multi-arm PEG (PEG₈) coupled to mercaptosilanated glass, and PEG-BSA nanogels coupled to mercaptosilanated glass. (B) Antibody adsorption onto uncoated, BSA-coated, multi-arm PEG-coated, and PEG-BSA nanogel-coated surfaces was quantified by TIRF imaging, and representative raw TIRF images are shown. (C) Molecule counts of antibody adsorption per unit area. On the uncoated surface, molecular density was so high that single-molecule counting was not possible. BSA-coated, multi-arm PEG-coated, and nanogel-coated surfaces show decreased antibody adsorption compared to the uncoated surfaces, with the nanogel-coated surface performing the best. Error bars represent standard deviation of triplicate substrates.

Figure 13



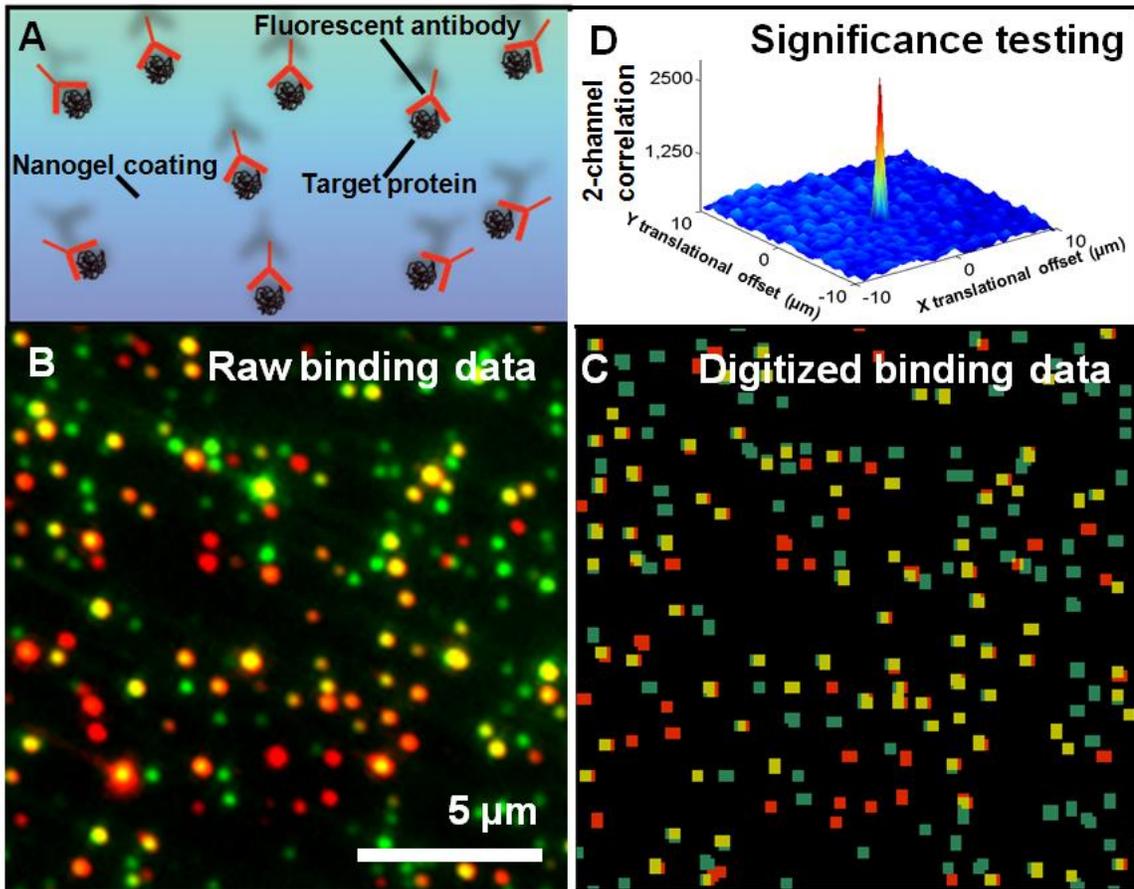
TIRF adsorption measurements for three types of biomolecules show improved performance by nanogel coatings compared to covalently-coupled BSA coatings. The nanogel coating reduces adsorption of (A-B) the multi-labeled protein, Cy5-labeled streptavidin, (C-D) the singly-labeled enzyme, mCherry-*E. coli* methionine aminopeptidase, and (E-F) the singly-labeled DNA, Cy5-labeled thrombin binding aptamer. (G) Molecule counts of biomolecule adsorption per unit area onto (black bars) covalently-coupled BSA coatings and (gray bars) nanogel coatings. Error bars represent standard deviation of duplicate substrates.

Figure 14



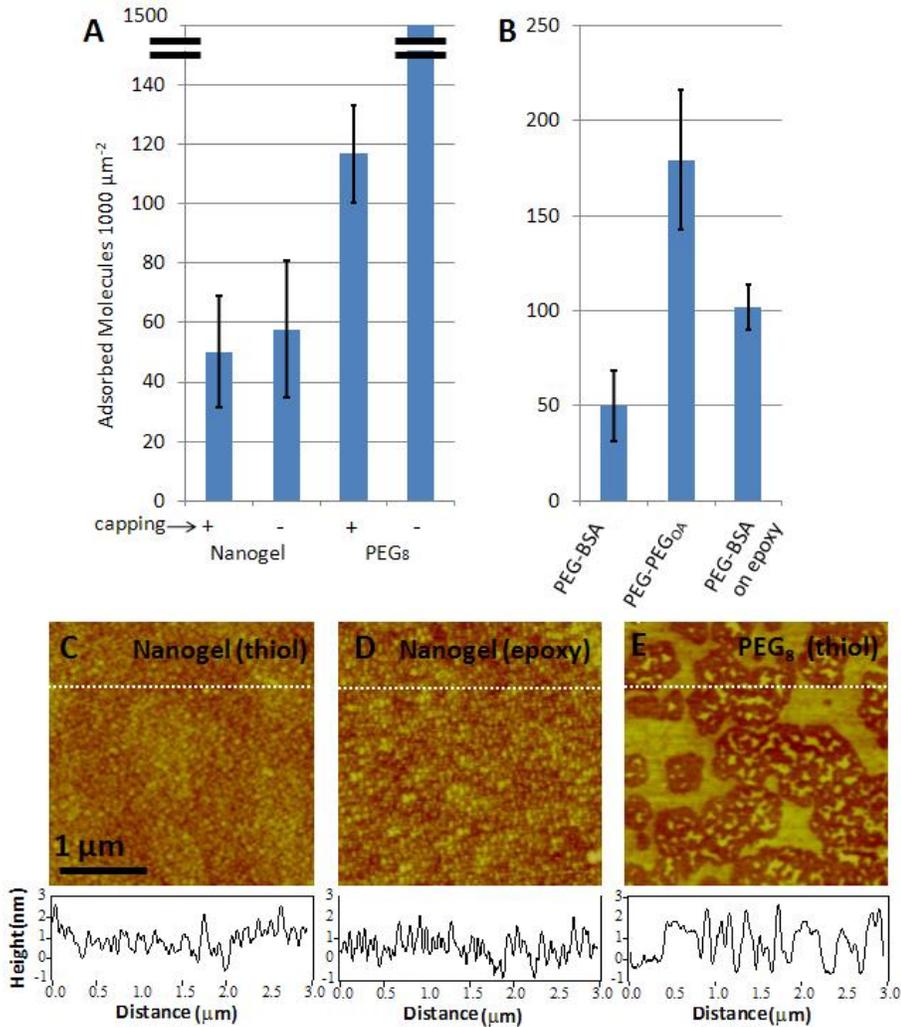
Nanogel-coated surfaces show resilience to a harsh surfactant environment compared to BSA-coated surfaces. Coated surfaces were exposed to antibody, washed with sodium dodecyl sulfate (SDS), re-exposed to antibody, and re-washed with SDS. Adsorption measurements were obtained in between each step. **(A)** Raw TIRF images of (top row) a covalently-coupled BSA-coated surface and (bottom row) a nanogel-coated surface over the course of a 5-step antibody adsorption experiment. (1.) Surfaces prior to antibody exposure. (2.) Surfaces after exposure to antibody. (3.) Surfaces after SDS wash. (4.) Surfaces after second exposure to antibody. (5.) Surfaces after second SDS wash. **(B)** The bar chart depicts the ratio of adsorption change after SDS exposure to adsorption change before SDS exposure, i.e. (Step 4 – Step 3):(Step 2 – Step 1). Error bars represent standard deviation of duplicate substrates.

Figure 15



Nanogel coatings can be utilized for sensitive and specific, digital antibody binding. (A) Cy3-target proteins were immobilized to a nanogel-coated surface and then bound by Cy5-antibody. (B) Merged raw image of red and green TIRF channels (15 μm x 15 μm). (C) Merged red and green TIRF channels after image processing. TIRF imaging was able to detect a high number of binding events (yellow) in which target proteins (green) were accessible to binding by antibodies (red). (D) The correlation between detection antibodies and individual analyte proteins tethered to nanogel coated-surfaces indicates high target protein accessibility and compatibility with TIRF detection.

Figure 16



(A) Adsorption onto nanogel-coated and multi-arm PEG coated (PEG₈) surfaces was measured with and without the use of a final BSA capping step. The nanogel-coated surfaces provided low antibody adsorption even without the capping step. (B) We measured adsorption onto the standard nanogel-coated surface (PEG-BSA). We also measured adsorption onto a surface coated with a nanogel solution formed with an alternative crosslinker to BSA (PEG-PEG_{0A}). Finally, we measured adsorption onto a surface that was coated by coupling PEG-BSA nanogels to epoxysilanated glass (as opposed to the standard, mercaptosilanated glass). We conclude that nanogel crosslinkers and coupling chemistries can have a significant effect on adsorption performance. (C-E) We assessed the topology of (C) the standard, thiol-reacted nanogel surface, (D) the epoxy-coupled nanogel surface, and (E) the multi-arm PEG-coated surface. Cross sectional height analysis is depicted in charts below. The thiol-reacted nanogel coating created the most connected surfaces while multi-arm PEG monolayer coatings generated the least connected surfaces.

CHAPTER FOUR: SENSITIVE SINGLE-MOLECULE PROTEIN QUANTIFICATION AND PROTEIN COMPLEX DETECTION IN A MICROARRAY FORMAT

This chapter embodies a manuscript that is being peer reviewed for the journal *Proteomics*. This chapter describes the development of a single-molecule antibody microarray. The experiments were designed by Robi D. Mitra and me, and the experiments and analysis were conducted by me.

ABSTRACT

Single-molecule protein analysis provides sensitive protein quantitation with a digital read-out and is promising for studying biological systems and detecting biomarkers clinically. However, current single-molecule platforms rely on the quantification of one protein at a time. Conventional antibody microarrays are scalable to detect many proteins simultaneously, but they rely on less-sensitive and less quantitative quantification by the ensemble averaging of fluorescent molecules. Here we demonstrate a single-molecule protein assay in a microarray format enabled by an ultra-low background surface and single-molecule imaging. The digital read-out provides a highly sensitive, low femtomolar limit of detection and 4 orders of magnitude of dynamic range through the use of hybrid digital-analog quantification. From crude cell lysate, we measured levels of p53 and MDM2 in parallel, proving the concept of a digital antibody microarray for use in proteomic profiling. We also applied the single-molecule microarray to detect the p53-MDM2 protein complex in cell lysate. Our study is promising for development and application of single-molecule protein methods because it represents a technological bridge between single-plex and highly multiplex studies.

TECHNICAL BRIEF

Single-molecule protein detection has the potential to improve systems biology and biomarker studies by providing highly sensitive quantification and a digital read-out. Single-molecule protein techniques are achieved by several methodologies, but common among them is the coupling of a single-molecule-sensitive detection modality with a method for eliminating background fluorescence. In one approach, single-molecule

protein quantitation is achieved through the dilution of antibody-target complexes in low cross-section capillaries followed by detection by a sensitive photon detector[90]. In another instance, femtoliter-volume wells are used to harbor single-molecule enzyme-linked immunoassays[91]. Additionally, total internal reflection fluorescence (TIRF) imaging has provided a platform for single-molecule quantification on planar surfaces. TIRF analysis is especially promising because of the reliability and affordability of TIRF optics and has been recently demonstrated for the digital quantification of proteins[76, 92-94] and lipopolysacharrides[95]. Despite the quantitative advantages of all of these single-molecule methods, they are currently low throughput in that they can analyze only one target at a time.

Microarrays are advantageous for proteome and interactome profiling because they scale reliably for dozens of protein targets while relying on minimal reagent volumes [57, 96] and so have proven valuable for quantifying the abundance of many proteins simultaneously[97-100] and for detecting pairs of interacting proteins[101-105]. However, microarrays have neither the sensitivity of detection nor the precise digital read-out provided by single-molecule methods. For these reasons, a single-molecule assay for proteins that has the scalability of a microarray has been a major goal[92, 93].

Here we perform a proof-of-principle demonstration of a single molecule antibody microarray. Our procedure begins by treating a glass substrate with a nanogel coating that forms a protein resistant, hydrogel barrier of around 75 nm thickness[43]. The coating is activated with a cross-linker, making it reactive with the exposed lysines of printed antibodies. Then, monoclonal capture antibodies are printed in specified locations (fields) onto the substrates. Substrates are sealed to a flow chamber, the surface

is passivated, and cell lysate is passed over the chip, binding target molecules to the antibody fields. Then, the chip is exposed to detection antibodies and finally to fluorescently-labeled streptavidin (Figure 17).

To obtain a single-molecule resolution read-out, total internal reflection fluorescence (TIRF) imaging is performed. The microscope is directed to the locations of the printed fields, and images are acquired within the center of each field. Non-overlapping viewing areas are acquired within each printed field to obtain intra-field replicates (multiple fields and multiple slides are imaged as well). Then, digital measurements are obtained by counting fluorescence objects within the printed fields and subtracting away background levels, which are measured outside of the fields.

We first sought to assess the performance difference between a digital and an analog microarray in terms of limit of detection (sensitivity) and dynamic range. To do this, we used the model assay that detects fluorescently labeled streptavidin by binding to a biotinylated protein printed on the surface (to emulate a protein target captured by a printed capture antibody). This model is a suitable estimate of performance since it incorporates (1) the printing/attachment efficiency of the capture molecule, (2) the pull-down of a protein target from solution, and (3) the non-specific binding of the protein target on the surface.

We obtained TIRF images for different concentrations of captured protein (Figure 18A). Then we analyzed the same raw data by two different methods to obtain two standard curves – one by single-molecule counting (digital) and one by conventional, ensemble intensity averaging (analog). The standard curve obtained from analog image processing provided a dynamic range from 10 pM to 100 pM (Figure 18B), and the curve

obtained from digital processing of the same data provided a dynamic range from 14 fM (0.74 pg mL^{-1}) to 33 pM (Figure 18C). So, given the same raw data, analysis by digital counting provided greater sensitivity than ensemble averaging by around 3 orders of magnitude. By combining the two standard curves (using the digital curve for the 14 fM to 33 pM range and the analog curve for the 33 pM to 100 pM range), the hybrid digital-analog curve provided around 4 orders of magnitude dynamic range. These results show that single-molecule microarrays, by providing a high resolution view of the printed antibody fields, can provide more sensitive detection than analog microarrays. Also, single-molecule microarrays can allow for extension of dynamic range by combining digital and analog standard curves. The dynamic range we observed here is well suited for the analysis of biological systems and serum biomarkers.

We applied the single-molecule antibody microarray to analyze the regulatory proteins p53 and MDM2. We generated p53/MDM2 microarrays as described above using capture antibodies for p53 and MDM2. We benchmarked the microarrays using dilutions of protein standards to validate each antibody pair individually and then tested for cross-reactivity of the antibodies in a parallelized (dual-plex) assay (Supporting Figure 1). To validate the single-molecule antibody microarray in a real world application, we quantified p53 and MDM2 protein levels in a well-characterized cell culture line, HCT116. This colorectal cancer line is known to respond to the DNA damage agent 5-fluorouracil by up-regulating p53 and its downstream targets[106, 107]. The p53/MDM2 single-molecule microarray was exposed to cell lysate, protein levels were detected in a single detection antibody step, and the slide was imaged and analyzed. Both p53 and MDM2 protein levels fell within the digital range of the assay

(Figure 19A). We found p53 protein expression to increase by 1.8-fold under DNA damage conditions and MDM2 protein expression to increase by 6.6-fold under DNA damage conditions (Figure 19B). These changes are in agreement with a recent study using similar induction conditions of cancer cells (2.3-fold and 5.5-fold, respectively) [108]. The results from this indicate the single molecule microarray is well suited for cellular profiling. Promisingly, the microarray format should make the assay readily adoptable to more highly multiplexed studies.

Antibody microarrays have been widely used for the discovery of proteome-wide interactions[101-105], and we hypothesized that a single-molecule antibody microarray could be used to detect a protein complex. To achieve this, we altered the detection step of the p53/MDM2 microarray protocol by incubating only one detection antibody at a time. In this way, we could quantify molecules located in the alternate field with respect to the detection antibody. To establish the specificity of this protocol, we analyzed solutions that contained one protein (e.g. p53) but lacked the other (e.g. MDM2). In these controls, we observed low signal from two distinct immunosandwich assays: anti-p53 detection antibody binding to anti-MDM2 capture fields and anti-MDM2 detection antibody binding to anti-p53 capture fields.

After establishing the levels of cross reactivity of the assay, we analyzed cell lysate from HCT116 cells. By the two independent immunosandwiches, we observed p53-MDM2 complex levels significantly above the controls (Figure 19C). The difference in magnitude between the two lysate measurements is indicative of the difference in affinity of the pairs of antibodies in the two different immunosandwiches. The mass of signal in both of the interaction assays indicates the presence of a protein complex, which

is consistent with the known protein-protein interaction between p53 and MDM2 *in vivo* and *in vitro* [109, 110]. Thus, we present a model of how a single-molecule antibody microarray may be used in protein-protein interaction studies.

Single molecule quantification of proteins should become of greater benefit to disease diagnostics and systems biology by providing quantitation of rare proteins in biological specimens. We used standard antibody microarray reagents, a nanogel-coated surface, and TIRF imaging to provide duplex digital molecule counting in a planar format that is suitable for higher degrees of multiplexing. As with all antibody microarrays, scaling up to greater numbers of targets requires optimization to minimize cross-reactivity. In this study, we focused on the platform itself – surface architecture, antibody immobilization, and detection – which dictates assay performance to a large degree [111-113]. Using a representative set of antibodies from a popular vendor, we were able to detect with a digital read-out regulatory proteins p53 protein and MDM2 from crude cell lysate and quantify small expression changes (1.8-fold). We also modified the assay to demonstrate its utility in a protein-protein interaction study. The low femtomolar sensitivity and 4-orders of magnitude dynamic range we observed provide a proof of principle for the use of single-molecule antibody microarrays in larger-scale protein quantification and protein-protein interaction studies.

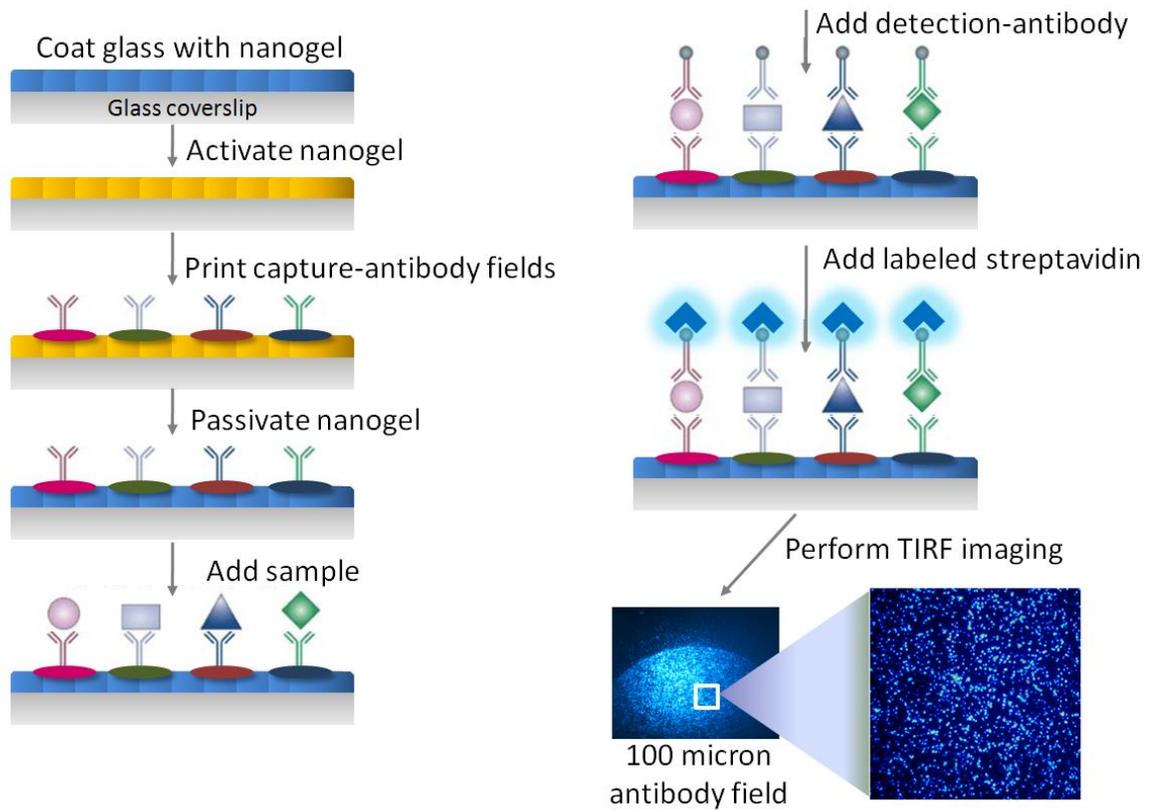
ACKNOWLEDGEMENTS

We thank Donald Elbert for helpful suggestions regarding nanogel synthesis. We thank David Mayhew for assistance with cell culture. We thank Scott Higdon, Aaron Spivak, and Robert Zeigler for helpful discussions regarding protein biochemistry. We

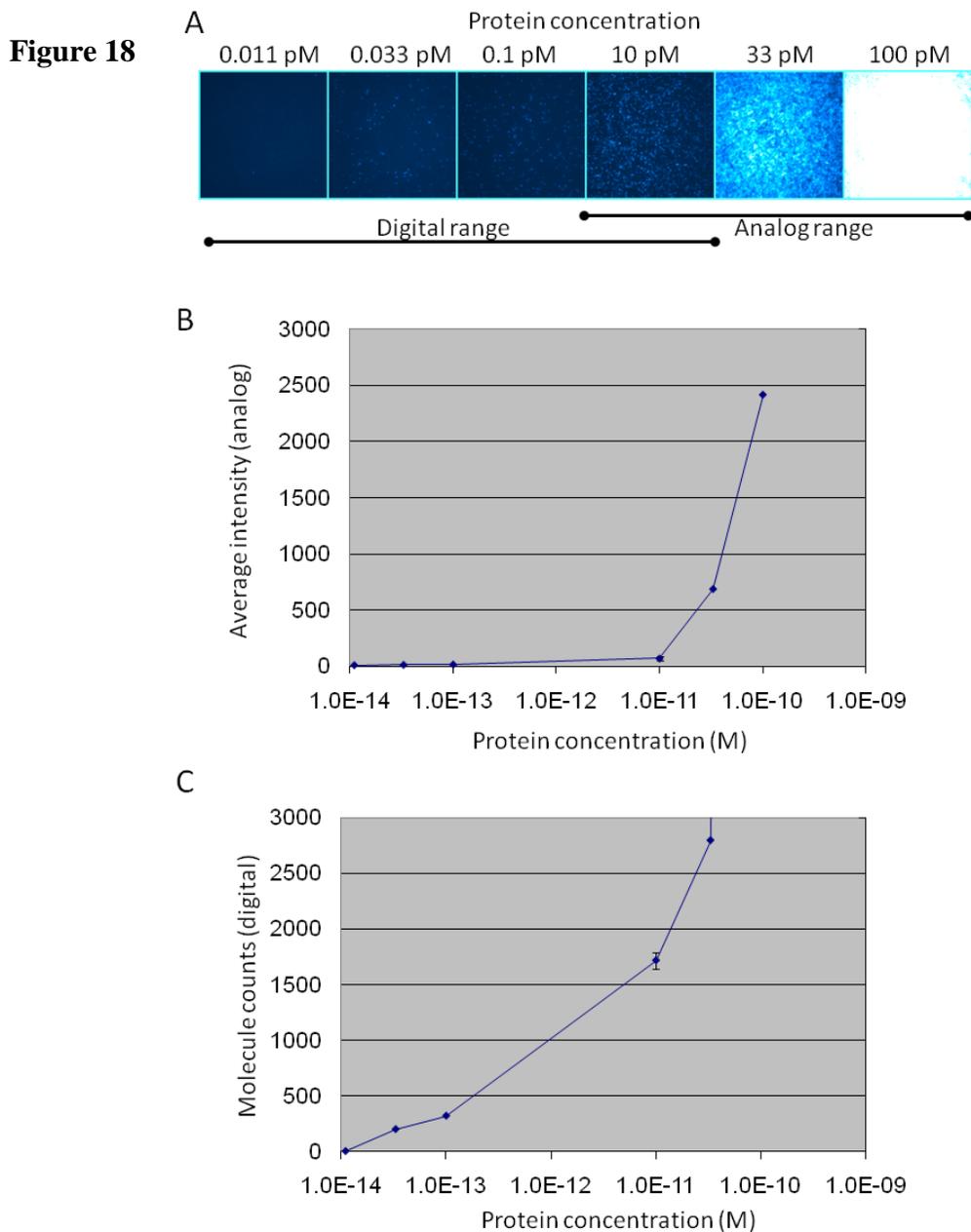
thank Seth Crosby, Chris Sawyer, and Neil Winegarden for advice regarding microarray printing. We thank Francesco Vallania for suggestions and critical reading of the manuscript. We thank Michael Brooks, Maximiliaan Schillebeeckx, Yue Yun, and Xuhua Chen for helpful discussions. This work was supported by an Innovative Molecular Analysis Technologies grant from the National Cancer Institute (1R21CA151197-01) and a Technology Development in Epigenetics grant from the National Institutes of Health (1R01DA025744-01).

FIGURES

Figure 17

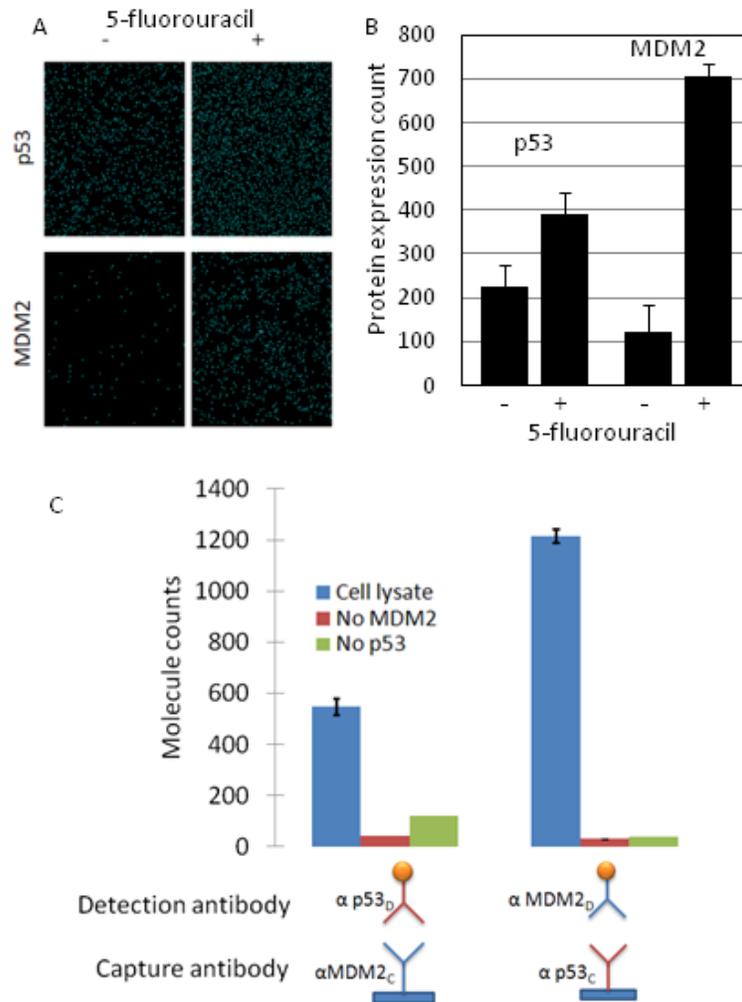


Scheme of the single-molecule antibody microarray method.



(A) Representative TIRF images of fluoro-streptavidin captured onto the single-molecule microarray printed with biotin capture fields. (B) Standard curve generated by analyzing the data by conventional, fluorescence intensity averaging (analog processing). (C) Standard curve generated by analyzing the data by discrete molecule counting (digital processing). The single-molecule microarray enables higher sensitivity and higher dynamic range than a microarray analyzed by analog processing.

Figure 19



(A) Single-molecule protein expression of p53 and MDM2 in cancer cells, with and without DNA damage induction. (B) Expression changes of 1.8-fold for p53 and 6.6-fold for MDM2 are detected by three biological replicates. (C) Two different digital sandwich immunoassays specifically detect the p53-MDM2 protein complex. Cell lysate and control samples were analyzed by distinct immunosandwich assays: one using printed p53 capture antibody, $\alpha p53_C$, with MDM2 detection antibody, $\alpha MDM2_D$, and the other using printed MDM2 capture antibody, $\alpha MDM2_C$, with p53 detection antibody, $\alpha p53_D$. The analysis of cell lysate provided a mass of signal above the levels of cross-reactivity determined from controls, indicating the presence of the p53-MDM2 complex.

CHAPTER FIVE: PROTEOMICS BY LINEAR MOTIF ANALYSIS AND PARALLEL PEPTIDE SEQUENCING

This chapter embodies methods and results that can be found in two patents (“Single Molecule Protein Screening,” PCT/US2009/066236, Mitra, RD, Tessler, LA, and “Methods of Polypeptide Identification and Compositions Thereof,” US11/674,642, Mitra, RD) as well as yet unpublished results. The *in silico* simulations for Dipeptide Motif Analysis were performed by Robi D. Mitra. The dipeptide antibody ELISAs, dipeptide correlation analysis, and the simulations for Parallel Peptide Sequencing were performed by me.

INTRODUCTION

Mass spectrometry and the immunoassay have been the workhorses of modern proteomics, but each technology has a clear profile of strengths as well as weaknesses, described in Chapter One. In short, mass spectrometry allows for the analysis of many proteins but lacks the sensitivity needed for most biomarkers. Immunoassays benefit from high sensitivity, but their multiplexing ability is limited even in microarray formats. Antibody microarrays have been mostly limited to analyzing dozens to hundreds of targets because they rely on an antibody (or pair of antibodies) for every target of interest. This “one-antibody/one-target problem” is a difficult issue for those interested in proteome-wide analysis, since it implies the need to make approximately 23,000 individual antibodies, validate them, and eliminate their cross reactivity. A technology with the sensitivity of an immunoassay but the multiplexed nature of mass spectrometry would be beneficial.

The concept of obtaining sequence information from individual molecules has spawned a revolution for DNA sequencing technologies, which perform DNA polymerase or DNA ligase reactions on individual molecules to obtain base pair information. We propose an analogous method that employs single-molecule detection to obtain sequence information from proteins. This method is based on the idea that individual amino acids or two amino acid (dipeptide) motifs may be detectable by binding a probe such as an antibody (or aptamer, or engineered protein). We also envision that such probes may be bound to targets, detected by TIRF imaging, and then removed by stringent washing to allow for subsequent rounds of binding. In this manner,

a small, “universal” set of antibodies (between 8 and 50) could be used to identify all human proteins.

We propose two major ways in which this technology may be implemented. In one embodiment of the technology, Dipeptide Motif Analysis, denatured protein molecules are analyzed for the presence or absence of dipeptides within their polypeptide sequence (Figure 20). In a second embodiment of the technology, Parallel Peptide Sequencing, sequential amino acid information is obtained from peptide digests of proteins (Figure 21). In both manifestations, polypeptides are analyzed by antibodies during multiple rounds of binding and washing, and the digital binding signatures for each molecule (enabled by single-molecule detection) reveal the unique identity of each molecule.

Here we present proof of concept data to support and inform the further development of Dipeptide Motif Analysis and Parallel Peptide Sequencing. The computational analyses, immunization protocols, and antibody validations shown here present hope for the development of high-throughput single-molecule protein detection.

DIPEPTIDE MOTIF ANALYSIS

Overview of the method

In Dipeptide Motif Analysis, denatured proteins are immobilized onto a surface and probed by antibodies in multiple rounds of binding (Figure 20). The antibodies are specific to common sequence motifs such as dipeptides. By analyzing the presence or absence of dipeptides within proteins and doing so at single-molecule resolution, the unique identity of the proteins as well as their abundances may be obtained.

Theoretical Performance

Dipeptide Motif Analysis relies on bioinformatic fact that short, linear amino acid sequences are moderately common throughout the proteome. For example, for many dipeptides, the chance of that dipeptide being present in any given protein is approximately 50% (Figure 22). For example, assuming that a complex mixture of proteins is denatured and immobilized on a surface with single-molecule separation, an antibody against a dipeptide motif (e.g. “proline-alanine”) should bind to ~50% of the proteins (assuming 100% binding occupancy). After the removal of that antibody by a stringent wash, a second antibody against a different motif (e.g. “serine-glutamate”) should bind to a different 50% of the proteins. After around 40 successive rounds of binding with different motif-recognizing antibodies, every molecule will have been probed and will reveal a binding signature that is unique to that protein.

This is demonstrated by an *in silico* experiment that shows the number of dipeptide antibody binding rounds that would be needed to uniquely identify all the proteins in the human genome (Figure 23). With 40 dipeptide antibody binding rounds, over 95% of the proteome can be uniquely identified. This *in silico* experiment also simulates the performance using an optimal set of antibodies compared to a randomized set. Those two sets behave similarly, probably due to the fact that the presence of dipeptide motifs throughout the genome is relatively random (Figure 24). Therefore, although 400 dipeptide motifs exist, with a set of only 40, one could, from an informatics perspective, perform a full proteome analysis.

After determining that the number of necessary dipeptide binding rounds is feasible, we wanted to investigate whether Dipeptide Motif Analysis would be robust to the type of errors expected in single-molecule experiments. For instance, 100% occupancy may not be possible (we reliably obtained 70-80% occupancy in our single molecule studies in Chapter Two). Additionally, non-specific binding could contribute to errors in the binding signatures. By simulation, we showed with a 2.5% false positive rate, the method can tolerate up to a 20% false negative rate. So, this method should be robust to noise, including false positive and false negative binding events (Figure 25).

Generation of dipeptide motif antibodies

With the theoretical framework established for Dipeptide Motif Analysis, we sought to investigate methods to generate and validate antibodies that could be used as dipeptide probes. Antibodies cannot be generated with *a priori* knowledge of their binding specificity. Instead, antibody recognition regions are generated by a biological process that is inherently stochastic. To complicate this phenomenon, polyclonal antibodies are comprised of a multitude of reactive species. Because of these confounding elements, polyclonal antibody specificity can only be determined empirically, by screening the antibody against a library of target molecules. As a result, no test for specificity is universal because specificity, as defined by such experiments, is only generalizable within the space of the test library.

Prior to deciding on dipeptide antibodies to generate, we sought to define the test library, which comprises of exemplar true and false targets for each antibody. We hypothesized that electrostatic forces would provide the strongest type of molecular

interaction that forms antibody specificity, therefore we limited the exemplar targets to those with specific electrostatic signatures. In this fashion we defined sixteen targets, comprised of the exhaustive combinations of lysine (K), arginine (R), glutamate (E), and aspartate (D) in dipeptide epitopes. These epitopes were synthesized to be internal to 9-mer peptides and flanked by randomized amino acids. In order to ensure that “accidental” exemplar dipeptides would not be formed, the amino acids flanking the dipeptide epitope did not contain amino acids that were used within the epitope. For example, this prevented the exemplar mixture X-X-K-R-X-X-G-G-C from being contaminated with X-R-K-R-X-X-G-G-C, which may have led to a false positive recognition by the RK antibody. These peptide mixtures were conjugated to carrier proteins and used in a rabbit immunization procedure (see Methods).

Purification of dipeptide motif antibodies

After obtaining polyclonal antisera to the sixteen dipeptides, we began assessing their specificity. We performed serum ELISAs for the KK, RR, EE, and DD antibodies against the KK, RR, EE and DD target peptides. We took this as a test case because we felt that if specificity was not possible between these cases, it would not be possible at all. At this stage, the cross-reactivity ELISAs demonstrated very little specificity, with only some specificity being produced from the KK antisera (Figure 26A).

We hypothesized that there may exist a high number of off-target antibodies within the antisera that was masking the specificity of the antibodies. We presumed this due to the complexity of the antigen (a highly randomized mixture containing only a small consensus epitope). We tested this by purifying the antibodies from the antisera and

subjecting them to multiple rounds of negative immuno-selection (see Methods). We performed this immunodepletion by running the antibodies through multiple columns containing peptide mixtures of the exemplar off-target dipeptide epitopes. Each antibody was passed through columns removed antibody species that reacted with all fifteen other dipeptide epitopes. Thus, after serial immunodepletion, the wash through from the columns should have been enriched for those antibodies that react only with the true exemplar dipeptide epitope.

After immunodepletion, the test set of KK, RR, EE, and DD antibodies showed improved specificity (Figure 26B). To test whether the entire set of antibodies would have any reactivity or specificity after purification and immunodepletion, I subjected them to ELISA tests against all sixteen dipeptide mixtures, each with a dilution series of antibody. All sixteen antibodies displayed reactivity in a concentration dependent manner and all showed at least some degree of specificity (exemplar curves in Figure 27). This demonstrated the success of the immunization approach and that specific antibody species remained within the polyclonal milieu after immunodepletion.

Analysis of antibody cross-reactivity

To analyze the global cross-reactivity all sixteen antibodies against the sixteen-dipeptide library, we used the highest concentration ELISA data points to create an antibody specificity heatmap (Figure 28). This heatmap demonstrates that the antibodies were highly specific in the epitope space of only two amino acids. Notably, most of the antibodies showed specificity with fewer than three side reactions.

The dipeptide antibodies achieved specificity to a startling degree. In one case particular, the antibodies were able to identify glutamate-aspartate (ED) as distinct from glutamate-glutamate (EE). This is impressive because these two dipeptides differ by only a single carbon bond. Additionally, the anti-ED antibody was able to identify *ED* as distinct from *DE*. This orientation specificity demonstrates the ability of the antibodies to recognize the side chains in the context of the amine bonds that form the peptide.

Some dipeptides displayed more cross-reactivity than others. For instance, those with lysine in the C-terminus tended to be more promiscuous in the second amino acid. To investigate the ability to improve specificity after the initial immunodepletion, we used the cross reactivity data to inform an additional immunodepletion step for an antibody. An ELISA after this second, targeted immunodepletion shows that, although affinity was reduced compared to the pre-depletion antibody, the cross reactivities were reduced (Figure 29). This is promising because it demonstrates our ability to tune the polyclonal antibody mixtures.

PARALLEL PEPTIDE SEQUENCING

Overview of the method

In Parallel Peptide Sequencing a protein sample is trypsin-digested, the resulting peptides are immobilized by cysteines, a small number of antibodies (or other similar probe) interrogates the N-termini the peptides, and then a chemical or enzymatic step removes the terminal residues (Figure 21). After cycling between antibody binding and cleavage, unique, identifying protein sequence information can be extracted. Then, by

counting the number of sequence reads mapping to each protein, each protein in a sample can be quantified.

Theoretical Performance

Parallel Peptide Sequencing relies on mapping sequence reads to a database of polypeptides in the genome. One practical question is, what read-lengths are needed to map back proteins reliably? This will dictate how many cycles of binding and cleavage will need to be carried out. I performed an *in silico* experiment by sampling terminal peptides from all of the proteins in the proteome (~23,000) given a particular read length n . Then, I asked how many of those reads were unique within the human proteome (Figure 30, blue).

By simulation, I found that with read lengths of eight amino acids and greater, one could map over 90% of the proteome uniquely. Moreover, for a unique 10-mer, the nearest peptide in the proteome is an average of 3.06 mismatches away. This means that a 10-mer can tolerate 1.5 mismatches within the read (15% error rate) and still be perfectly mappable. For a 15-mer and 20-mer, the error tolerance is 21% and 24% respectively. Taken together, this means that short reads can be highly mappable and robust to sequencing errors.

To test the dependence on the types of peptide digestion, I performed simulations of peptide digestion within the middle of proteins (30 residues and 60 residues internal) as well as with trypsin digestion. With random protein digestion, the mapping ability was slightly reduced. Here, a read length of eight could map 85% of the proteome (Figure 30, red and green), down from 90%. When using a trypsin digestion however, the mapping

ability was improved (Figure 30, blue). 8-mer reads from a trypsin digest could map 92% of the proteome, and 6-mers are sufficient to map 89%. The performance improvement by trypsin digestion is due to the fact that trypsin cuts at the C-terminal end of lysines and arginines. Because every sequence read must be terminal or preceded by a K or R, the information contained in each read is increased.

These simulations provide a feasible estimate for the number of cleavage rounds that would be needed. However, the simulations assumed all twenty amino acids were distinguishable from one another. Even though the data for dipeptide antibodies was promising, it may be difficult to produce probes to detect and distinguish all twenty amino acids since some amino acids are small and similar, such as glycine and alanine. So next, I performed a more conservative simulation that assumed that only *four types* of amino acids would be able to be distinguished: neutral hydrophobic, neutral hydrophilic, acidic, and basic. To conduct this simulation, all proteins in the proteome were converted into this four-letter sequence space.

The results of the *in silico* experiment using a quaternary alphabet show that twenty residues would need to be sequenced in order to map 90% proteins uniquely (Figure 30, orange). This is very promising for the possibility of Parallel Peptide Sequencing, since it greatly reduces the challenge of finding a sufficient number of terminal amino acid antibodies.

CONCLUSIONS

We have shown by simulation that proteomics by Dipeptide Motif Analysis and Parallel Peptide Sequencing are feasible in terms of determination of protein identity, the

required number of probe binding cycles, and robustness to sequencing error. For Dipeptide Motif Analysis, we have additionally shown the generation of dipeptide-specific antibodies to be feasible.

Hurdles remain for both technologies. For Dipeptide Motif Analysis, it was found to be difficult for the dipeptide antibodies to recognize peptides that were longer than the peptides used in the immunization and to recognize peptides not attached to carrier protein (data not shown). This inability to bind may be due to length dependence of the antigen recognition region or context dependence of the carrier molecule. The inability to bind may alternatively be due to secondary structure of peptides, but this is unlikely since exhaustive combinations of heat, SDS, Triton X-100, and urea could not improve binding. A valuable technological milestone would be to show that dipeptide antibodies could recognize their respective epitopes in many sequence contexts. New antibody generation techniques may be required for this goal[114, 115].

Smaller hurdles lie with Parallel Peptide Sequencing. A collaboration with Jim Havranek and Ben Borgo of the Havranek Lab has been fruitful for the development of proteins with engineered mutations that can bind to specific N-terminal amino acids (data not shown). What remains to be shown is the effective binding to peptides that are on a solid substrate. We have achieved success in conducting several types of cleavage reactions in bulk experiments including Edman Degradation, Barrett's Modification, and substrate-dependent cleavage by an engineered enzyme. What is left to be shown is the adaptation of one of these to a solid surface such as the nanogel surface.

MATERIALS AND METHODS

Producing polyclonal antisera against dipeptide epitopes

Semi-random peptide mixtures containing dipeptide sequence epitopes were generated by solid-phase peptide synthesis. The sequence of the peptide mixtures (from C to N termini) was *X-X-epitope-X-X-G-G-C*. Each mixture contained one of the sixteen following epitopes: KK, KR, KE, KD, RK, RR, RE, RD, DK, DR, DE, DD, and EK, ER, EE, ED. Positions labeled *X* denote positions in which equal amount of all twenty amino acids were used in the synthesis step. The G-G sequence was used as a spacer between the variable region of the peptides and the carrier protein, and the terminal cysteine acted as the linker for carrier conjugation. Each of the sixteen dipeptide mixtures was conjugated to carrier protein and used to generate antisera in rabbits. Two rabbits were used per dipeptide epitope protocol for assurance of immune response. All peptide synthesis and animal protocols were conducted by 21st Century Biochemicals, Inc. (Marlborough, MA).

Affinity purification

The sixteen antisera were subjected to affinity purification to separate the antibodies from serum proteins. Columns were prepared with resin coated in the targeted peptide mixture and antisera was bound, washed, then eluted. For example the KK antisera was bound, washed and eluted from a column packed with *X-X-K-K-X-X-G-G-C* resin.

Immunodepletion

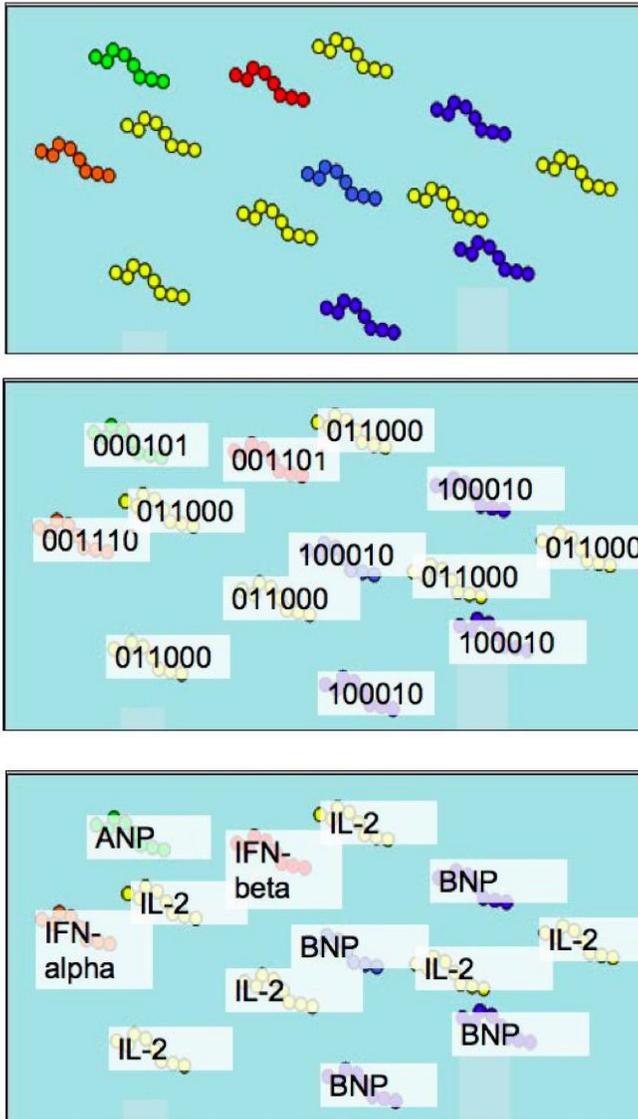
We synthesized sixteen negative selection (immunodepletion) peptide mixtures with which to remove off-target antibody binding and to maximize the specificity within this group of sixteen antibodies against the sixteen epitopes. These peptides were of the sequence *Z-Z-epitope-Z-Z-G-G-C*, where *epitope* denotes one of the sixteen dipeptide epitopes and *Z* denotes randomized flanking residues that excluded the amino acids K, R, E and D. For each affinity purified antibody, we depleted it against beads containing fifteen *other* dipeptide epitope negative selection peptides. These depletions were split up over four columns to provide enough exposure of all negative selection peptides.

ACKNOWLEDGEMENTS

We thank Jordan Fishman and 21st Century Biochemicals in Marlborough, MA for their expertise in custom antibody generation and their flexibility to work on challenging new peptide and antibody methods.

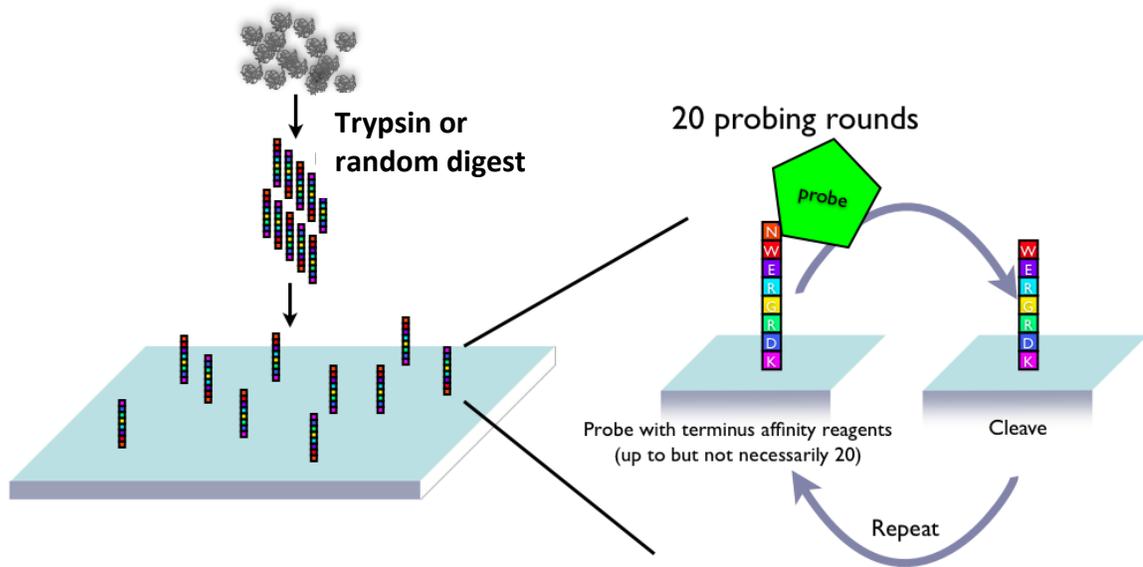
FIGURES

Figure 20



A schematic of digital proteomics by Dipeptide Motif Analysis. (A) Millions of denatured proteins are covalently attached to a substrate. (B) In successive rounds of binding and washing, a set of antibodies recognize dipeptide motifs internal to the proteins. Zeros indicate the absence of a motif and ones indicate the presence of a motif. (C) Binding signatures are mapped to unique polypeptide sequences in the proteome database.

Figure 21



A schematic of digital proteomics by Parallel Peptide Sequencing. A protein sample is digested and peptides are covalently attached to the analysis substrate. The N-termini of peptides are probed by rounds of probe binding and washing. After the terminal amino acid is identified, the terminus is removed by chemical or enzymatic cleavage, and the next amino acid is identified.

Figure 22

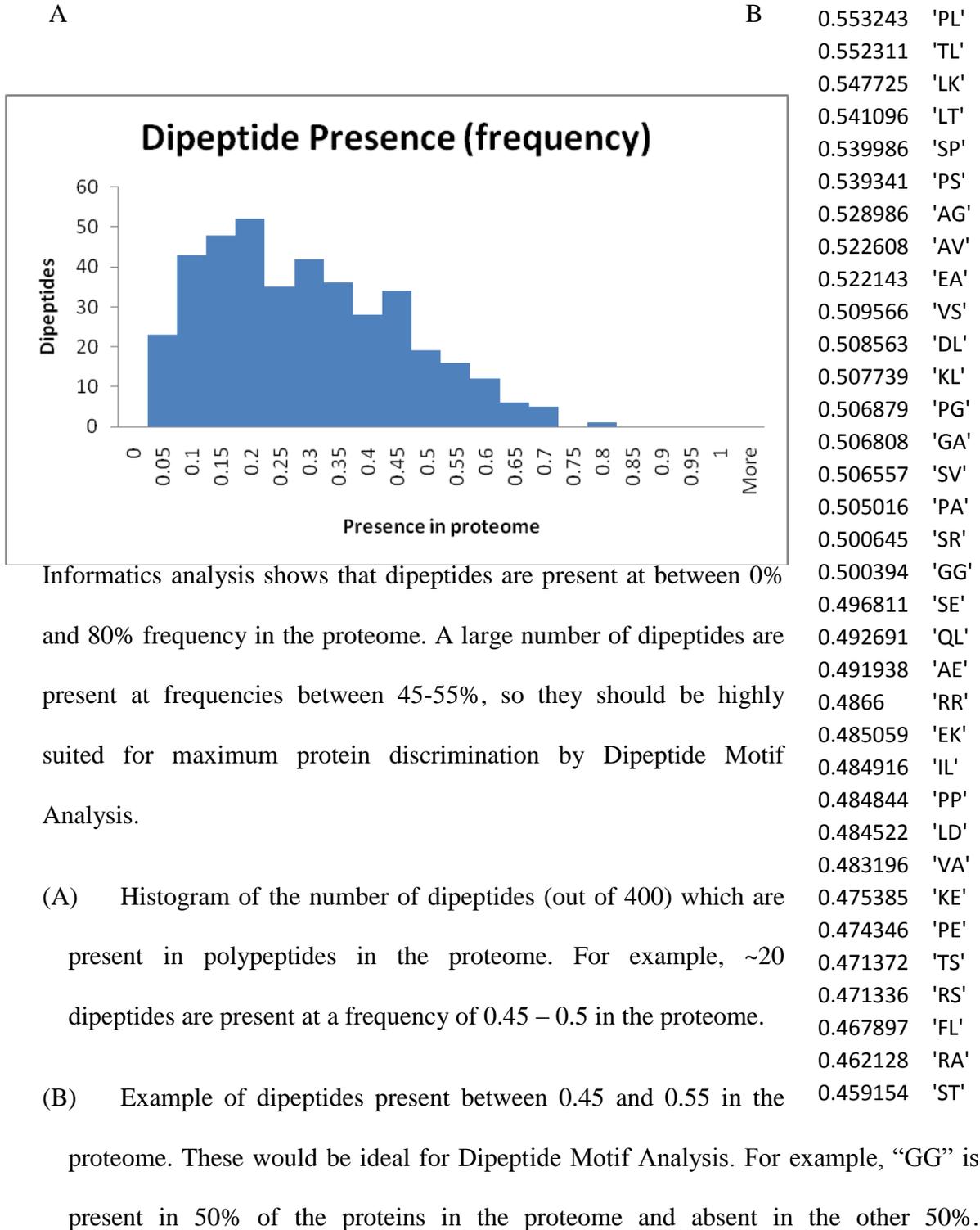
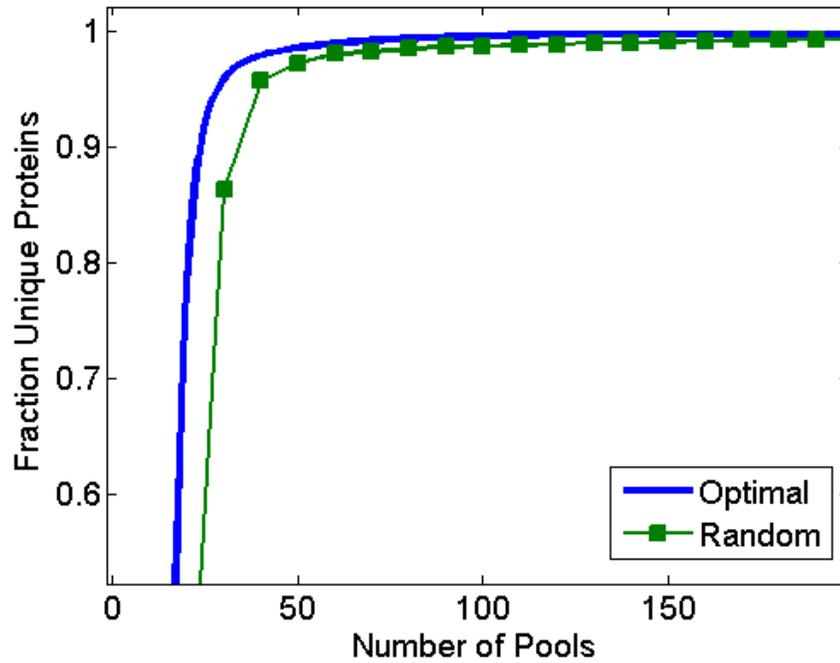
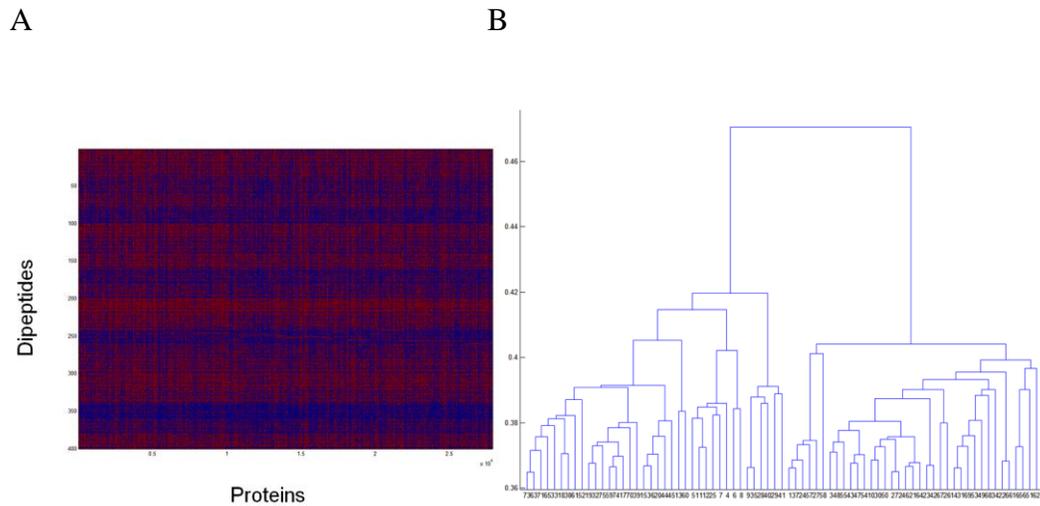


Figure 23



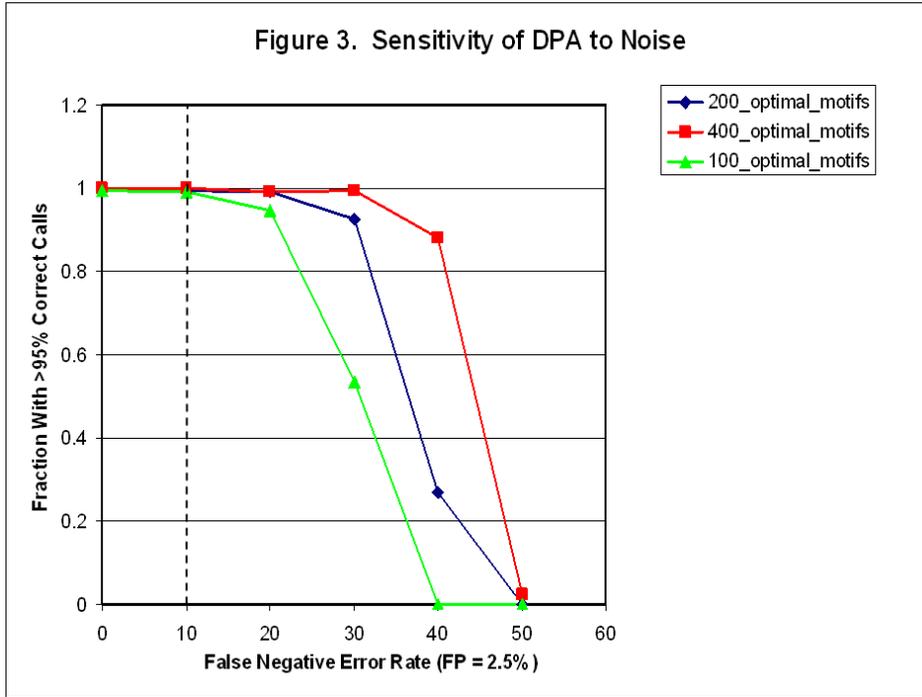
Simulation showing the fraction of proteins in the proteome that would have a unique dipeptide motif binding signature using an optimal pool of dipeptide probes (blue line) and a random pool of dipeptide probes (green line). (Mitra, unpublished)

Figure 24



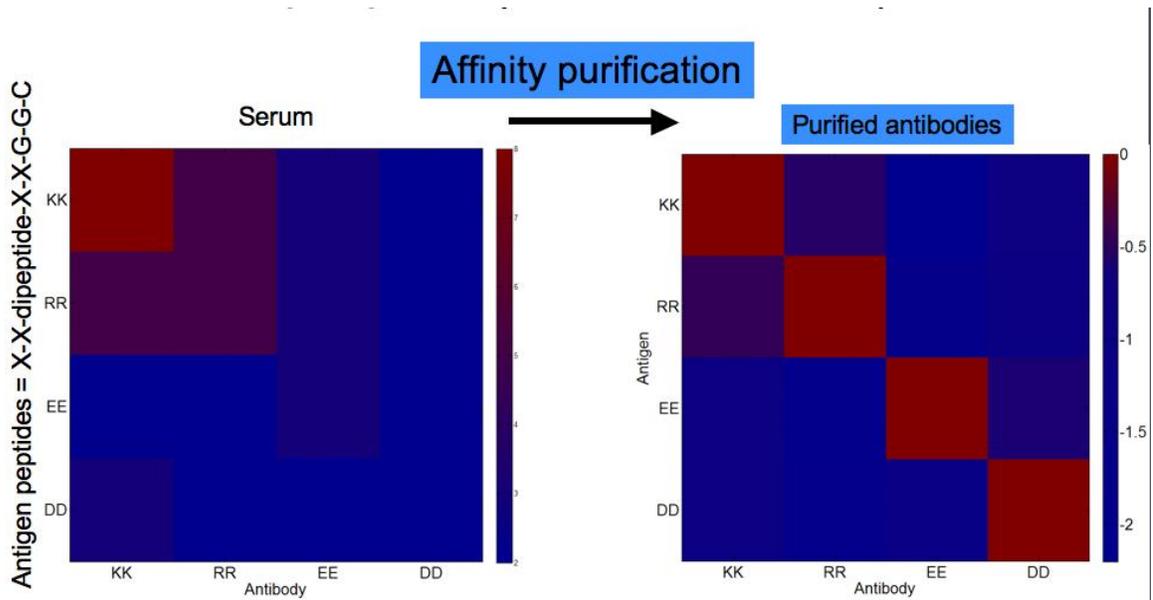
The presence of dipeptide motifs is mostly uniform throughout the proteome. A clustering of the presence (red) and absence (blue) of the 400 dipeptides across ~23,000 proteins shows that no two dipeptides co-occur substantially. The only co-occurrence is at the level of individual amino acids. For example, proteins that contain a tyrosine are more likely to contain tyrosine-containing dipeptides than proteins without tyrosine. (B) Hierarchical clustering based on presence/absence Hamming distance shows no substantial clustering.

Figure 25



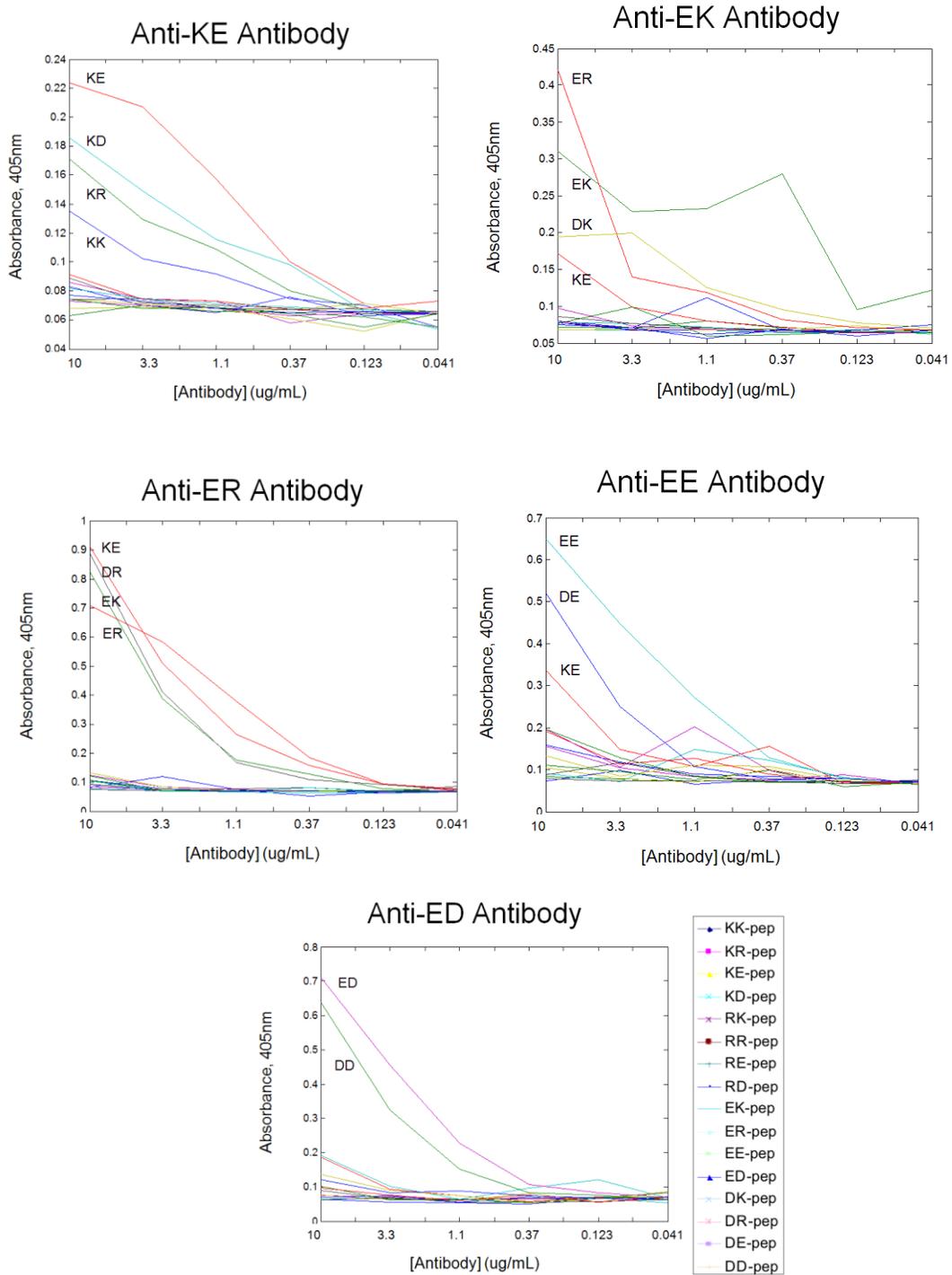
Simulation of unique mapping for Dipeptide Motif Analysis including a 2.5% false positive error rate and varying false negative error rates (Mitra, unpublished). Up to a 20% false negative rate is tolerable for a 100 motif binding assay.

Figure 26



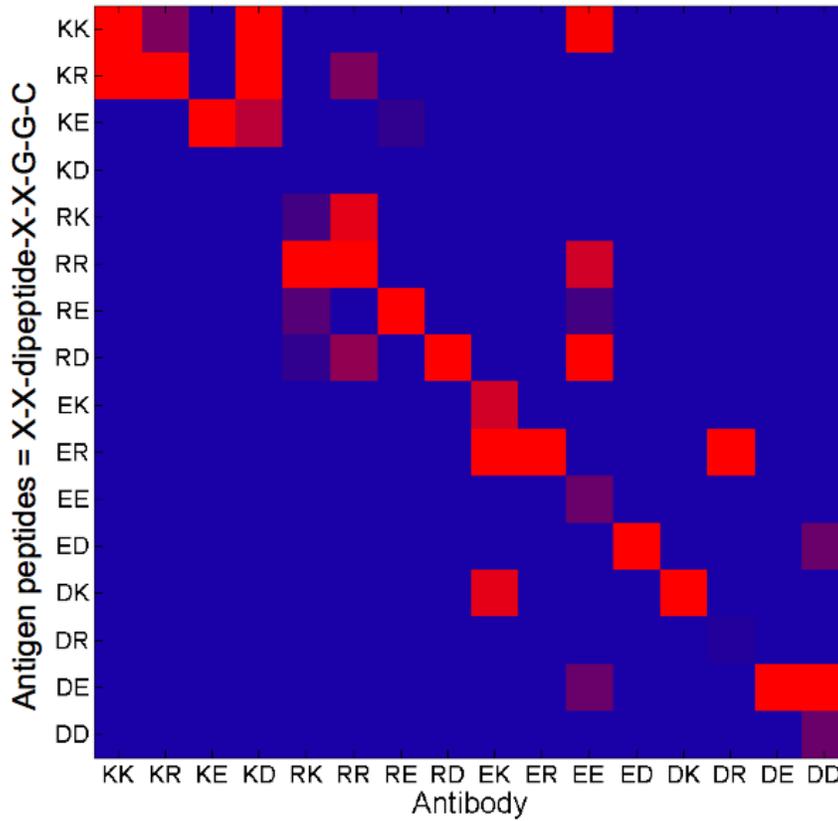
ELISAs performed on serum and on purified/immunodepleted antibodies for four test cases (KK, RR, EE, and DD motifs). Perfect specificity of antibodies for their respective antigens would be indicated by a descending diagonal line. Serum from immunized rabbits showed little specificity. After antibody purification and immunodepletion, specificity was achieved.

Figure 27



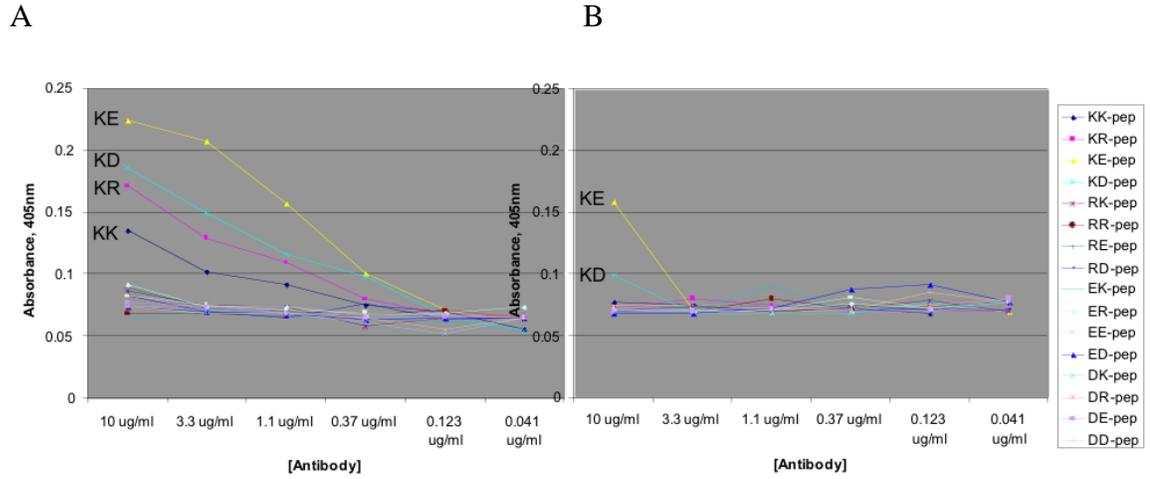
Five ELISA that were representative of the sixteen ELISAs to test specificity of the dipeptide antibodies against a panel of sixteen dipeptides.

Figure 28



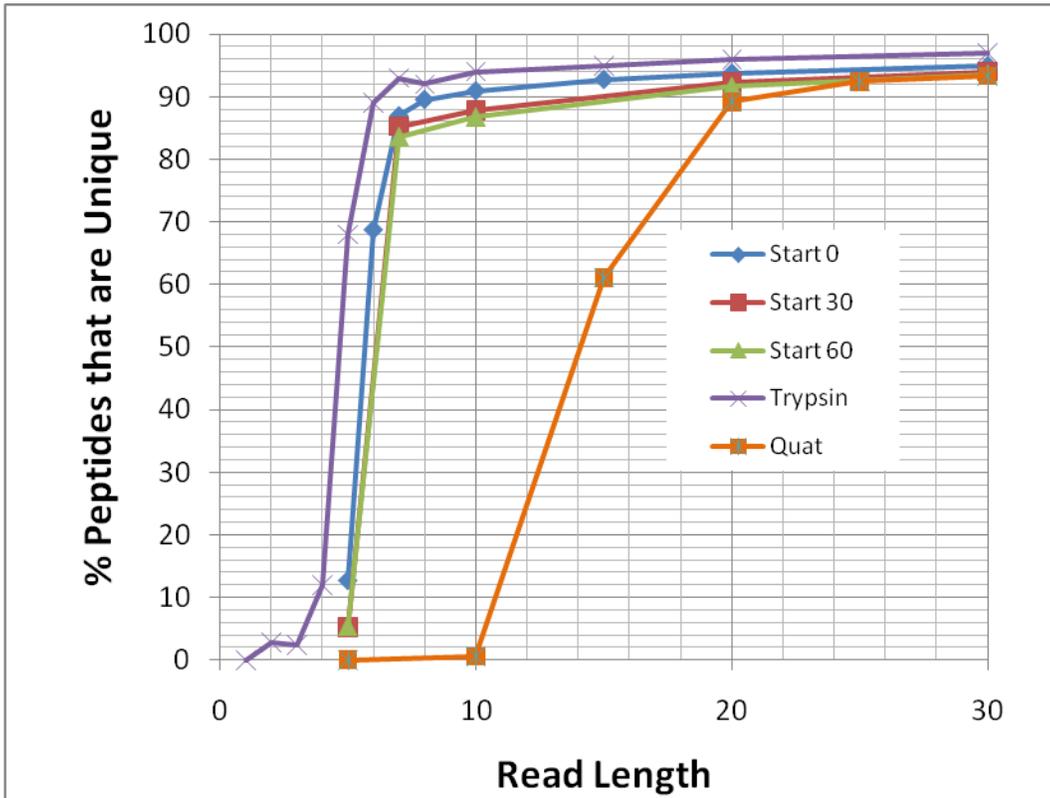
Summary of specificity of sixteen antibodies against sixteen dipeptides. Red boxes along the descending diagonal indicate perfect specificity, and red boxes off of the diagonal indicate cross-reactivity. Most antibodies showed cross reactivity with fewer than three peptides.

Figure 29



Specificity of the polyclonal antibodies is “tunable.” Cross-reactivity profile by ELISA for the Anti-KE antibody before and after a second round of immunodepletion. (A) Initial purification and immunodepletion yielded some cross-reactivity with KD, KR and KK. (B) After a second round of immunodepletion against KD, KR and KK, the cross-reactivity (as well as the affinity) was reduced.

Figure 30



Computer simulations of Parallel Peptide Sequencing show that only 8 residues need to be sequenced (with full, twenty amino acid resolution) in order to uniquely map reads to 90% of the human proteome (blue). With the amino acid sequence reduced to a quaternary alphabet, 20 residues need to be sequenced to uniquely map reads to 90% of the human proteome (orange). Sequencing within the middle of a protein reduced mappability slightly (red and green). Sequencing from trypsin-generated peptides improves mappability (purple). The remaining 10% of the proteome is difficult to map due to homology among small sets of proteins.

CHAPTER SIX: CONCLUSIONS AND FUTURE IMPLICATIONS

The biomedical diagnostics world is moving toward single-molecule detection faster than ever before. Before the initiation of this thesis, only one group had examined single-molecule detection in an immunoassay, and that was not in a chip-based/solid-phase format. Since the publication of our study in *Analytical Chemistry*, at least nine other groups have published on the subject with one of them being a venture capital funded company. I believe this trend will continue, that more academic and commercial laboratories will focus on single-molecule immunoassay applications, and that existing biotechnology companies will begin to adopt these techniques into their platform technologies. For instance, one DNA sequencing company, Pacific Biosciences, has begun exploring the translation of proteins with single ribosome resolution.

Single-molecule immunoassays are increasing in number, but there is still a dearth of science and engineering progress in the area of single amino acid and dipeptide recognition. Although two groups have written articles on prospective research in sequence and motif analysis [114, 115], only one group has published a molecule, an aptamer, that can detect short peptide motifs (tripeptides) [116]. Thus, I believe our lab and our university collaborators hold the scientific know-how, funding, and intellectual property rights to develop the proteomic technologies introduced in Chapter Five into research and commercial tools.

The findings in this thesis can be applied immediately to the quantification of proteins in cell lysate and serum. For experiments in which multiplexing is not necessary,

the direct immobilization method presented in Chapter Two provides an easy and sensitive way to monitor single-molecule antibody binding and could be used to monitor the *in vitro* binding of other interacting molecules. The nanogel surface coating characterized in Chapter Three should provide the biophysics community a way to conduct *in vitro* binding assays with higher occupancy and lower background. The single-molecule antibody microarray in Chapter Four will allow multiplexed experiments that are at least as sensitive as their analog counterparts, and with optimization, may be made more sensitive. Moreover, microarray facilities can easily adapt their printing methods to suit the nanogels protocol.

Protein analysis has proven to be a challenging and highly rewarding subject. This thesis has given me the chance to learn much in the areas of microscopy, surface chemistry, antibody development, and biochemistry. My enthusiasm for this line of research has only increased over my graduate career, and I hope this thesis inspires others to apply and develop protein analysis methods and single-molecule detection methods in biology and biomedicine.

REFERENCES

1. Mengel, M.B., W.L. Holleman, and S.A. Fields, *Fundamentals of clinical practice*. 2nd ed. 2002, New York: Kluwer Academic/Plenum Publishers. xxx, 837 p.
2. Zastrow, R.C., *Putting the Papanicolaou smear in perspective*. Arch Pathol Lab Med, 1997. **121**(3): p. 225-6.
3. Dinant, G.J., F.F. Buntinx, and C.C. Butler, *The necessary shift from diagnostic to prognostic research*. BMC Fam Pract, 2007. **8**: p. 53.
4. Cohn, B.A., G.A. Kaplan, and R.D. Cohen, *Did early detection and treatment contribute to the decline in ischemic heart disease mortality? Prospective evidence from the Alameda County Study*. Am J Epidemiol, 1988. **127**(6): p. 1143-54.
5. Waugh, N., et al., *Screening for type 2 diabetes: literature review and economic modelling*. Health Technol Assess, 2007. **11**(17): p. iii-iv, ix-xi, 1-125.
6. Greenwald, P., *A favorable view: progress in cancer prevention and screening*. Recent Results Cancer Res, 2007. **174**: p. 3-17.
7. Read, C., et al., *Early lung cancer: screening and detection*. Prim Care Respir J, 2006. **15**(6): p. 332-6.
8. Jemal, A., et al., *Annual report to the nation on the status of cancer, 1975-2005, featuring trends in lung cancer, tobacco use, and tobacco control*. J Natl Cancer Inst, 2008. **100**(23): p. 1672-94.
9. Stewart, S.L., et al., *Cancer mortality surveillance--United States, 1990-2000*. MMWR Surveill Summ, 2004. **53**(3): p. 1-108.
10. Ju, J.H., et al., *Changes in disease pattern and treatment outcome of colorectal cancer: a review of 5,474 cases in 20 years*. Int J Colorectal Dis, 2007. **22**(8): p. 855-62.
11. Omenn, G.S., *Strategies for plasma proteomic profiling of cancers*. Proteomics, 2006. **6**(20): p. 5662-73.
12. Jemal, A., et al., *Cancer statistics, 2009*. CA Cancer J Clin, 2009. **59**(4): p. 225-49.
13. Etzioni, R., et al., *The case for early detection*. Nat Rev Cancer, 2003. **3**(4): p. 243-52.

14. Petricoin, E.F., et al., *Clinical proteomics: Translating benchside promise into bedside reality*. Nature Reviews Drug Discovery, 2002. **1**(9): p. 683-695.
15. Anderson, L., *Candidate-based proteomics in the search for biomarkers of cardiovascular disease*. Journal of Physiology-London, 2005. **563**(1): p. 23-60.
16. Tirumalai, R.S., et al., *Characterization of the low molecular weight human serum proteome*. Mol Cell Proteomics, 2003. **2**(10): p. 1096-103.
17. Surinova, S., et al., *On the development of plasma protein biomarkers*. J Proteome Res. **10**(1): p. 5-16.
18. Archakov, A., et al., *Gene-centric view on the human proteome project: the example of the Russian roadmap for chromosome 18*. Proteomics: p. n/a-n/a.
19. Srinivas, P.R., et al., *Proteomics in early detection of cancer*. Clin Chem, 2001. **47**(10): p. 1901-1911.
20. Marvin, L.F., M.A. Roberts, and L.B. Fay, *Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry in clinical chemistry*. Clinica Chimica Acta, 2003. **337**(1-2): p. 11-21.
21. Picotti, P., et al., *Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics*. Cell, 2009. **138**(4): p. 795-806.
22. Diamandis, E.P., *Mass Spectrometry as a diagnostic and a cancer biomarker discovery tool - Opportunities and potential limitations*. Molecular & Cellular Proteomics, 2004. **3**(4): p. 367-378.
23. Leng, S.X., et al., *ELISA and multiplex technologies for cytokine measurement in inflammation and aging research*. J Gerontol A Biol Sci Med Sci, 2008. **63**(8): p. 879-84.
24. Macbeath, G., *Protein arrays: sandwich approach for analyzing complex solutions*. CSH Protoc, 2007. **2007**: p. pdb prot4633.
25. Fredriksson, S., et al., *Multiplexed protein detection by proximity ligation for cancer biomarker validation*. Nature Methods, 2007. **4**(4): p. 327-329.
26. Darmanis, S., et al., *Sensitive Plasma Protein Analysis by Microparticle-based Proximity Ligation Assays*. Molecular & Cellular Proteomics, 2010. **9**(2): p. 327-335.
27. Schwenk, J.M., et al., *Antibody suspension bead arrays within serum proteomics*. J Proteome Res, 2008. **7**(8): p. 3168-79.

28. Schwenk, J.M., et al., *Determination of binding specificities in highly multiplexed bead-based assays for antibody proteomics*. Mol Cell Proteomics, 2007. **6**(1): p. 125-32.
29. Burbulis, I., et al., *Quantifying small numbers of antibodies with a 'near-universal' protein-DNA chimera*. Nat Methods, 2007. **4**(12): p. 1011-3.
30. Walter, N.G., et al., *Do-it-yourself guide: how to use the modern single-molecule toolkit*. Nature Methods, 2008. **5**(6): p. 475-489.
31. Yang, W.D., J. Gelles, and S.M. Musser, *Imaging of single-molecule translocation through nuclear pore complexes*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(35): p. 12887-12892.
32. Kim, H.D., et al., *Mg²⁺-dependent conformational change of RNA studied by fluorescence correlation and FRET on immobilized single molecules*. Proc Natl Acad Sci U S A, 2002. **99**(7): p. 4284-9.
33. Conchello, J.A. and J.W. Lichtman, *Optical sectioning microscopy*. Nat Methods, 2005. **2**(12): p. 920-31.
34. Levene, M.J., et al., *Zero-mode waveguides for single-molecule analysis at high concentrations*. Science, 2003. **299**(5607): p. 682-686.
35. Gorris, H.H., T.M. Blicharz, and D.R. Walt, *Optical-fiber bundles*. Febs Journal, 2007. **274**(21): p. 5462-5470.
36. Sako, Y. and T. Yanagida, *Single-molecule visualization in cell biology*. Nat Rev Mol Cell Biol, 2003. **Suppl**: p. SS1-5.
37. Axelrod, D., *Total internal reflection fluorescence microscopy in cell biology*. Methods Enzymol, 2003. **361**: p. 1-33.
38. Mitra, R.D., et al., *Fluorescent in situ sequencing on polymerase colonies*. Anal Biochem, 2003. **320**(1): p. 55-65.
39. Shendure, J., et al., *Accurate multiplex polony sequencing of an evolved bacterial genome*. Science, 2005. **309**(5741): p. 1728-32.
40. *Human genome: Genomes by the thousand*. Nature. **467**(7319): p. 1026-7.
41. Wold, B. and R.M. Myers, *Sequence census methods for functional genomics*. Nature Methods, 2008. **5**(1): p. 19-21.
42. Marioni, J.C., et al., *RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays*. Genome Res, 2008. **18**(9): p. 1509-17.

43. Scott, E.A., et al., *Protein adsorption and cell adhesion on nanoscale bioactive coatings formed from poly(ethylene glycol) and albumin microgels*. *Biomaterials*, 2008. **29**(34): p. 4481-93.
44. Mattheyses, A.L., S.M. Simon, and J.Z. Rappoport, *Imaging with total internal reflection fluorescence microscopy for the cell biologist*. *J Cell Sci*. **123**(Pt 21): p. 3621-8.
45. Ghaemmaghami, S., et al., *Global analysis of protein expression in yeast*. *Nature*, 2003. **425**(6959): p. 737-741.
46. Cohen, A.A., et al., *Dynamic Proteomics of Individual Cancer Cells in Response to a Drug*. *Science*, 2008. **322**(5907): p. 1511-1516.
47. Omenn, G.S., et al., *Overview of the HUPO Plasma Proteome Project: Results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database*. *Proteomics*, 2005. **5**(13): p. 3226-3245.
48. Harris, T.D., et al., *Single-molecule DNA sequencing of a viral genome*. *Science*, 2008. **320**(5872): p. 106-109.
49. Roy, R., S. Hohng, and T. Ha, *A practical guide to single-molecule FRET*. *Nature Methods*, 2008. **5**(6): p. 507-516.
50. Sauer, M., et al., *Detection and identification of individual antigen molecules in human serum with pulsed semiconductor lasers*. *Applied Physics B-Lasers and Optics*, 1997. **65**(3): p. 427-431.
51. Li, L., et al., *Quantitative counting of single fluorescent molecules by combined electrochemical adsorption accumulation and total internal reflection fluorescence microscopy*. *Anal Chem*, 2008. **80**(11): p. 3999-4006.
52. Loscher, F., et al., *Counting of single protein molecules at interfaces and application of this technique in early-stage diagnosis*. *Analytical Chemistry*, 1998. **70**(15): p. 3202-3205.
53. Moul, J.W., *Population screening for prostate cancer and emerging concepts for young men*. *Clin Prostate Cancer*, 2003. **2**(2): p. 87-97.
54. Munkarah, A., M. Chatterjee, and M.A. Tainsky, *Update on ovarian cancer screening*. *Curr Opin Obstet Gynecol*, 2007. **19**(1): p. 22-6.
55. Wallrabe, H. and A. Periasamy, *Imaging protein molecules using FRET and FLIM microscopy*. *Current Opinion in Biotechnology*, 2005. **16**(1): p. 19-27.
56. Gunderson, K.L., et al., *Decoding randomly ordered DNA arrays*. *Genome Research*, 2004. **14**(5): p. 870-877.

57. Berglund, L., et al., *A Genecentric Human Protein Atlas for Expression Profiles Based on Antibodies*. *Molecular & Cellular Proteomics*, 2008. **7**(10): p. 2019-2027.
58. Lange, K. and M. Rapp, *Influence of intermediate aminodextran layers on the signal response of surface acoustic wave biosensors*. *Anal Biochem*, 2008. **377**(2): p. 170-5.
59. Jung, S., et al., *Biotin-functionalized cellulose-based monolayers as sensitive interfaces for the detection of single molecules*. *Chembiochem*, 2006. **7**(6): p. 900-3.
60. Heyes, C.D., et al., *Biocompatible surfaces for specific tethering of individual protein molecules*. *Journal of Physical Chemistry B*, 2004. **108**(35): p. 13387-13394.
61. Sofia, S.J., V. Premnath, and E.W. Merrill, *Poly(ethylene oxide) grafted to silicon surfaces: Grafting density and protein adsorption*. *Macromolecules*, 1998. **31**(15): p. 5059-5070.
62. Akkoyun, A. and U. Bilitewski, *Optimisation of glass surfaces for optical immunosensors*. *Biosensors & Bioelectronics*, 2002. **17**(8): p. 655-664.
63. Piehler, J., et al., *Surface modification for direct immunoprobes*. *Biosensors & Bioelectronics*, 1996. **11**(6-7): p. 579-590.
64. Piehler, J., et al., *A high-density poly(ethylene glycol) polymer brush for immobilization on glass-type surfaces*. *Biosensors & Bioelectronics*, 2000. **15**(9-10): p. 473-481.
65. Taylor, S., et al., *Impact of surface chemistry and blocking strategies on DNA microarrays*. *Nucleic Acids Research*, 2003. **31**(16): p. e78.
66. Masson, J.F., et al., *Nondestructive monitoring of the photochromic state of dithienylethene monolayers by surface plasmon resonance*. *Langmuir*, 2005. **21**(16): p. 7413-7420.
67. Lange, K. and M. Rapp, *Influence of intermediate aminodextran layers on the signal response of surface acoustic wave biosensors*. *Analytical Biochemistry*, 2008. **377**(2): p. 170-175.
68. Yanker, D.M. and J.A. Maurer, *Direct printing of trichlorosilanes on glass for selective protein adsorption and cell growth*. *Mol Biosyst*, 2008. **4**(6): p. 502-4.
69. Agrawal, A., et al., *Nanometer-scale mapping and single-molecule detection with color-coded nanoparticle probes*. *Proc Natl Acad Sci U S A*, 2008. **105**(9): p. 3298-3303.

70. Li, H.T., et al., *Molecule by molecule direct and quantitative counting of antibody-protein complexes in solution*. Analytical Chemistry, 2004. **76**(15): p. 4446-4451.
71. Kingsmore, S.F., *Multiplexed protein measurement: technologies and applications of protein and antibody arrays*. Nature Reviews Drug Discovery, 2006. **5**(4): p. 310-320.
72. Braslavsky, I., et al., *Sequence information can be obtained from single DNA molecules*. Proc Natl Acad Sci U S A, 2003. **100**(7): p. 3960-4.
73. Heyes, C.D., et al., *Synthesis, patterning and applications of star-shaped poly(ethylene glycol) biofunctionalized surfaces*. Molecular Biosystems, 2007. **3**(6): p. 419-430.
74. Mitra, R.D., et al., *Fluorescent in situ sequencing on polymerase colonies*. Analytical Biochemistry, 2003. **320**(1): p. 55-65.
75. Yeung, Y.G. and E.R. Stanley, *A solution for stripping antibodies from polyvinylidene fluoride immunoblots for multiple reprobings*. Anal Biochem, 2009. **389**(1): p. 89-91.
76. Tessler, L.A., J.G. Reifengerger, and R.D. Mitra, *Protein quantification in complex mixtures by solid phase single-molecule counting*. Anal Chem, 2009. **81**(17): p. 7141-8.
77. Rasnik, I., S.A. McKinney, and T. Ha, *Nonblinking and longlasting single-molecule fluorescence imaging*. Nature Methods, 2006. **3**(11): p. 891-893.
78. Mikhailenko, S.V., Y. Oguchi, and S. Ishiwata, *Insights into the mechanisms of myosin and kinesin molecular motors from the single-molecule unbinding force measurements*. J R Soc Interface. **7 Suppl 3**: p. S295-306.
79. Groll, J., et al., *Biofunctionalized, ultrathin coatings of cross-linked star-shaped poly(ethylene oxide) allow reversible folding of immobilized proteins*. Journal of the American Chemical Society, 2004. **126**(13): p. 4234-4239.
80. Wacker, B.K., et al., *Delivery of sphingosine 1-phosphate from poly(ethylene glycol) hydrogels*. Biomacromolecules, 2006. **7**(4): p. 1335-43.
81. Wayment, J.R. and J.M. Harris, *Biotin-avidin binding kinetics measured by single-molecule imaging*. Anal Chem, 2009. **81**(1): p. 336-42.
82. Roy, R., et al., *SSB protein diffusion on single-stranded DNA stimulates RecA filament formation*. Nature, 2009. **461**(7267): p. 1092-7.
83. Wang, H., Y.S. Yeh, and P.F. Barbara, *HIV-1 nucleocapsid protein bends double-stranded nucleic acids*. J Am Chem Soc, 2009. **131**(42): p. 15534-43.

84. Wu, J.Y., M.D. Stone, and X. Zhuang, *A single-molecule assay for telomerase structure-function analysis*. *Nucleic Acids Res*, 2009. **38**(3): p. e16.
85. Deniz, A.A., S. Mukhopadhyay, and E.A. Lemke, *Single-molecule biophysics: at the interface of biology, physics and chemistry*. *J R Soc Interface*, 2008. **5**(18): p. 15-45.
86. Reifengerger, J.G., et al., *Myosin VI undergoes a 180 degrees power stroke implying an uncoupling of the front lever arm*. *Proc Natl Acad Sci U S A*, 2009. **106**(43): p. 18255-60.
87. Kapranov, P., et al., *New class of gene-termini-associated human RNAs suggests a novel RNA copying mechanism*. *Nature*. **466**(7306): p. 642-6.
88. Flusberg, B.A., et al., *Direct detection of DNA methylation during single-molecule, real-time sequencing*. *Nat Methods*. **7**(6): p. 461-5.
89. Elbert, D.L. and J.A. Hubbell, *Conjugate addition reactions combined with free-radical cross-linking for the design of materials for tissue engineering*. *Biomacromolecules*, 2001. **2**(2): p. 430-41.
90. Todd, J., et al., *Ultrasensitive flow-based immunoassays using single-molecule counting*. *Clin Chem*, 2007. **53**(11): p. 1990-5.
91. Rissin, D.M., et al., *Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations*. *Nature Biotechnology*, 2010. **28**(6): p. 595-U25.
92. Li, L., et al., *Single-molecule-counting protein microarray assay with nanoliter samples and its application in the dynamic protein expression of living cells*. *Biosensors & Bioelectronics*, 2011. **26**(8): p. 3688-3691.
93. Lee, S., et al., *An ultra-sensitive nanoarray chip based on single-molecule sandwich immunoassay and TIRFM for protein detection in biologic fluids*. *Analyst*, 2009. **134**(5): p. 933-8.
94. Tessler, L.A., et al., *Nanogel surface coatings for improved single-molecule imaging substrates*. *J R Soc Interface*, (Epub Feb 16, 2011).
95. Schmidt, R., et al., *Single-Molecule Detection on a Protein-Array Assay Platform for the Exposure of a Tuberculosis Antigen*. *J Proteome Res*. **10**(3): p. 1316-1322.
96. Haab, B.B., et al., *Immunoassay and antibody microarray analysis of the HUPO Plasma Proteome Project reference specimens: systematic variation between sample types and calibration of mass spectrometry data*. *Proteomics*, 2005. **5**(13): p. 3278-91.

97. Ingvarsson, J., et al., *Detection of pancreatic cancer using antibody microarray-based serum protein profiling*. Proteomics, 2008. **8**(11): p. 2211-2219.
98. Kopf, E., D. Shnitzer, and D. Zharhary, *Panorama Ab Microarray Cell Signaling kit: a unique tool for protein expression analysis*. Proteomics, 2005. **5**(9): p. 2412-6.
99. Nielsen, U.B. and B.H. Geierstanger, *Multiplexed sandwich assays in microarray format*. J Immunol Methods, 2004. **290**(1-2): p. 107-20.
100. Yu, X., N. Schneiderhan-Marra, and T.O. Joos, *Protein microarrays for personalized medicine*. Clin Chem. **56**(3): p. 376-87.
101. Bergsma, D., et al., *Antibody-array interaction mapping, a new method to detect protein complexes applied to the discovery and study of serum amyloid P interactions with kininogen in human plasma*. Mol Cell Proteomics. **9**(3): p. 446-56.
102. Hurst, R., et al., *Protein-protein interaction studies on protein arrays: effect of detection strategies on signal-to-background ratios*. Anal Biochem, 2009. **392**(1): p. 45-53.
103. Macbeath, G., *Protein arrays: labeling the protein and probing the array for protein-protein interactions*. CSH Protoc, 2007. **2007**: p. pdb prot4630.
104. Mattoon, D.R. and B. Schweitzer, *Profiling protein interaction networks with functional protein microarrays*. Methods Mol Biol, 2009. **563**: p. 63-74.
105. Shoemaker, B.A. and A.R. Panchenko, *Deciphering protein-protein interactions. Part I. Experimental techniques and databases*. PLoS Comput Biol, 2007. **3**(3): p. e42.
106. Riley, T., et al., *Transcriptional control of human p53-regulated genes*. Nature Reviews Molecular Cell Biology, 2008. **9**(5): p. 402-412.
107. Wei, C.L., et al., *A global map of p53 transcription-factor binding sites in the human genome*. Cell, 2006. **124**(1): p. 207-219.
108. Sun, X.X., M.S. Dai, and H. Lu, *5-fluorouracil activation of p53 involves an MDM2-ribosomal protein interaction*. J Biol Chem, 2007. **282**(11): p. 8052-9.
109. Moll, U.M. and O. Petrenko, *The MDM2-p53 interaction*. Mol Cancer Res, 2003. **1**(14): p. 1001-8.
110. Momand, J., et al., *The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation*. Cell, 1992. **69**(7): p. 1237-45.

111. Kingsmore, S.F., *Multiplexed protein measurement: technologies and applications of protein and antibody arrays*. Nat Rev Drug Discov, 2006. **5**(4): p. 310-20.
112. Pang, S., et al., *A comparability study of the emerging protein array platforms with established ELISA procedures*. J Immunol Methods, 2005. **302**(1-2): p. 1-12.
113. Poetz, O., et al., *Protein microarrays for antibody profiling: specificity and affinity determination on a chip*. Proteomics, 2005. **5**(9): p. 2402-11.
114. Poetz, O., et al., *Proteome wide screening using peptide affinity capture*. Proteomics, 2009. **9**(6): p. 1518-1523.
115. Wingren, C., P. James, and C.A.K. Borrebaeck, *Strategy for surveying the proteome using affinity proteomics and mass spectrometry*. Proteomics, 2009. **9**(6): p. 1511-1517.
116. Niu, W.Z., N. Jiang, and Y.H. Hu, *Detection of proteins based on amino acid sequences by multiple aptamers against tripeptides*. Analytical Biochemistry, 2007. **362**(1): p. 126-135.
117. Hornbeck, P., *Enzyme-linked immunosorbent assays*. Curr Protoc Immunol, 2001. **Chapter 2**: p. Unit 2 1.
118. Tessler, L.A., et al., *Nanogel surface coatings for improved single-molecule imaging substrates*. J R Soc Interface, (Epub Feb 18, 2011).
119. Vallania, F.L., et al., *High-throughput discovery of rare insertions and deletions in large cohorts*. Genome Res. **20**(12): p. 1711-8.

APPENDIX ONE: SUPPLEMENTAL INFORMATION

Supporting Information

for

“Protein quantification in complex mixtures by solid phase

single-molecule counting”

Analyzing Images by Iterative Thresholding

We developed software to determine the positions of single fluorescent molecules that overcomes the limitations of the nonuniform illumination region inherent to the Nikon TIRF Illumination. We created a Metamorph Journal that reads in a 12-bit TIFF and acts as follows. We iterated over intensity thresholds from 0 to 4,050 by increments of 50. For each intensity threshold, we defined objects as sets of pixels that 1) have intensity values are greater than the threshold, 2) are contiguous with other pixels within that object and 3) comprise an area between 2 and 20 pixels. We found the limits of 2 and 20 to give the best precision. Then for each object found at each intensity threshold, we outputted the X-Y locations of the object centroids. We imported this set of 2-D points into MATLAB. Then we created a binary output image, of ones and zeros, which contained a 3 x 3 pixel square of ones centered on each of the 2-D points. These binary

squares were used to represent the locations of fluorescent molecules for analyzing their positions and abundance.

To measure sensitivity and specificity of iterative thresholding (IT), we obtained several raw images of Cy5 antibody molecules, analyzed them by IT and by single value thresholding (SVT), and outputted analyzed versions using each method. To obtain sensitivity, we measured the fraction of objects present in the SVT image that were also present in the IT image. To obtain specificity, we measured the fraction of objects present in the IT image that were present in the raw image. False positives with respect to the raw image were identifiable by the lack of likeness to a Gaussian point-spread function.

Measuring Protein Detection Efficiency

For analyzing the dual-color assays, the acquired Cy5 image (antibody) and the corresponding Cy3 image (protein) were processed by the IT algorithm into binary images. Then, we chose a 316 pixel x 316 pixel (31.6 μm x 31.6 μm) region from the Cy5 and Cy3 images to calculate the antibody-protein correlation. The correlation between red and green channels was calculated as follows (Equation 1):

$$C = \sum_{i=1}^{316} \sum_{j=1}^{316} R_{i,j} G_{i,j} \quad (1)$$

Here, R and G are matrices of ones and zeros, representing the binary image of size 316 x 316 pixel². To correct for stage shifting, we allowed the Cy5-Cy3 image pair to be offset with respect to each other in order to find the alignment that produced the maximum correlation (the true alignment). Once the true alignment was found, our software counted

the number of proteins that overlapped antibodies and divided that by the total number of proteins. We defined this ratio as the fractional accessibility or binding efficiency.

Testing for Random Correlations and Specificity of Binding

To determine whether the Cy5-Cy3 image correlation was a random event, we compared the correlation derived from Equation 1 to the background distribution of correlations for Cy5-Cy3 image pairs that were offset in the X and Y directions. To do this, we chose a 316 x 316 pixel² region from the Cy3 image as the “base”. We then computed the correlation of the base with regions of the Cy5 image that were misaligned by a translational offset. We scanned offsets between -100 to +100 pixels (with respect to the true alignment) in both the X and Y directions. After these 40,000 correlations were computed, they were plotted as a function of the X and Y offsets. To interpret the Z-axis of the correlograms, level surfaces correspond to the background distribution of correlation values, and peaks correspond to correlations that are significantly nonrandom. In binding experiments in which the antibody was specific for its target (Paper Figure 2), a high peak was seen around the offset (0, 0). Therefore correlation for the true alignment was nonrandom, and binding was specific. By contrast, when a nonmatching protein target was used, no peak appeared (Figure S3), indicating randomness between Cy3 and Cy5 channels (and no specific binding).

Estimating Nonspecific Binding Based on the Observed Antibody Density

We used the total number of antibodies remaining on the surface after washing to estimate the frequency of antibody-ligand correlations that occurred merely by chance overlap of molecules – the false positive (FP) rate. We define the FP rate as the probability that a randomly chosen pixel will be within a radius 2.5 pixels from an antibody pixel. This probability follows a Poisson process, where the parameter lambda is the frequency of antibody pixels out of the total number of pixels. Therefore, for the number of antibodies on the surface A , and total pixel area of the image T ,

$$FP = 1 - \exp\left(-\frac{2.5^2 \pi A}{T}\right) \quad (2)$$

Quantifying Total IgG in Rabbit Serum by ELISA

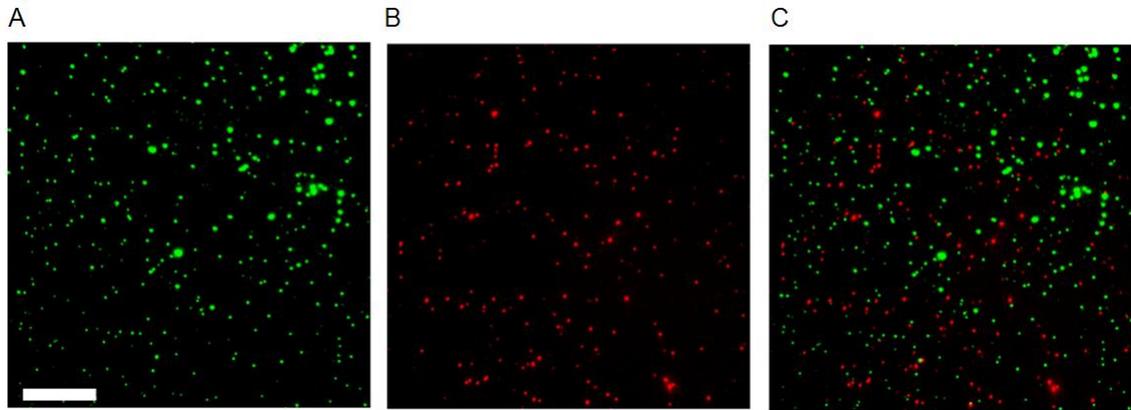
The three rabbit serum samples (see Methods) were used as coating antigens. The detection antibody was polyclonal anti-rabbit antibody conjugated to alkaline-phosphatase (Abcam, Cambridge, MA). Polystyrene microtiter plates (Immulon 2HB) were obtained from Thermo Fisher Scientific (Waltham, MA). Washes were performed using Labsystems Multidrop 384 (Beckman Coulter, Fullerton, CA). Detection of the fluorogenic substrate, (4-methylumbelliferyl phosphate, Sigma Aldrich, St. Louis) was performed on the microtiter plate fluorimeter Synergy HT (Biotek, Winooski, VT).

An indirect enzyme-linked immunosorbent assay (ELISA) was performed as described.[117] For each serum sample, two dimensional titrations were performed to determine the optimal dilutions of coating antigen and detection antibody. A dynamic

range of detection that spanned the signals of all three serum samples was achieved using the following dilutions. Coating antigens – 1:64,000. Detection antibody – 625 ng/ml.

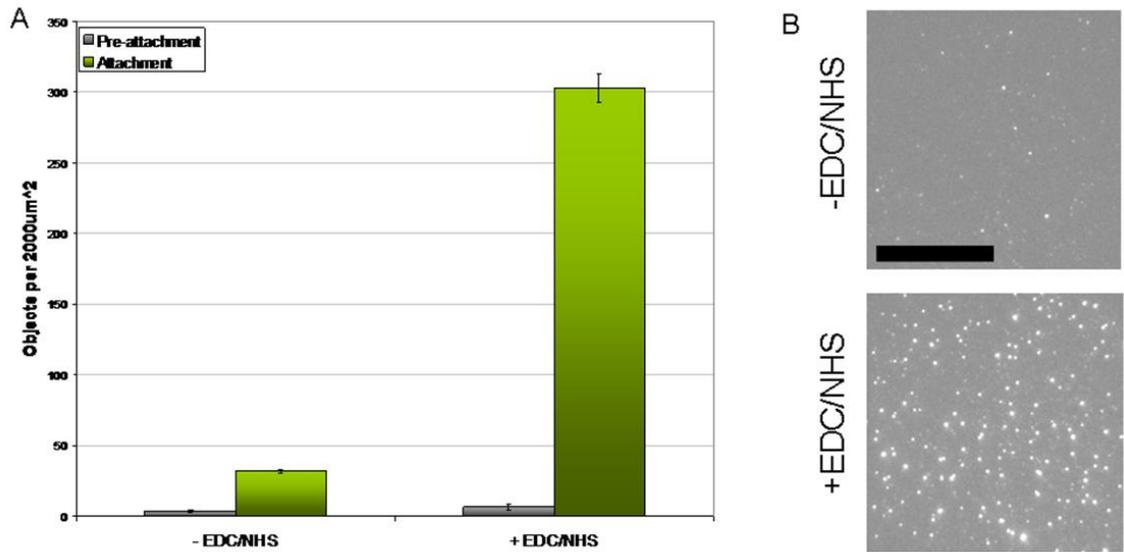
SUPPLEMENTAL FIGURES

Figure S1.



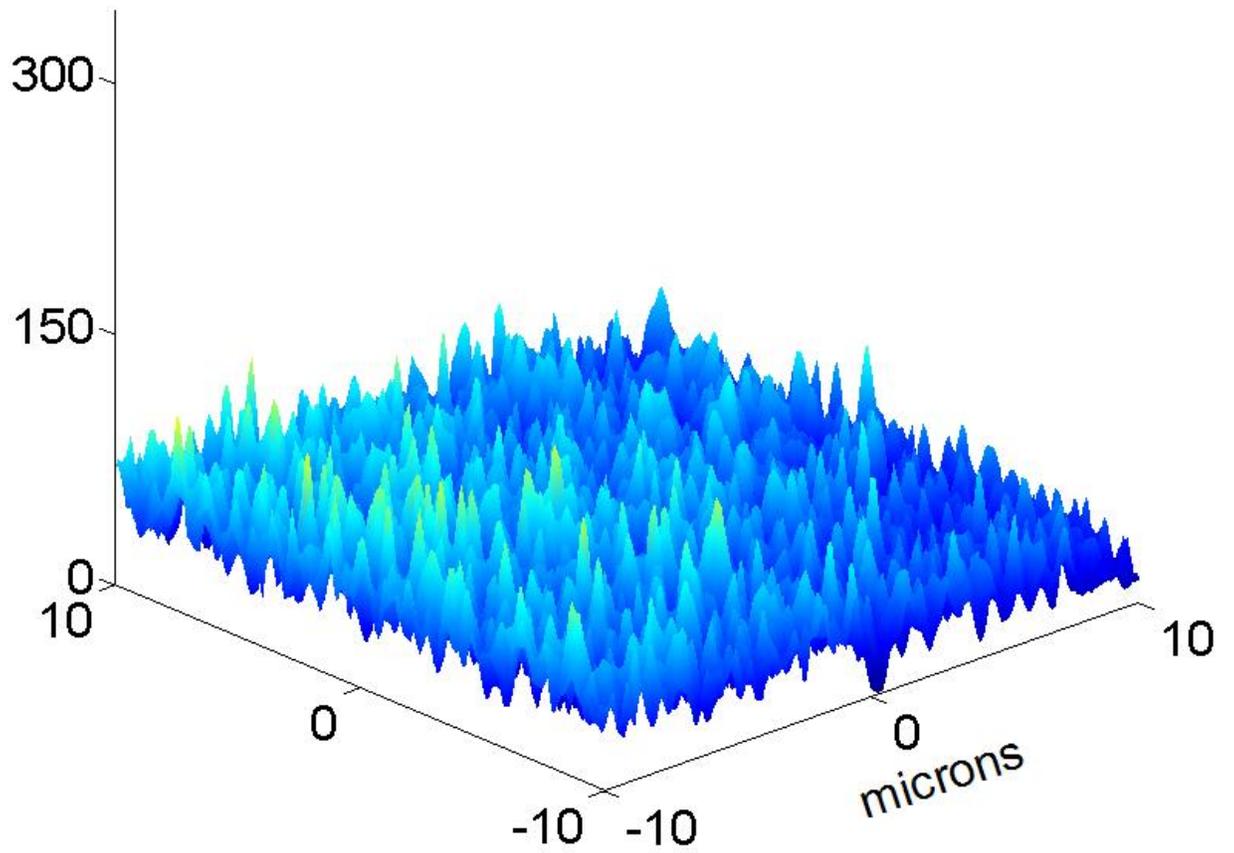
Demonstrating single-antibody detection. To test whether fluorescence objects were in fact single molecules, we mixed antibodies of two different colors together and quantified number of instances in which we observed two overlapping objects on the surface. We observe no significant overlap ($p = 0.73$ Fisher's Exact Test) between antibodies labeled with (A) Cy3 or (B) Cy5 when (C) images are merged (scale bar = 10 μm). This indicates that each fluorescence object represents a single antibody molecule.

Figure S2



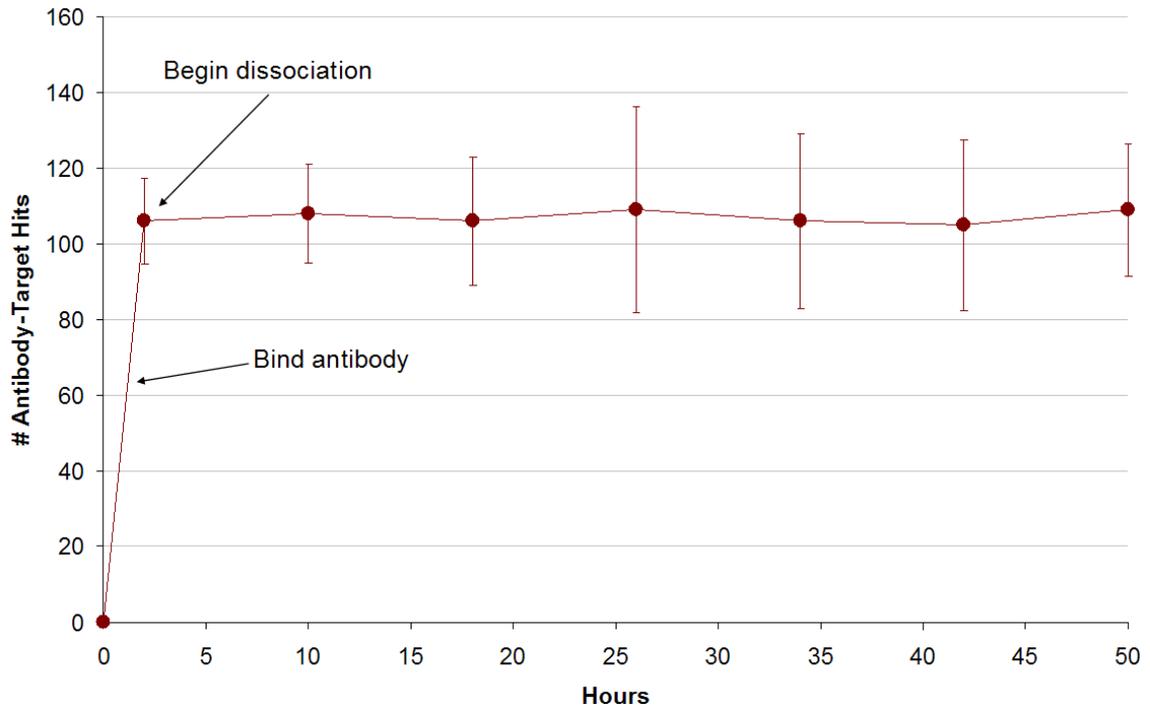
Attachment efficiency. The EDC/NHS heterobifunctional crosslinking system can effectively activate BSA molecules on the surface to immobilize target proteins. (A) The number of protein molecules attached to the surface per 2,000 μm^2 with and without EDC/NHS surface activation. (B) Images of protein molecules attached to the surface (top) without EDC/NHS surface activation and (bottom) with EDC/NHS (scale bar = 10 μm).

Figure S3



Negative control for binding. When we perform protein detection using a nonspecific target protein, the correlogram analysis shows a random distribution of correlations, indicating no specific binding.

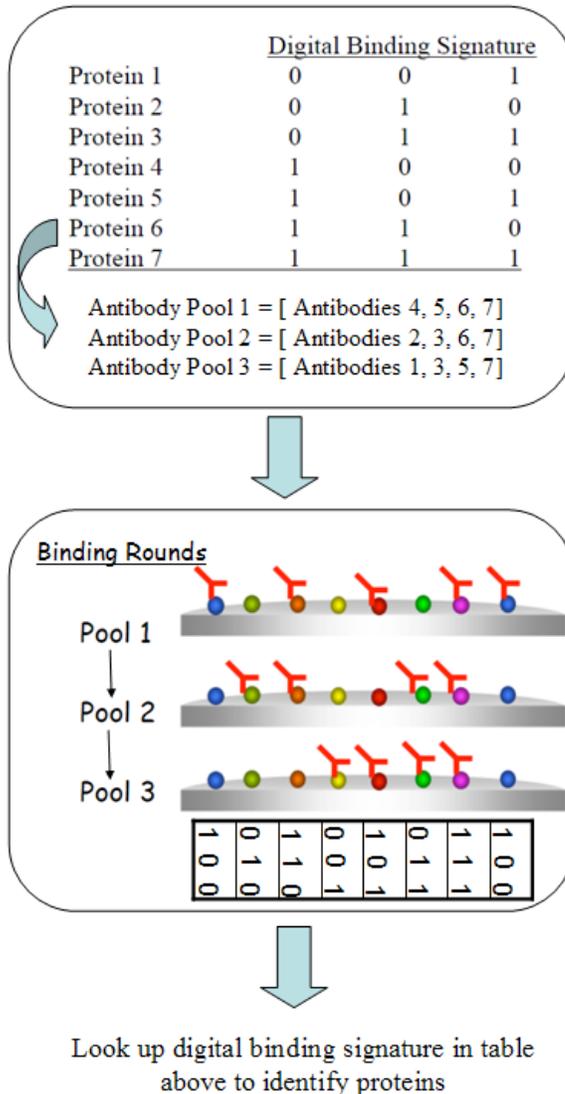
Figure S4



Dissociation of surface-bound antibody:target complexes is insignificant over 48 hours. We performed antibody binding onto immobilized Cy3-targets and counted the number of antibody:target complexes. We washed the flow cell over 48 hours and analyzed the number of complexes every 8 hours. Here we plot the number of complexes over time. We do not see a decay of the number of antibody:target complexes over time, so there is likely an antibody-surface interaction.

Figure S5

Efficient multiplexing. Here we illustrate an efficient strategy for multiplexed protein detection in which n proteins may be quantified in $\left\lceil \frac{\log_2 n}{c} \right\rceil$ binding rounds, where c is the number of independent fluorescence channels used for antibody detection. This logarithmic encoding is based on the method by Gunderson et al, used to decode bead-based random microarrays.[56] Here we describe a 7-plex assay, using 1 fluorescence channel, as a small-scale example.



First, each protein is assigned a unique digital signature. Next, fluorescent antibodies for each protein are pooled into combinations that are determined by the columns of the digital signatures. (In the example above, the three columns of the signatures dictate the compositions of the three “antibody pools”.) Then, immobilized proteins are probed by one antibody pool per binding round. In each binding round, proteins of different species are bound and detected. In between binding rounds, antibodies are stripped. After probing with all of the

antibody pools (three in this example), the history of binding at each position on the slide is analyzed. In this manner, each position on the flow cell becomes represented by a binding signature. Finally, the pre-assigned digital signatures are used to decode the flow cell positions into protein identities. Moreover, the number of occurrences of each signature is counted to determine protein abundance.

To illustrate the decoding procedure, in this example there are two positions on the flow cell that have the binding history 1-0-0 (i.e. bound in round 1, unbound in round 2, and unbound in round 3). This signature corresponds to Protein 4, so the number of instances of that signature on the flow cell (two), indicates the abundance of Protein 4.

Supporting Information

for

“Nanogel surface coatings for improved single-molecule detection substrates”

Design of SM antibody adsorption experiments

To quantify the non-specific adsorption of fluorescent antibody onto coated surfaces, we generated coated surfaces in flow cells and then exposed them to dilute solutions of labeled antibody. We visualized the antibody molecules that adsorbed to the surfaces by a total internal reflection fluorescence (TIRF) microscope that is capable of detecting single fluorescent molecules.

Several steps were taken to enable the accuracy and precision of the antibody adsorption measurements. We used antibodies that were labeled with an average of eight fluorophores, so that almost all were fluorescently labeled. This can be seen by approximating the dye-labeling reaction as a Poisson process. Eq. 1 is used to compute the fraction of antibody molecules that were labeled in our experiments:

$$P(N = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{Eq. 1}$$

If $\lambda = 8$ (the labeling efficiency reported by the supplier), then only ~0.03% of the antibody molecules contained no label ($k = 0$, in the equation above), and so ~99.97% of the antibodies were labeled with at least one dye.

To maintain precision, all surfaces in an experiment were exposed to antibody pipetted from an identical aliquot. This eliminated variation otherwise caused by

antibody dilution errors. We found that despite using low-protein-binding tubes for storing the antibody, there was noticeable loss of antibody within the storage aliquots over time due to adsorption to the surfaces of the tubes. To minimize this variation, we parallelized experiments whenever possible. Because of the limited number of flow cells at our disposal, there were some experiments for which we could not test all surfaces and replicates simultaneously. In these cases, common control surfaces were included in the experiments to remove (by normalization) the variation caused by antibody loss in storage tubes.

SM imaging

SM imaging was performed on a Nikon TE-2000 inverted microscope fitted for total internal reflection fluorescence (TIRF) (Nikon, Melville, NY). A 640 nm, 40 mW laser was used for fluorescence excitation of Cy5 dyes, and a 532 nm, 75 mW laser was used for excitation of Cy3 dyes. (Cube-40C and Compass 215M, Coherent, Santa Clara, CA). Illumination of the surface was controlled through a computer animated shutter (Prior Scientific, Rockland, MA). The laser beams were coupled into one end of an optical fiber cable with the other end of the cable attached to the TIRF illuminator. Before reaching the objective, the beams were passed through a D635/30 band pass filter for the red channel and HQ545/30 band pass filter for the green channel (Chroma, Brattleboro, VT). Objective type total internal reflection was achieved through a 60x TIRF oil objective with numerical aperture 1.49 (Nikon, Melville, NY). The surface modifications were performed in a glass flow cell (FSC2, Bioprotechs, Butler, PA) which was mounted onto the microscope stage. The incident angle of the laser beam was

adjusted to generate total internal reflection, which creates an evanescent wave that decays exponentially at the glass-water interface into the flow cell to a distance of about 300 nm. To prevent fluorophore photobleaching and blinking, an oxygen scavenger system based on glucose oxidase and catalase and a blink-reduction system (Trolox) was added to the flow cell prior to image acquisition.[77] Emitted photons from the labeled antibody molecules were collected by the objective and passed through a dichroic mirror (custom Cy3/Cy5, Semrock, Rochester, NY) and emission filters (LP02-647RU-25, Semrock, Rochester, NY for the red channel, and HQ610/75, Chroma, Brattleboro, VT for the green channel). Light from each of the channels was then detected sequentially by a charge coupled device (CoolSnap ED, Roper Scientific, Tucson, AZ) which imaged a 140 μm by 100 μm (1400 px x 1000 px) region of the surface. Images of the flow cell surface were acquired with 0.5 second exposure. For adequate sampling, images were acquired at five positions across the length of the 3 mm x 35 mm flow-cell with independent surfaces analyzed in duplicate or triplicate. Therefore, between 500 and 10,000 molecules were sampled for each reported data point.

Image processing

We used custom image processing software described previously to determine the number of fluorescent molecules present on the surface.[76] Two notable traits of the raw data are non-uniform illumination across the field of view (due to our TIRF apparatus) and fluorescence objects of different sizes (due to multi-labeling of protein and antibodies by fluorophores). The image processing described below has been validated to normalize

for both non-uniformities; we normalize for non-uniform illumination, and we filter out molecules that are overlapping other molecules.

Briefly, we employed a Metamorph Journal (Molecular Devices, Downington, PA) that reads in a raw, 12-bit TIFF image and performs iterative intensity thresholding to determine fluorescence objects in an image. We iterated over intensity thresholds by increments of 50, with each iteration defining a threshold for local intensities (“local threshold”). (For the Cy5 channel we iterated over the range of 175 to 4025, and for the Cy3 channel, we iterated over the range of 200 to 4050. This normalizes for emission intensity differences between the two channels.) For each local threshold, we defined objects as sets of pixels that: 1) have intensity values that are greater than the local threshold, 2) are contiguous with other pixels within that object, and 3) comprise an area between 2 and 16 pixel² (20 and 160 nm²). The contiguity requirement creates a list of potential objects while the area requirement filters out those that are likely to be combinations of multiple objects. For each object found at each local threshold, we outputted the X-Y locations of the object centroids. This set of 2-D points was imported into a custom MATLAB script (MathWorks, Natick, MA), which created binary images of the objects. These objects were counted by the software to generate a measure of molecules per unit area.

Correlation analysis

To determine whether fluorescence objects in the red and green channels (antibodies and target proteins) were overlapping significantly or simply by chance, correlogram analysis was performed as previously described.[76] We found that XY

translations were perfectly suited to identify proper alignment, and additional channel distortion played a negligible role. In brief, the red and green channel images were first analyzed (see Image Processing) to produce binary matrices, indicating the positions of each fluorescence object. A sub-matrix of the green channel was chosen as the reference matrix and compared to a sub-matrix of the red channel, using the correlation function

$$C = \sum_{i=1}^{316} \sum_{j=1}^{316} R_{i,j} G_{i,j}$$

Eq. 2

in which R and G are sub-matrices representing 316 x 316 pixel sub-images of the red and green images, respectively. This correlation was determined for truly-aligned sub-matrices of the red and green channels, as well as for red and green sub-matrices that were mis-aligned by a translational offset. We sampled offsets between -100 pixels and +100 pixels in X and Y dimensions with respect to the true alignment. A correlogram was created by plotting the correlation C as a function of X and Y translational offsets. A distinguishable peak near the origin indicates that the true alignment produced a correlation that would not be expected given a random overlapping of objects.

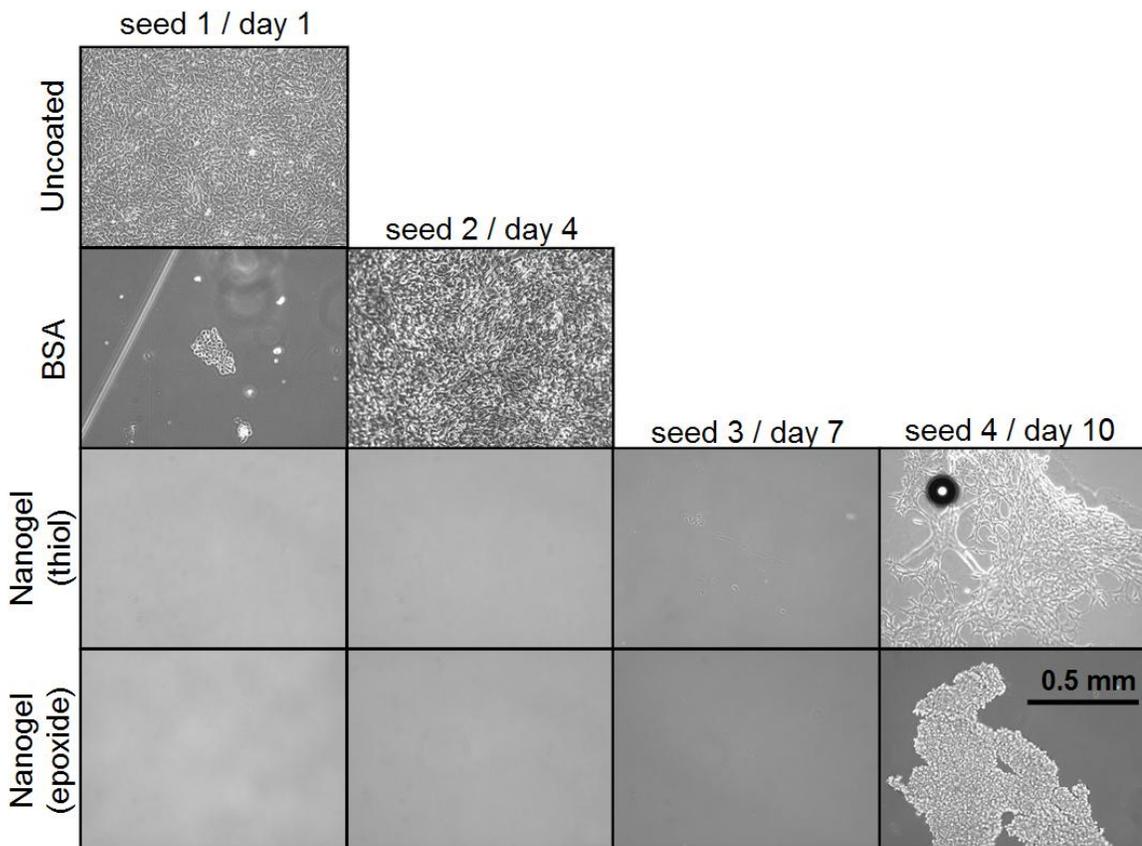
Since no anti-photobleaching system is perfect, some molecules on the surface may be missed. Any of these false negative events would reduce the number of observed overlaps. Thus, the height of the peak in the correlogram is a lower bound of the true level of correlation.

Cell adhesion assay and time course

Here we analyzed the nanogel-coated surfaces by measuring cell adhesion. In this method, the surface is exposed to fibroblast cells in medium containing physiological levels of serum proteins (100 mg/mL). Because fibroblasts require the adhesion of serum proteins before they can attach to a surface, cell growth serves as a measure of the amount of adsorbed serum protein.

We generated BSA-coated and nanogel-coated surfaces, exposed them to fibroblasts in serum-containing growth media, and imaged the surfaces by phase contrast microscopy the next day. Images of the surfaces after cell seeding is shown in Figure S1, column 1. Fields of view were chosen to depict the region of each surface with the most cell adhesion. The uncoated control surfaces were covered with a monolayer of cells after only one exposure to cells and medium. In contrast, the BSA-coated surfaces were largely free of cells, with cells covering ~1% of the surface while PEG-BSA nanogel-coated surfaces (both epoxide- and thiol-reacted) displayed virtually no cell adhesion. Therefore, although the cell adhesion measurements provide less dynamic range than the SM antibody adsorption measurements taken previously, we observed remarkably similar results between the methods.

Figure S6



We next wanted to investigate the stability of the surfaces. We performed a time course of cell adhesion by continuing to expose the surfaces to serum proteins and fibroblasts every 3 days, until we observed cells covering a significant portion of each surface. Images of the surfaces over the time course (days 4, 7, and 10) are shown in Figure S6, columns 2-4. BSA-coated surfaces maintained low adhesion until after the second seeding (day 4), at which point they were completely covered with monolayers of fibroblasts. In contrast, the PEG-BSA nanogel-coated surfaces resisted serum protein adsorption through three rounds of seeding. On day 10, after the 4th cell seeding, the thiol-reacted nanogel surfaces lost their resistance to cell adhesion and were found to contain large networks of cells. The epoxide-reacted nanogel surfaces also began to fail

after the 4th seeding. Therefore, the BSA coating was stable for approximately 4 days whereas the nanogel coatings resist adsorption of serum protein mixtures for approximately 10 days before deteriorating. We determine that the nanogel surface will be more useful than BSA surfaces in solid-phase protein studies that require extended durations.

Despite the differences between the SM adsorption assay and the cell adhesion assay we were happy to find that data at the first time point were similar between the two methods. This lends support to the relevance of the SM adsorption measurements to higher concentrations.

Atomic force microscopy

AFM (Nanoscope V multimode SPM, Veeco Instruments) was utilized in tapping mode. We examined morphologies of the PEG-BSA nanogel-coated samples on both epoxysilanated and mercaptosilanated glass surfaces without BSA capping. *In situ* surface imaging was performed using silicon nitride tips (NP-10, Nanoprobe, Veeco) with a nominal spring constant of 0.32 N m^{-1} , and scan rates of 1-1.2 Hz. Force curve analysis was performed with a spring constant of 40 N m^{-1} , Poisson's ratio of 0.5, and semi-vertical tip angle of 25° . Roughness analysis and force curve analysis were performed on two replicate surfaces for each coating type.

Supporting Information

For

“Sensitive single-molecule protein quantification and protein complex detection in a microarray format”

MATERIALS AND METHODS

Coating glass with nanogel

The microarray chips were formed from a glass coverslip (Fisher Scientific, Pittsburg, PA) that was treated with nanogel coating by epoxy coupling[118]. The nanogel coating protocol described previously was streamlined by substituting PEG-octovinylsulfone with PEG-4 maleimide (Creative PEGworks, Winston Salem, NC). This modification greatly reduced the nanogel preparation time, since PEG-4 maleimide is available commercially. After BSA and Tris capping, the chips were activated with EDC and NHS[76] (Sigma Aldrich, St. Louis, MO), washed in water and dried with argon.

Capture and detection

Capture antibodies for p53 and MDM2 (R&D Systems, Minneapolis, MN) were printed by hand in triplicate onto the nanogel coating using a microarray pin (Arrayit Corporation, Sunnyvale, CA) and allowed to dry 15 minutes. To form a flow chamber, the chips were affixed, by a silicon gasket interface (Grace Biolabs, **Bend, OR**), onto a standard glass slide (1” x 3”) fitted with two 1 mm holes. The flow chamber was inserted onto the microscope stage, and the positions of the antibody fields were saved via Metamorph software (Molecular Devices, Sunnyvale, California) using the white light

image of the dried fields. The activated surface was quenched with Tris for 10 minutes and washed with Tween-20.

The six-hour, semi-automated workflow went as follows. The flow chambers (three technical replicates) were loaded with the protein sample (cell extract or a protein standard), exposed for 2 h, washed with PBS, loaded with p53 and MDM2 biotinylated detection antibody (only one or the other for protein-protein interaction detection), exposed for 2 h, washed with PBS, loaded with Cy5-labeled streptavidin, exposed for 2 h, and washed with PBS.

To measure the intrinsic limit of detection and dynamic range, we fabricated microarray chips with biotinylated-BSA printed onto the surface. After washing with Tween-20, flow chambers were loaded with different concentrations of Cy5-streptavidin. After 2 hr exposure, the chips were washed with PBS, imaged, and analyzed.

Cell culture and protein extraction

In brief, we grew HCT116 cells under normal and 5-fluorouracil DNA damage conditions, with three biological replicates per condition. Cells were harvested and equalized for cell number. Cells were then lysed, and lysate was flowed onto microarrays we had fabricated with p53 and MDM2 monoclonal capture antibody fields (three technical replicates per microarray).

HCT116 cells were grown in DMEM + FBS medium and passaged into six flasks (Sigma Aldrich, St. Louis, MO). Cells were grown to a confluence over 12 hours, then 25 μ M 5-fluorouracil was added to three of the flasks. Cells were grown for an additional 12 hours and then trypsinized. Cells were rinsed twice in PBS and resuspended in lysis

buffer (1 mM EDTA, 0.5% Triton X-100, 1 mM DTT, 150 mM NaCl, one Complete protein extraction tablet (Roche, Basel, Switzerland)) and normalized to the same cell concentration (10^7 cells / mL). Cells remained on ice for 15 minutes and were centrifuged at 2000 x g for 5 minutes. Supernatant was used as input for the single-molecule microarrays.

Imaging

Total internal reflection fluorescence (TIRF) imaging was performed at the locations of the antibody fields, the images were processed by Metamorph and MATLAB (Mathworks, Natick, MA)[76]. Molecule counts of fluorescence objects were obtained as well as average fluorescence intensity per unit area. To average using ensemble intensities, we used 400 x 400 pixel viewing areas from the center of the TIRF images and calculated the average intensity over each pixel. To obtain digital counts we applied custom software to identify discrete fluorescence objects, and summed the number of objects

SUPPORTING FIGURES

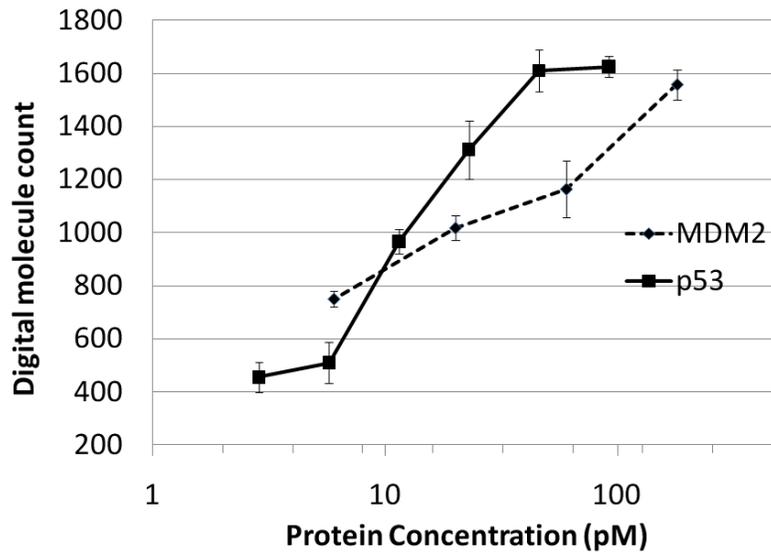


Figure S7. Using a set of paired, monoclonal antibodies, we performed single-molecule microarray analysis on protein standards. Using digital analysis by fluorescence object counting, we detected p53 and MDM2 down to the low picomolar range. This matches the limit of detection using these antibodies in an optimized ELISA.

APPENDIX TWO: DEEP SEQUENCING DETECTS RARE GENETIC POLYMORPHISMS ASSOCIATED WITH RESPIRATORY DISTRESS SYNDROME IN NEWBORNS

This appendix will contain a manuscript in preparation by Yue Yun and me. At the time of this defense, the manuscript is still in early stages. For this collaborative project we performed massively-parallelized Illumina DNA sequencing to find mutations correlated with neonatal respiratory distress syndrome (RDS). Additionally this work attempts to directly address the Rare Variants Hypothesis regarding heritability of common diseases. To summarize the state of the project we:

- Quantified and normalized DNA samples for 850 individuals using epMotion robotics.
- Prepared and submitted custom DNA libraries for over 50 lanes of pooled-sample sequencing on the Illumina Genome Analyzer and HiSeq 2000.
- Created a streamlined software package, utilizing existing software (SPLINTER[119]) and developing custom software for the identification of single nucleotide polymorphisms (SNPs) from 54 billion nucleotides of data.
- Submitted DNA samples to and analyzed results from Taqman and Sequenom genotyping platforms for the 850 individuals.