Characterization and Analysis of a Novel Platform for Profiling the Antibody Response

by

Rebecca Halperin

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved October 2011 by the
Graduate Supervisory Committee:

Stephen Johnston, Chair
Andrew Bordner
Thomas Taylor
Phillip Stafford

ARIZONA STATE UNIVERSITY

December 2011

ABSTRACT

Immunosignaturing is a new immunodiagnostic technology that uses random-sequence peptide microarrays to profile the humoral immune response. Though the peptides have little sequence homology to any known protein, binding of serum antibodies may be detected, and the pattern correlated to disease states. The aim of my dissertation is to analyze the factors affecting the binding patterns using monoclonal antibodies and determine how much information may be extracted from the sequences. Specifically, I examined the effects of antibody concentration, competition, peptide density, and antibody valence. Peptide binding could be detected at the low concentrations relevant to immunosignaturing, and a monoclonal's signature could even be detected in the presences of 100 fold excess naive IgG. I also found that peptide density was important, but this effect was not due to bivalent binding. Next, I examined in more detail how a polyreactive antibody binds to the random sequence peptides compared to protein sequence derived peptides, and found that it bound to many peptides from both sets, but with low apparent affinity. An in depth look at how the peptide physicochemical properties and sequence complexity revealed that there were some correlations with properties, but they were generally small and varied greatly between antibodies. However, on a limited diversity but larger peptide library, I found that sequence complexity was important for antibody binding. The redundancy on that library did enable the identification of specific sub-sequences recognized by an antibody. The current immunosignaturing platform has little repetition of sub-sequences, so I evaluated several methods to infer antibody epitopes. I found two methods that had modest

prediction accuracy, and I developed a software application called GuiTope to facilitate the epitope prediction analysis. None of the methods had sufficient accuracy to identify an unknown antigen from a database. In conclusion, the characteristics of the immunosignaturing platform observed through monoclonal antibody experiments demonstrate its promise as a new diagnostic technology. However, a major limitation is the difficulty in connecting the signature back to the original antigen, though larger peptide libraries could facilitate these predictions.

ACKNOWLEDGMENTS

I am grateful Stephen Albert Johnston for serving as my advisor throughout my graduate school experience, as well as the opportunity to work on these exciting projects. He has provided direction and support for everything that I have done here. Nothing in this dissertation would have been possible without his big ideas and willingness to take risks.

I would like to thank the rest of my committee for their mentorship. Phillip Stafford provided valuable feedback and guidance on an almost daily basis. Tom Taylor's input has provided a critical outside perspective that has influenced me to think more quantitatively. Andrew Bordner has consistently provided insightful questions and constructive commentary. Also, Neal Woodbury was a member of my committee for the first few years, until new responsibilities conflicted with his duties as committee member. He also provided a great deal of feedback that influenced the way I approached the projects.

The development of the immunosignaturing technology has truly been a group effort, involving many people over the course of several years. D. Mitch Magee was involved in the early development of the technology, including the immunization experiments presented in chapter 2. J. Bart Legutki has been instrumental in optimization of the assay and development of quality control measures, as well as providing some of the data in chapter 2. He also published the first paper on the technology, paving the way for future publications from the group. John Lainson has played a critical role in preparing the slides and the spotting plates as well as printing the arrays. Chris Diehnelt, Sergei Servosky, and Phillip Stafford

have also been involved in the array production, including array design and peptide handling issues.

I would also like to thank my collaborators Abner Notkins and Ying Xiong. They provided the polyreactive antibody and the idea for that project, as well as valuable discussion and feedback on the manuscript.

The production, use, and analysis of the BioMicroChip were also a group effort. Nidhi Gupta, Kevin Brown, and all of the others who worked at the PolyTech and HealthTel were responsible for the production of the chips. Bart also played an important role in the assay development and optimization. Neal and Phillip made critical suggestions with respect to the data analysis.

I would also like to thank everyone who helped with the creation of GuiTope. Jack Emery played an invaluable role in setting up the first version of the program and convincing me that I could write software. Krupa Navalkar suggested the dipeptide inversion approach and also provided feedback on the software and manuscript. Bart provided data supporting the dipeptide inversion model and tested the software. Kevin gave critical suggestions on implementation and feedback on the manuscript.

All of my current and former colleagues at CIM have made this a great place to work. I would like to thank all of the other faculty members, including Chris Diehnelt, Zhan-Gong Zhao, Kathy Sykes, Alex Borovkov, Sergei Svarovsky, and Mitch Magee, who have always asked challenging questions and willing to offer advice. My fellow graduate always provided interesting discussion and camaraderie, including Hoojoon Lee, Luihui Shen, Lucas Restrepo, Krupa Navalkar, Muskan

Kukreja, Hu Duan (Tiger), and Kurt Whittmore. All of the lab members have always been willing to help out wherever needed including Bart Legutki, John Charles Rodenberry, Valeriy Domenyuk, Nidhi Gupta, Paul Belcher, Andrey Loskutov, Loren Howell, Kathy Boltz, Maria Gonzalez, Tien Olsen, Brian Chase, and Kaitlin Kroening. I want to give a special thanks to my lab neighbors who have been a constant source of support and encouragement, including John Lainson, Lori Phillips, Nikki Hackett, and Alex Carpenter. Kevin Brown and Preston Hunter provided much needed assistance with technical issues. Pattie Madjidi and Penny Gwynne deserve a huge amount of thanks for keeping the lab running.

TABLE OF CONTENTS

Page

CHAPTER Page

APPENDIX Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

PREFACE

Chapter 2, "Antibodies as Biomarkers", describes a number of experiments demonstrating the immunosignaturing concept. The introduction lays out the justification for using antibodies as biomarkers and explains the advantages of the immunosignaturing method over other approaches. The experimental data presented starts from the simplest case of monoclonal antibody profiles, and works through several levels of increasing complexity including various mixing experiments, responses to immunizations, finally presenting human immune response data. By illustrating these different levels of complexity the manuscript attempts to reconstruct how the immunosignaturing method works.

This project was the result of the effort of a number of individuals. S. A. Johnston developed the concept of immunosignaturing. P. Stafford and S. A. Johnston were involved in the experiment design for all of the experiments in this project. P. Stafford also performed most of the statistical analysis and drafted the manuscript. J. B. Legutki was involved in development of the assay and quality control procedures, performing the mouse infections and immunizations, and assaying the human influenza patients. D. M. Magee contributed to the design and analysis of the immunization and blocking experiments. J. Galgiani was involved in the experiments utilizing the human Valley Fever sera samples. All co-authors contributed to manuscript revisions.

My role in this project was primarily in performing and analyzing the results of the experiments involving monoclonal antibodies. I performed the experiments shown in Figure 2.2 that demonstrate that monoclonal antibodies have unique

signatures on the array. I carried out the dilution series experiments in Figure 2.3 as well as estimated the half maximal binding concentrations. I designed the custom arrays shown in Figure 2.4, ran those arrays, and analyzed the data. I also designed, executed, and analyzed the Fab vs. IgG experiment in Figure 2.5, and the competition experiment in Figure 2.6. I also contributed to writing parts of the manuscript pertaining to the monoclonal experiments and revising the manuscript as a whole.

There are a few places where my interpretation of the data differs from what is presented in the manuscript. Figure 2 shows the monoclonal binding patterns as median normalized data in a heatmap form. While this does nicely illustrate that antibodies have unique patterns on the array, this view of the data fails to show how different the antibodies in the number of peptides that they bind above background. I believe that this is an important phenomenon to understand in predicting the capabilities and limitations of immunosignaturing. I have made some observations of trends in which antibodies bind more peptides than others that I discuss in Chapter 7. I also think that the dilution experiments (Figure 3) and competition experiment (Figure 6) demonstrate that the antibodies are not typically saturating the binding sites on the peptides at the concentrations we typically use. I think that this is an important point, as it would imply that patterns of different antibodies should be additive, which simplifies the interpretation of immunosignaturing results. I also discuss in more detail how the concentration at which these experiments were run compares to the antibody concentrations you would expect in a real immune response in my Conclusions Chapter.

CHAPTER 1

INTRODUCTION

Antibodies have a number of characteristics that make them appealing as biomarkers. The immune system is constantly monitoring changes in the body and new antibodies are produced when anything recognized as foreign is found. Though the newly recognized antigen may be at a very low concentration, a B-cell that produces an antibody that binds to it may rapidly clonally expand, thus providing a natural amplification of the signal. While antibody response is most classically thought of in the context of infectious and autoimmune disease, antibodies are increasingly being discovered in other diseases including cardiovascular disease, neurodegenerative disease, and cancer (Brettschneider et al. 2005; Cho-Chung 2006; Erkkila et al. 2000). In the case of autoimmune and other chronic diseases, the antibody response may develop in the early stage of the disease, well before the onset of symptoms, thus enabling pre-symptomatic diagnosis (Arbuckle et al. 2003; Chapman et al. 2007; Nell et al. 2005; Whittingham et al. 1997). Antibodies are rather stable, which facilitates sample processing and enables the use of archived serum samples. Here I will review the current literature concerning harnessing the information in the antibody response, including the production, binding mechanisms, and cross reactivity of antibodies, current technology for discovering and detecting disease relevant antibodies, and methods for determining antibody targets and epitopes.

Production and Function of Antibodies

Each immunoglobulin molecule is composed of two identical heavy chains and two identical light chains. Heavy chains are composed of three or four constant domains and one variable domain. Light chains contain only one constant domain and one variable domain. The variable regions of the heavy and light chains form the paratope, or antigen binding site. The constant regions determine the antibody isotype that is responsible for the effector function. Each immunoglobulin is a "Y" shaped molecule with two identical paratopes. The variable domains are produced through a unique process of gene rearrangement. Antibody production consists of two stages: non-antigen dependent and antigen dependent. The quantitative aspects of each stage are summarized in Table 1. In the non-antigen dependent stage, the pre-B-cell undergoes a process of gene rearrangement to generate the variable regions of the antibody. This process is capable of producing on the order of 10^{16} unique B-cell receptors (Schroeder and Cavacini). However, only 10-20% of these rearrangements result in functional B-cell receptors; those with non-functional B-cell receptors undergo apoptosis. Each immature B-cell expresses a single arrangement so that all antibody molecules expressed by each B-cell have identical paratopes. The naïve B-cells have a half-life of only four days without receiving further survival signals, which results in a turnover of a million cells per day. At this point the antibody is primarily expressed as a membrane bound form and serves as a B-cell receptor. The immature B-cell will not divide or further differentiate unless it is stimulated by binding of antigen to its receptor.

Table 1.1 Quantitative Summary of Antibody Production

Stage	**Diversity of BCR**	**Population Dynamics**	**Quantity of Soluble Antibody**
Immature B-cells	10^{16} possible arrangements, 10-20% result in functional BCR	Constant population turns over at a rate of a million per day	Negligible
Activated B-cells	A small number of B-cells clonally expand, typically have low affinity for the antigen	A population of ~10^{7} short lived plasma peaks about a week after antigen exposure	~100ug/ml
Germinal Center B-cells	Undergo somatic hypermutation accumulating 2.8 mutations per day, 1% of mutations are favorable	Process takes about two weeks and results in about a million clonally related B-cells	Negligible
Affinity Matured B-Cells	Have an average of nine somatic mutations, and around 100 fold improved affinity to the antigen	Long lived plasma cells may survive in the bone marrow for decades, capacity is about a billion cells, with 0.1% specific to a particular antigen	~10ug/ml per antigen
Innate-like B Cells (B1 Cells)	Germline encoded BCR, some biases exist compared to immature BCR repertoire	Represent a small percentage of circulating B-cells, secret antibody without antigen dependent signaling	~1mg/ml

When a B-cell receptor binds an antigen, the response depends on the context and presence or absence of other signals. B-cells are screened for self-reactivity and those that are self-reactive undergo apoptosis or enter an anergic state. When the antigen binding occurs in the context of particular co-stimulatory signals,

it can either become short lived plasma cell that will secrete antibody of the IgM isotype for a few days before undergoing apoptosis (Smith et al. 1996) or enter a germinal center where it undergoes affinity maturation and further differentiation. Germinal centers, which are specialized sites where B-cells undergo affinity maturation, are usually seeded within a few days after the initiation of the immune response (Allen, Okada, and Cyster 2007). During the process of somatic hypermutation, mutations into the variable regions of the antibody are introduced at a rate of up to 10^{-3} mutations per base pair per cell cycle (Schroeder and Cavacini). B-cells whose modified receptor has sufficient affinity to compete for binding to the antigen and also receive appropriate signals from T-cells are selected for survival (Allen, Okada, and Cyster 2007). These B-cells may either undergo further rounds of mutation or differentiate into memory B-cells or plasma cells. At this time class switch recombination may also occur to generate IgG or IgA antibodies (Schroeder and Cavacini). Plasma cells are terminally differentiated cells that are responsible for secreting large quantities of antibody. In the peripheral circulation, plasma cells will only survive for several weeks. However, those that migrate to specialized niches in the bone marrow or sites of chronic inflammation may continue to produce antibody for years. The bone marrow can support approximately 10^9 plasma cells, and approximately 10^6 will enter the bone marrow from any given immune response. These 10^6 plasma cells are capable of maintaining serum antibody concentrations around 67nM for a specific antigen (Radbruch et al. 2006). The half life of IgG ranges from one to three weeks depending on the isotype (Morell, Terry, and Waldmann 1970).

The primary function of antibodies is generally thought to be to recognize pathogens, and either destroys them and/or alerts other branches of the immune system. The functions that antibodies are capable of include complement fixation, opsonization, neutralization, and signaling through Fc receptors. The function is dependent on the antibody isotype and the context. In some cases neutralizing antibodies are sufficient to prevent or clear an infection. Anti-self antibodies are present in many autoimmune diseases and may play a causal role in some. The role of antibodies in other disease that are not typically thought to be immune mediated is just beginning to be elucidated. There have been many reports of the existence of antibodies to tumor-associated autoantigens in cancer patients, but their role in controlling or enabling tumor growth is still unclear (Reuschenbach, von Knebel Doeberitz, and Wentzensen 2009). Similarly, a number of autoantibodies have been identified in association with Alzheimer 's disease, though the extent to which each has pathogenic or protective role in disease progression requires further study (Colasanti et al. 2010). Conversely, in atherosclerosis, a pathogenic role of anti-oxidized low density lipoprotein IgG antibodies and a protective role of IgM antibodies to the same targets have been demonstrated, but their potential as biomarkers of disease progression remains controversial (Hansson and Hermansson 2011).

Antibody Recognition Mechanisms

On a conceptual level, antibody-antigen interactions may be represented as distances in a shape space. The more complementary the antibody is to the antigen,

the closer the two are in shape space, and this distance would be proportional to the affinity of the interaction. This shape space complementary is not meant to literally represent geometric complementary as in a lock and key binding model, but can represent any type of binding interaction. Lapedes and Farber developed a method to transform affinity data to coordinates in a shape space (Lapedes and Farber 2001). This method was later applied to describe the relationship between genetic distance and antigenic distance between strains of the influenza virus (Smith et al. 2004). They also found that the dimensionality of the shape space was approximately five dimensions. A comparison between the shape space distance and the number of mutations between two strains of influenza viruses found that the relationship was fairly linear, though some mutations had a disproportionately large effect. If these abstract dimensions could be related to physiochemical descriptors such as geometric complementary, electrostatic interactions, etc. the shape space concept would be an even more powerful framework for understanding and predicting antibody recognition, particularly for understanding and predicting the cross-reactivity of unrelated entities.

The current understanding of the biophysical basis of antibody recognition is based primarily on crystal structures of antibody-antigen complexes. Antibody binding generally fits the lock and key model with the epitope and paratope having complementary shape and local charge. There are some examples were induced fit has been found to be important to antibody recognition (Braden and Poljak 1995). Particularly, there is some evidence that germline encoded variable regions have more conformational flexibility which allows for lower specificity and lower affinity

binding and the process of affinity maturation increases the rigidity of the antibody (Manivel et al. 2000). Additionally, the antibody structure may also depend on the type of antigen that it recognizes. For example, antibodies that recognize peptides tend to have a more grooved paratope compared to antibodies that bind proteins (Chen, Van Regenmortel, and Pellequer 2009). For the purpose of using antibodies as biomarkers, it will be important to understand the biophysical characteristics of disease relevant antibodies to optimize detection.

Antibodies are "Y" shaped molecules with two identical binding sites at the ends of the arms. The hinge is flexible allowing the antibody to bind two antigens appropriately spaced apart. The affinity of the antibody to a target that allows both arms to bind will approach the product of the affinity of each individual arm. This bivalent binding is physiologically important allowing antibodies to be especially effective at binding to the repetitive features on the surface of bacteria and viruses. Recently, an example of heterobivalent binding was observed where an anti-HIV monoclonal antibody had moderate affinity for the HIV spike protein and low affinity for the viral lipid membrane, which led to enhanced affinity for the virus by being able to bind both simultaneously (Mouquet et al. 2010). When measuring antibody affinity *in vitro* it is important to keep in mind whether the assay facilitates two arm binding or only permits one arm binding (Kaufman and Jain 1992).

Antibody Specificity and Polyreactive Antibodies

Much of our conceptual framework for understanding the adaptive immune response lies around the central principle of distinguishing self from non-self. It is

commonly held that the immune system is capable of generating antibodies that recognize virtually any molecular shape. A great diversity of antibody shapes is present in the immature B-cell population. Those that recognize shapes found in the self are eliminated, altered, or suppressed at checkpoints both during and after maturation. Only when a naïve B-cell encounter foreign antigen does it clonally expand and differentiate to produce soluble antibody, as well as trigger other arms of the immune response. Under this simplified framework, the production of antibodies against self-molecules would only be expected as a pathological feature of autoimmune diseases.

There is a class of antibodies that does not fit neatly into this framework. These antibodies, known as natural antibodies, are produced in the absence of foreign immune stimulus. Many of these antibodies will react with self-antigens. They often have multiple specificities. Typically, they are of the IgM isotype, though IgG and IgA natural antibodies have also been identified. Produced by a specialized subset of B-cells, called B1 cells, they tend to show less deviation from the germline sequences. They may play a role in fighting pathogens, as they have been shown to bind to some pathogen associated molecular patterns, such as bacterial lipopolysaccharide (LPS), so may serve an innate-like function (Zhou, Tzioufas, and Notkins 2007). Another function may be anti-cancer activity as they can recognize the altered glycosylation patterns of tumors (Vollmers and Brandlein 2007). Apoptotic cells may also be targeted by natural antibodies, serving to help clean up cell debris. Merbl et al. used a microarray approach to simultaneously measure binding of total natural immunoglobulin to several hundred potential auto antigens

simultaneously, which gives a much more global view of what types of antigens natural antibodies recognize (Merbl et al. 2007). They were able to identify a common pattern of autoantigen reactivity in newborns, suggesting these antibodies play an important role early in life. The importance and mechanisms of these beneficial functions is just beginning to be elucidated.

Though antibodies raised to a specific target are often expected to only bind to the target, "specific" antibodies are frequently found to cross-react with unrelated targets. For example monoclonal and polyclonal antibodies to yeast proteins were found to cross-react to other yeast proteins to varying degrees when tested against a whole yeast proteome microarray (Michaud et al. 2003). A similar result was found when commercially available monoclonals were used to probe a human proteome array (Kijanka et al. 2009). A variable level of cross-reactivity has also been observed between a number of unrelated pathogens (Vigil, Davies, and Felgner 2010). While these examples with protein arrays may be the most recent and comprehensive looks at antibody cross-reactivity, the first observations of antibody cross-reactivity were identified decades ago. In the 1980's several studies demonstrated that monoclonal antibodies raised to viruses could cross-react with normal human tissues and proteins (Fujinami et al. 1983; Srinivasappa et al. 1986). The role of such molecular mimicry in triggering autoimmune disease is still a matter of much research and debate (Blank, Barzilai, and Shoenfeld 2007; Oldstone 1998; Tsuchiya and Williams 1992).

Antibody Profiling Technologies

The potential of antibodies as biomarkers is increasingly being recognized and the number of technologies to detect them is growing. An ideal antibody profiling platform would be able to detect antibodies to any type of target. It should also be able to detect many antibodies in parallel, as well as being amenable to process a large number of sera samples for clinical studies. The dynamic range should be large to be able to detect antibodies varying greatly in affinity and abundance. The reproducibility of the measurements is also critical. Biologically relevant information about the antibodies target would be ideal. Finally, the cost may not be prohibitive. There are a number of technologies that meet some several of these requirements (Table 2), but none has yet to emerge that satisfies them all.

Table 1.2 Antibody profiling technologies

Assay	**Number of markers/ samples per assay**	**Dynamic Range/ Reproducibility**	**Types of antibody targets detected**	**Antigen/ Epitope information**	**Assay Cost/ Serum Amount**
Spotted Protein Arrays	hundreds to thousands of proteins/ dozens of samples	Low Dynamic Range/High Batch to Batch Variability	protein targets, partially denatured, no PTM's	Antigens tested	High/ µLs
Cell Lysate Arrays	several thousand fractions/ dozens of samples	Low Dynamic Range/High Batch to Batch Variability	protein targets, partially denatured, with PTMS and disease relevant mutations/ isoforms	not directly, may use other techniques such as mass spec to identify proteins in the lysate	Mid/ µLs

Table 1.2 Antibody profiling technologies (continued)

Assay	Number of markers/ samples per assay	Dynamic Range/ Reproducibility	Types of antibody targets detected	Antigen/ Epitope information	Assay Cost/ Serum Amount
NAPPA	hundreds to thousands of proteins/ dozens of samples	Low Dynamic Range/Moderate Reproducibility	protein targets, functionally folded, little PTMs	Antigens tested	Mid/ µLs
Luminex	Up to a hundred proteins /dozens of samples	Low Dynamic Range/Good Reproducibility	protein targets, partially denatured, no PTM's	Antigens tested	High/ µLs
LIPS	One protein at a time/dozens of samples	High Dynamic Range/Good Reproducibility	protein targets, folded though some interference from tag	Antigens tested	Mid/ µLs
phage display random peptide library	~10^9peptides/ one sample at a time	not quantitative, sequences of top hits only	mimotopes of any	motifs sometimes enable prediction of linear epitope	Low/ mLs
T7-Pep/Phip Seq	~10^6 peptides/ several samples at time	Low Dynamic Range/Moderate Reproducibility	36 amino acid protein fragments	linear epitopes	Mid/ µLs
Epitope Peptide Arrays	thousands to tens of thousands of peptides/ dozens of samples	Moderate Dynamic Range/Good Reproducibility	<20 amino acid peptides	linear epitopes	Low/ µLs
Immuno-signaturing	10^4 peptides/ dozens of samples	Moderate Dynamic Range/Good Reproducibility	mimotopes of any	little directly	Low/ µLs

Protein microarrays are a common approach to profiling the humoral immune response, which are solid phase assays utilizing a library of recombinant proteins. Human proteins may be used for autoantibody discovery or pathogen proteins for infectious disease applications (Bacarese-Hamilton, Gray, and Crisanti 2003; Mattoon et al. 2005; Robinson et al. 2002). In the case of infectious disease, the pathogen must be identified and sequenced before a microarray may be designed. If the pathogen has a large proteome, extensive bioinformatics analysis may be done to predict which proteins are most likely to be immunogenic in order to identify a reasonable number to express and purify (Vigil, Davies, and Felgner 2010). For both human and pathogen protein arrays, membrane proteins are particularly challenging, though specialized techniques have been developed to print them (Fang, Frutos, and Lahiri 2002). In either case antibody detection is limited to the antigen set available, including the presence or absence of the appropriate conformation, multimeric state, and post-translational modifications. A variation on the standard protein array called Nucleic Acid Programmable Protein Arrays (NAPPA) spots the DNA encoding the protein, an *in vitro* transcription/translation reaction is performed on the surface, and the protein product is captured on the same spot. This method has better success at making membrane proteins and preserving protein functions than traditional protein arrays (Ramachandran et al. 2008). An approach that retains the post-translational modification involves fractionating the proteins from a cell lysate and spotting the fractions in microarray format (Caiazzo et al. 2011). However, these arrays are limited in their reproducibility and it may not be straightforward to identify the reactive antigen in the fraction (Anderson and LaBaer 2005). Another approach to

detect antibodies to post-translational modifications is to chemically synthesize modified peptides (Kracun et al. 2010; Papini 2009)

An alternative to including all of the possible antibody targets in a library is to detect antibodies through their cross-reactivity. The most common method for identifying molecular mimics of antibody targets is through random peptide library phage display. This method has enabled the discovery of peptides that cross-react with antibodies to many types of targets, including proteins, polysaccharides, and DNA (Meloen, Puijk, and Slootstra 2000). Phage display random peptide libraries typically consists 10^9 different peptide sequences. Through an iterative process of selection and amplification, peptides that bind to the antibody are identified. However, those phage displaying peptides that facilitate growth have a clear advantage, and can out compete those that have peptides that are better binders (Derda et al. 2011). These peptides, known as mimotopes, have also been discovered using smaller synthetic peptide libraries (Meloen, Puijk, and Slootstra 2000). These peptide mimotopes may serve as reagents to detect disease specific antibodies.

Peptide microarrays arrays provide a more efficient alternative to quickly and simultaneously measure the activity of a peptide library. Peptide microarrays have been shown to be a reliable semi-quantitative screening tool for both identifying binders and discovering diagnostic signatures. Fitting binding curves yields a reasonable approximation of SPR affinity measurements (Tapia et al. 2007). Off rates could also be estimated using peptide microarrays using time resolved imaging (Greving et al. 2010). They have been most commonly used to map linear antibody

epitopes, as well as study protein-protein binding, and enzyme substrate interactions (Reimer, Reineke, and Schneider-Mergener 2002). In these cases peptide sequences are typically based on existing protein sequences and substitution variants of those sequences. There has been some use of peptide microarrays for immunodiagnostic applications, either utilizing known epitopes or for epitope discovery (Uttamchandani and Yao 2008). Mimotope peptides discovered from a library selection approach have also shown diagnostic potential

Another mimotope based approach called immunosignaturing utilizes a random-sequence peptide microarray to profile the humoral immune response. By combining the high throughput advantages of the microarray approaches and the unbiased sampling of the phage display library, this approach has the potential to profile the antibody response to any disease. By screening sera directly on a random-sequence peptide microarray, diagnostic patterns may be discovered. The first proof of concept study on influenza has demonstrated that the immunosignature is reproducible and stable over time (Legutki et al. 2010). The immunosignaturing technique has also been successfully applied to distinguish Alzheimer's (Restrepo et al. 2011), valley fever, type 1 diabetes, and several types of cancer (unpublished data). The immunosignaturing approach has some clear advantages in being amenable to processing large numbers of samples for clinical studies. An individual slide costs less than one hundred dollars to produce and process compared to several thousand dollars for commercially available protein arrays. Each slide has two replicate arrays and less than one microliter of sample is required per replicate. The current CIM lab

setup enable the processing of up to 48 samples per day and additional slide processing stations could easily increase that number.

While these protein and peptide library screening approaches may be very useful in discovering novel antibody biomarkers, they are not necessarily ideal assays for use in a clinical setting. A clinical diagnostic requires a higher standard of reproducibility and shouldn't require specialized equipment or skills to run. Additionally, current regulations require that only markers that have shown to be informative are included in the test. ELISA's are standard clinical assays to detect antibodies, but require a large amount of material and are not amenable for multiplexing. Luminex technology requires less sample and antigen than ELISA's and also enables multiplexed measurements. Proteins or peptides are coupled to color coded beads, and antibody binding is detected with secondary labeled with another color. About a hundred different colored beads are available allowing up to a hundred different antigens to be measured simultaneously (Burbelo et al. 2010). While there are some luminex assays that are used as clinical diagnostics, it does require specialized equipment and requires large amounts of purified antigen. Another standard clinical assay is a radiobinding assay, which uses a radiolabeled antigen in an immunoprecipitation assay. This solution phase assay has demonstrated superior sensitivity, specificity, and reproducibility for detecting antibodies to several autoimmune diseases compared to solid phase ELISA assays (Liu and Eisenbarth 2007). However, these assays have declined in use because of the extra precautions and licensing required for using radioactivity. A more recently developed solution phase assay called luciferase immunoprecipitation (LIPS), should

have the advantages in sensitivity and specificity of the radiobinding assay without requiring radioactivity (Burbelo et al. 2007). This assay involves generating recombinant proteins with a luciferase tag for use in an immunoprecipitation assay. LIPS has shown to be amenable to screening small viral proteomes (Burbelo et al. 2011) as well as for adaptation to a microfluidic point of care device (Zubair et al. 2011). Even the most efficient methods of recombinant protein production cannot compare to peptide synthesis in scalability, stability, ease of characterization, and production cost. Perhaps a solution phase assay to measure antibody binding to peptides could be a good option for a clinical immunodiagnostic.

Antibody Target Discovery and Epitope Mapping

When antibodies are discovered that serve as reliable biomarkers of a particular disease, understanding their role in the disease etiological could be of critical importance to developing new preventative or therapeutic strategies. The first step in determining the antibodies role would likely be to identify the molecule that it was generated against. While some of the techniques to profile the antibody response directly uncover the biologically relevant interactions, others simply generate a signature or set of markers. Because of the cross-reactive nature of antibodies, the identification of a molecule that an antibody interacts with does not necessarily imply that the antibody was raised against that molecule. However, there is no straightforward way to observe which antigen actually initiates the antibody response, so identifying the molecules that the antibody interacts with is the logical first step in analyzing the antibody response.

Protein microarrays are an obvious method to identify the target of an unknown antibody response. However, only targets that are present on the array with the relevant conformations and post-translational modifications may be identified. Another common approach is 2D-electrophoresis, immunoblot, mass spectrometry strategy. Proteins from a relevant cell lysate are separated in two dimensions by electrophoresis, transferred to a membrane and probed with the serum of interest. The blotting pattern is compared to a control blot and spots that are different are removed and identified by mass spectrometry. This method is prone to error because the most abundant protein in a spot is not necessarily the antigenic protein. A recent paper has demonstrated that antigens may be identified more accurately if multiple separation methods are used and the overlap between those methods is identified (Mun et al. 2010). This method does not appear to be amenable to processing a large number of clinical samples.

Another group has shown that phage display peptide libraries may be used to identify an antibody response to an unknown pathogen. In this technique known as epitope mediated antigen prediction (E-MAP), antibodies are panned against a random peptide phage display library. Motifs are identified from the peptide sequences and used to search against a protein database (Bastas et al. 2008). This method has been used to identify the target antigen in a multiple myeloma (Sompuram et al. 2008). A more straightforward approach to mapping epitopes also using phage display is to use cDNA library rather than a random peptide library. An important limitation of the cDNA approach is that only a small proportion of the library contains in frame translations of the cDNA. A recent method developed by

Larman *et al.* overcomes this limitation by constructing the library with synthetic oligonucleotides so that all of the protein fragments are designed to be in frame (2011). They also developed a selection scheme that eliminates the traditional rounds of amplification that introduce bias. Phage immunoprecipitation paired with deep sequencing enabled a more efficient identification of binders (Larman et al. 2011). While this approach appears more amenable to high throughput analysis than traditional phage display, the deep sequencing would significantly increase the cost. Both of these phage display techniques will only capture linear epitopes.

Once the antibody target has been identified, it is often of interest to locate the exact binding site, or epitope of the antibody. The most straightforward method to map epitopes is to simply have peptides synthesized tiling the protein sequence and measure binding to each peptide. However, this method is only able to map linear epitopes. Crystallography is probably the most definitive method to map conformational epitopes or those involving post-translational modifications or non-protein targets. However, crystallography is quite time consuming and not always successful. There has been great interest in using phage display random peptide libraries to map conformational epitopes. Many algorithms and software applications have been developed to address the problem of aligning a linear sequence to the structure of a protein, but the accuracy of these methods is still rather low when tested on a benchmark dataset (Sun et al. 2011). These methods are only applicable when the structure of the protein has previously been solved.

The characteristics of antibodies and previous clinical studies presented here demonstrate the promise of antibodies as biomarkers. The immunosignaturing

technology has clear advantages as representing a universal platform for detecting any disease. The cost and standardization of the assay should enable clinical studies of the size necessary to properly validate a biomarker. The major limitation of the immunosignaturing approach is it does not provide a straightforward link to biologically relevant antigens. In this respect, other antibody profiling technologies could be extremely complementary. For example, sera samples identified to have a common immunosignature could be pooled before being used to probe a protein or lysate array. It may also be possible to use the peptides identified in the immunosignature to capture antibodies of a common specificity for further characterization. If a sophisticated bioinformatics method could be developed to predict epitopes from the peptide sequences, it would greatly enhance the information content of the immunosignaturing approach.

Project Overview

A novel approach to profiling the humoral immune response called immunosignaturing was conceived by S. A. Johnston and is being developed by many others in the Center for Innovations in Medicine. The immunosignaturing concept arose from the observation that antibodies are generally not perfectly specific and may cross-react with a variety of unrelated targets. Peptides in particular are adept at molecular mimicry and it is possible to find peptide ligands for antibodies raised against both protein and non-protein targets (Meloen, Puijk, and Slootstra 2000). The current immunosignaturing platform consists of a library of 10,000 random sequence peptides covalently attached to a glass microscope slide.

The peptides are twenty amino acids long with a three amino acid constant linker. The remaining seventeen positions are randomized, with equal frequencies of all amino acids except for Cys, which is used for attachment. The immunosignaturing assay is similar to a miniaturized ELISA: diluted serum is incubated with the microarray and antibody binding is detected with species specific fluorescently labeled secondary antibody. The goal is to identify a common binding pattern among sera samples that are diagnostic of the disease of interest. Preliminary studies in influenza and Alzheimer's have already demonstrated the feasibility of this concept (Legutki et al. 2010; Restrepo et al. 2011). The focus of my project has been to characterize the immunosignaturing platform using monoclonal antibodies with an emphasis on extracting information from the peptide sequences.

The first question that I addressed is how monoclonal antibodies behave on the array and the extent to which monoclonals bind to the random sequence peptides (if at all). Since it will be important to know what types of antibodies may be easier or harder to detect in immunosignaturing experiments, I also looked for trends explaining the differences observed between monoclonals. The concentration of specific antibody clones can be quite low in the sera, so it was important to look at the sensitivity of the assay. The peptide spacing and the effect of antibody bivalency were other issues I examined. The serum contains a mixture of 10^9 different antibodies of different specificities, so it is important to know how antibodies behave in a complex mixture. Mixtures of monoclonal antibodies were used to examine how they may compete. Monoclonal spiked in to an excess of pooled naïve

immunoglobulin simulate detecting a disease specific antibody among normal serum antibodies.

Polyreactive antibodies are an important part of the antibody repertoire, and it is important to know how they contribute to an immunosignature. It has been shown that polyreactive antibodies tend to bind to more targets but with weaker affinity than antibodies raised to a specific target. Chapter 3 addresses how this comparison holds up for random-sequence peptides. I also address how peptides tiling natural protein sequences compare to the random-sequence peptides in binding to the polyreactive antibody.

One advantage of working with peptides is that many of the physiochemical properties can be calculated directly from the peptide sequences. For the purpose of immunosignaturing, it is important to know to what extent the overall properties of the peptides drive binding compared to the extent to which the binding is truly sequence dependent. The ability to discriminate between different antibodies is essential to the immunosignaturing concept, so I also analyze how the properties correlate with sequence specificity (Chapter 4). Since antibodies typically recognize an epitope of 5-12 amino acids in length, there are potentially several antibody binding sites on each peptide. However, the 10K immunosignaturing platform only covers about 5% of all possible 5mer sequences, which does not permit analysis by n-mers of relevant lengths. I did have the opportunity to work with a prototype of a 100K peptide chip with limited amino acid diversity and shorter peptides, which gave me the opportunity to see how short 5mer sequences were able to drive binding.

It would be of great utility to be able to predict the epitope(s) driving the immune response from the immunosignaturing data. In order to address the feasibility of this task, I used a set of known epitope monoclonal antibodies. I first examined whether peptides with sequence similarity to the epitope would be more likely to bind that antibody. Then I identified two approaches to predict the epitopes. The first was a sequence alignment approach using the RELIC program (Mandava et al. 2004) and the second was a motif-based approach using the glam2 program (Frith et al. 2008). I used these results to extrapolate the feasibility of using the random-sequence peptide array data to simply map a monoclonal epitope or to search a database for an unknown antigen.

While the RELIC program was found to have some utility in mapping, it was found to have a number of critical limitations, particularly its availability. The demand for a program to analyze peptide and protein sequences in a similar manner with improved flexibility and statistical analysis, motivated me develop the application GuiTope described in chapter 6. This program is designed to identify regions of similarity between a set of peptides and protein(s) of interest. It allows the user to access to all of the relevant parameters in a graphical user interface, and will estimate the statistical significance of the results. In addition to demonstrating its utility in analyzing my peptide array data, I also show its utility in analyzing phage display data.

References

Allen, C. D., T. Okada, and J. G. Cyster. 2007. Germinal-center organization and cellular dynamics. *Immunity* 27, no. 2: 190-202.

Anderson, Karen S. and Joshua LaBaer. 2005. The sentinel within: Exploiting the immune system for cancer biomarkers. *Journal of Proteome Research* 4, no. 4: 1123-1133.

Arbuckle, M. R., M. T. McClain, M. V. Rubertone, R. H. Scofield, G. J. Dennis, J. A. James, and J. B. Harley. 2003. Development of autoantibodies before the clinical onset of systemic lupus erythematosus. *N Engl J Med* 349, no. 16: 1526-33.

Bacarese-Hamilton, T., J. Gray, and A. Crisanti. 2003. Protein microarray technology for unraveling the antibody specificity repertoire against microbial proteomes. *Curr Opin Mol Ther* 5, no. 3: 278-84.

Bastas, G., S. R. Sompuram, B. Pierce, K. Vani, and S. A. Bogen. 2008. Bioinformatic requirements for protein database searching using predicted epitopes from disease-associated antibodies. *Mol Cell Proteomics* 7, no. 2: 247-56.

Blank, M., O. Barzilai, and Y. Shoenfeld. 2007. Molecular mimicry and auto-immunity. *Clinical Reviews in Allergy & Immunology* 32, no. 1: 111-8.

Braden, B. C. and R. J. Poljak. 1995. Structural features of the reactions between antibodies and protein antigens. *FASEB J* 9, no. 1: 9-16.

Brettschneider, S., N. G. Morgenthaler, S. J. Teipel, C. Fischer-Schulz, K. Burger, R. Dodel, Y. Du, H. J. Moller, A. Bergmann, and H. Hampel. 2005. Decreased serum amyloid beta(1-42) autoantibody levels in alzheimer's disease, determined by a newly developed immuno-precipitation assay with radiolabeled amyloid beta(1-42) peptide. *Biol Psychiatry* 57, no. 7: 813-6.

Burbelo, P. D., K. E. Bren, K. H. Ching, E. S. Gogineni, S. Kottilil, J. I. Cohen, J. A. Kovacs, and M. J. Iadarola. 2011. Lips arrays for simultaneous detection of antibodies against partial and whole proteomes of hcv, hiv and ebv. *Mol Biosyst* 7, no. 5: 1453-62.

Burbelo, P. D., K. H. Ching, E. R. Bush, B. L. Han, and M. J. Iadarola. 2010. Antibody-profiling technologies for studying humoral responses to infectious agents. *Expert Rev Vaccines* 9, no. 6: 567-78.

Burbelo, P. D., K. H. Ching, T. L. Mattson, J. S. Light, L. R. Bishop, and J. A. Kovacs. 2007. Rapid antibody quantification and generation of whole proteome antibody response profiles using lips (luciferase immunoprecipitation systems). *Biochemical & Biophysical Research Communications* 352, no. 4: 889-95.

Caiazzo, R. J., Jr., D. J. O'Rourke, T. J. Barder, B. P. Nelson, and B. C. Liu. 2011. Native antigen fractionation protein microarrays for biomarker discovery. *Methods Mol Biol* 723: 129-48.

Chapman, C., A. Murray, J. Chakrabarti, A. Thorpe, C. Woolston, U. Sahin, A. Barnes, and J. Robertson. 2007. Autoantibodies in breast cancer: Their use as an aid to early diagnosis. *Ann Oncol* 18, no. 5: 868-73.

Chen, S. W., M. H. Van Regenmortel, and J. L. Pellequer. 2009. Structure-activity relationships in peptide-antibody complexes: Implications for epitope prediction and development of synthetic peptide vaccines. *Curr Med Chem* 16, no. 8: 953-64.

Cho-Chung, Y. S. 2006. Autoantibody biomarkers in the detection of cancer. *Biochim Biophys Acta* 1762, no. 6: 587-91.

Colasanti, T., C. Barbati, G. Rosano, W. Malorni, and E. Ortona. 2010. Autoantibodies in patients with alzheimer's disease: Pathogenetic role and potential use as biomarkers of disease progression. *Autoimmunity Reviews* 9, no. 12: 807-11.

Derda, R., S. K. Tang, S. C. Li, S. Ng, W. Matochko, and M. R. Jafari. 2011. Diversity of phage-displayed libraries of peptides during panning and amplification. *Molecules* 16, no. 2: 1776-803.

Erkkila, A. T., O. Narvanen, S. Lehto, M. I. Uusitupa, and S. Yla-Herttuala. 2000. Autoantibodies against oxidized low-density lipoprotein and cardiolipin in patients with coronary heart disease. *Arterioscler Thromb Vasc Biol* 20, no. 1: 204-9.

Fang, Y., A. G. Frutos, and J. Lahiri. 2002. Membrane protein microarrays. *Journal of the American Chemical Society* 124, no. 11: 2394-5.

Frith, M. C., N. F. Saunders, B. Kobe, and T. L. Bailey. 2008. Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Computational Biology* 4, no. 4: e1000071.

Fujinami, R. S., M. B. Oldstone, Z. Wroblewska, M. E. Frankel, and H. Koprowski. 1983. Molecular mimicry in virus infection: Crossreaction of measles virus phosphoprotein or of herpes simplex virus protein with human intermediate filaments. *Proceedings of the National Academy of Sciences of the United States of America* 80, no. 8: 2346-50.

Greving, M. P., P. E. Belcher, C. D. Cox, D. Daniel, C. W. Diehnelt, and N. W. Woodbury. 2010. High-throughput screening in two dimensions: Binding intensity and off-rate on a peptide microarray. *Analytical Biochemistry* 402, no. 1: 93-5.

Hansson, G. K. and A. Hermansson. 2011. The immune system in atherosclerosis. *Nature Immunology* 12, no. 3: 204-12.

Kaufman, E. N. and R. K. Jain. 1992. Effect of bivalent interaction upon apparent antibody affinity: Experimental confirmation of theory using fluorescence photobleaching and implications for antibody binding assays. *Cancer Res* 52, no. 15: 4157-67.

Kijanka, G., S. Ipcho, S. Baars, H. Chen, K. Hadley, A. Beveridge, E. Gould, and D. Murphy. 2009. Rapid characterization of binding specificity and cross-reactivity of antibodies using recombinant human protein arrays. *J Immunol Methods* 340, no. 2: 132-7.

Kracun, S. K., E. Clo, H. Clausen, S. B. Levery, K. J. Jensen, and O. Blixt. 2010. Random glycopeptide bead libraries for seromic biomarker discovery. *J Proteome Res* 9, no. 12: 6705-14.

Lapedes, A. and R. Farber. 2001. The geometry of shape space: Application to influenza. *Journal of Theoretical Biology* 212, no. 1: 57-69.

Larman, H. B., Z. Zhao, U. Laserson, M. Z. Li, A. Ciccia, M. A. Gakidis, G. M. Church, S. Kesari, E. M. Leproust, N. L. Solimini, and S. J. Elledge. 2011. Autoantigen discovery with a synthetic human peptidome. *Nat Biotechnol* 29, no. 6: 535-41.

Legutki, J. B., D. M. Magee, P. Stafford, and S. A. Johnston. 2010. A general method for characterization of humoral immunity induced by a vaccine or infection. *Vaccine* 28, no. 28: 4529-37.

Liu, E. and G. S. Eisenbarth. 2007. Accepting clocks that tell time poorly: Fluid-phase versus standard elisa autoantibody assays. *Clinical Immunology* 125, no. 2: 120-6.

Mandava, S., L. Makowski, S. Devarapalli, J. Uzubell, and D. J. Rodi. 2004. Relic--a bioinformatics server for combinatorial peptide analysis and identification of protein-ligand interaction sites. *Proteomics* 4, no. 5: 1439-60.

Manivel, V., N. C. Sahoo, D. M. Salunke, and K. V. Rao. 2000. Maturation of an antibody response is governed by modulations in flexibility of the antigen-combining site. *Immunity* 13, no. 5: 611-20.

Mattoon, D., G. Michaud, J. Merkel, and B. Schweitzer. 2005. Biomarker discovery using protein microarray technology platforms: Antibody-antigen complex profiling. *Expert Rev Proteomics* 2, no. 6: 879-89.

Meloen, R. H., W. C. Puijk, and J. W. Slootstra. 2000. Mimotopes: Realization of an unlikely concept. *J Mol Recognit* 13, no. 6: 352-9.

Merbl, Y., M. Zucker-Toledano, F. J. Quintana, and I. R. Cohen. 2007. Newborn humans manifest autoantibodies to defined self molecules detected by antigen microarray informatics. *Journal of Clinical Investigation* 117, no. 3: 712-8.

Michaud, G. A., M. Salcius, F. Zhou, R. Bangham, J. Bonin, H. Guo, M. Snyder, P. F. Predki, and B. I. Schweitzer. 2003. Analyzing antibody specificity with whole proteome microarrays. *Nat Biotechnol* 21, no. 12: 1509-12.

Morell, A., W. D. Terry, and T. A. Waldmann. 1970. Metabolic properties of igg subclasses in man. *Journal of Clinical Investigation* 49, no. 4: 673-80.

Mouquet, H., J. F. Scheid, M. J. Zoller, M. Krogsgaard, R. G. Ott, S. Shukair, M. N. Artyomov, J. Pietzsch, M. Connors, F. Pereyra, B. D. Walker, D. D. Ho, D. B. Wilson, M. S. Seaman, H. N. Eisen, A. K. Chakraborty, T. J. Hope, J. V. Ravetch, H. Wardemann, and M. C. Nussenzweig. 2010. Polyreactivity increases the apparent affinity of anti-hiv antibodies by heteroligation. *Nature.* 467, no. 7315: 591-5.

Mun, J., Y. H. Kim, J. Yu, J. Bae, D. Y. Noh, M. H. Yu, and C. Lee. 2010. A proteomic approach based on multiple parallel separation for the unambiguous identification of an antibody cognate antigen. *Electrophoresis* 31, no. 20: 3428-36.

Nell, V. P., K. P. Machold, T. A. Stamm, G. Eberl, H. Heinzl, M. Uffmann, J. S. Smolen, and G. Steiner. 2005. Autoantibody profiling as early diagnostic and prognostic tool for rheumatoid arthritis. *Ann Rheum Dis* 64, no. 12: 1731-6.

Oldstone, M. B. 1998. Molecular mimicry and immune-mediated diseases. *FASEB Journal* 12, no. 13: 1255-65.

Papini, A. M. 2009. The use of post-translationally modified peptides for detection of biomarkers of immune-mediated diseases. *J Pept Sci* 15, no. 10: 621-8.

Radbruch, A., G. Muehlinghaus, E. O. Luger, A. Inamine, K. G. Smith, T. Dorner, and F. Hiepe. 2006. Competence and competition: The challenge of becoming a long-lived plasma cell. *Nature Reviews. Immunology* 6, no. 10: 741-50.

Ramachandran, N., J. V. Raphael, E. Hainsworth, G. Demirkan, M. G. Fuentes, A. Rolfs, Y. Hu, and J. LaBaer. 2008. Next-generation high-density self-assembling functional protein arrays. *Nature Methods* 5, no. 6: 535-8.

Reimer, U., U. Reineke, and J. Schneider-Mergener. 2002. Peptide arrays: From macro to micro. *Curr Opin Biotechnol* 13, no. 4: 315-20.

Restrepo, Lucas, Phillip Stafford, D. Mitch Magee, and Stephen Albert Johnston. 2011. Application of immunosignatures to the assessment of alzheimer's disease. *Annals of Neurology* 70, no. 2: 286-295.

Reuschenbach, M., M. von Knebel Doeberitz, and N. Wentzensen. 2009. A systematic review of humoral immune responses against tumor antigens. *Cancer Immunology, Immunotherapy* 58, no. 10: 1535-44.

Robinson, W. H., C. DiGennaro, W. Hueber, B. B. Haab, M. Kamachi, E. J. Dean, S. Fournel, D. Fong, M. C. Genovese, H. E. de Vegvar, K. Skriner, D. L. Hirschberg, R. I. Morris, S. Muller, G. J. Pruijn, W. J. van Venrooij, J. S. Smolen, P. O. Brown, L. Steinman, and P. J. Utz. 2002. Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nat Med* 8, no. 3: 295-301.

Schroeder, H. W., Jr. and L. Cavacini. Structure and function of immunoglobulins. *Journal of Allergy & Clinical Immunology* 125, no. 2 Suppl 2: S41-52.

Smith, D. J., A. S. Lapedes, J. C. de Jong, T. M. Bestebroer, G. F. Rimmelzwaan, A. D. Osterhaus, and R. A. Fouchier. 2004. Mapping the antigenic and genetic evolution of influenza virus. *Science* 305, no. 5682: 371-6.

Smith, K. G., T. D. Hewitson, G. J. Nossal, and D. M. Tarlinton. 1996. The phenotype and fate of the antibody-forming cells of the splenic foci. *European Journal of Immunology* 26, no. 2: 444-8.

Sompuram, S. R., G. Bastas, K. Vani, and S. A. Bogen. 2008. Accurate identification of paraprotein antigen targets by epitope reconstruction. *Blood* 111, no. 1: 302-8.

Srinivasappa, J., J. Saegusa, B. S. Prabhakar, M. K. Gentry, M. J. Buchmeier, T. J. Wiktor, H. Koprowski, M. B. Oldstone, and A. L. Notkins. 1986. Molecular mimicry: Frequency of reactivity of monoclonal antiviral antibodies with normal tissues. *Journal of Virology* 57, no. 1: 397-401.

Sun, P., W. Chen, Y. Huang, H. Wang, Z. Ma, and Y. Lv. 2011. Epitope prediction based on random peptide library screening: Benchmark dataset and prediction tools evaluation. *Molecules* 16, no. 6: 4971-93.

Tapia, V., J. Bongartz, M. Schutkowski, N. Bruni, A. Weiser, B. Ay, R. Volkmer, and M. Or-Guil. 2007. Affinity profiling using the peptide microarray technology: A case study. *Analytical Biochemistry* 363, no. 1: 108-18.

Tsuchiya, N. and R. C. Williams, Jr. 1992. Molecular mimicry--hypothesis or reality? *Western Journal of Medicine* 157, no. 2: 133-8.

Uttamchandani, M. and S. Q. Yao. 2008. Peptide microarrays: Next generation biochips for detection, diagnostics and high-throughput screening. *Curr Pharm Des* 14, no. 24: 2428-38.

Vigil, A., D. H. Davies, and P. L. Felgner. 2010. Defining the humoral immune response to infectious agents using high-density protein microarrays. *Future Microbiology* 5, no. 2: 241-51.

Vollmers, H. P. and S. Brandlein. 2007. Natural antibodies and cancer. *Journal of Autoimmunity* 29, no. 4: 295-302.

Whittingham, S., S. L. Byron, J. Tuomilehto, P. Z. Zimmet, M. A. Myers, G. Vidgren, M. J. Rowley, S. J. Feeney, P. Koskela, E. Tuomilehto-Wolf, and I. R. Mackay. 1997. Autoantibodies associated with presymptomatic insulin-dependent diabetes mellitus in women. *Diabet Med* 14, no. 8: 678-85.

Zhou, Z. H., A. G. Tzioufas, and A. L. Notkins. 2007. Properties and function of polyreactive antibodies and polyreactive antigen-binding b cells. *Journal of Autoimmunity* 29, no. 4: 219-28.

Zubair, A., P. D. Burbelo, L. G. Vincent, M. J. Iadarola, P. D. Smith, and N. Y. Morgan. 2011. Microfluidic lips for serum antibody detection: Demonstration of a rapid test for hsv-2 infection. *Biomed Microdevices*.

CHAPTER 2

ANTIBODIES AS BIOMARKERS OF HEALTH STATUS

Abstract

Identifying new, effective biomarkers for diseases is proving to be a challenging problem. We have proposed that antibodies may offer a solution to this problem. The physical features and abundance of antibodies make them ideal as biomarkers. Additionally, antibodies are often elicited early in the ontogeny of different chronic and infectious diseases. We reported that antibodies from patients with infectious disease and separately those with Alzheimer's disease display a characteristic and reproducible "immunosignature" on a microarray of 10,000 random sequence peptides. Here we investigate the physical and chemical parameters underlying how immunosignaturing works. We first show that a variety of monoclonal and polyclonal antibodies raised against different classes of antigens produce distinct profiles on this microarray and the relative affinities are determined. A proposal for how antibodies bind the random sequences is tested. Sera from vaccinated mice and people suffering from a fugal infection are individually assayed to determining the complexity of signals that can be distinguished. Based on these results, we propose that this simple, general and inexpensive system could be optimized to generate a new class of antibody biomarkers for a wide variety of diseases.

Introduction

The effort to make medicine preventative should include the development of systems to detect disease before the appearance of major symptoms. The value of early detection is widely accepted and has been the spur to develop new biomarkers of disease that enable earlier diagnosis and treatment. Over 100,000 biomarkers have been reported in the literature to date (Kurian et al. 2009) yet there are only 43 approved by the FDA (Amur et al. 2008) including 19 genomic markers for drug use (FDA 2010). This low return on investment for biomarker discovery suggests that new approaches are needed. Here we characterize a method that has recently been proposed as an alternative for biomarker discovery.

Discovery of biomarkers for early diagnosis of disease poses exceptional demands. For example, in the case of cancer, in order to detect a small number of cells initiating the disease one has to overcome the blood dilution problem. For example, if 10^6 initiating cancer cells release 1000 molecules each of a biomarker into five liters of blood at steady state, the concentration of this biomarker would only be $3x10^{-14}$M. Clearly, it would be an advantage if the response to the biomarker could be amplified. Antibodies are ideal in this sense. An activated B cell produces 5000-20,000 antibodies per minute (Cenci and Sitia 2007; Sulzer et al. 1993) and the cell itself replicates every ~70hr (Cooperman et al. 2004) with a lifespan of up to 4 ½ months (Forster and Rajewsky 1990; Hao and Rajewsky 2001) leading to ~10^{11} amplification of specific signal in a week. Unpurified antibodies are stable in blood, unlike other biomarkers, allowing even historical samples to be used in testing (Geijersstam et al. 1998).

There are three key issues relative to using antibodies as biomarkers of early disease. Do they respond to diseases other than infections? Do they respond early in the course of disease? Can these antibodies be identified with a simple and inexpensive detection system?

There are reports in diabetes (Bonifacio et al. 2000), arthritis (Thurlings et al. 2006), and cancer (Stockert et al. 1998) that the humoral response is activated specifically and early in these chronic diseases. A number of autoantibodies have been identified that appear often years before the disease is diagnosed (Arbuckle et al. 2003b; Hampton 2003; Scofield 2004). In the case of Type I diabetes, antibodies against GAD, IA2 and insulin are found in various combinations well before the onset of clinical disease (Leslie, Atkinson, and Notkins 1999). In patients with paraneoplastic syndrome (PNS), specific neurological symptoms appear years before a cancer is detected (Darnell and Posner 2003; Gultekin et al. 2000; Voltz 2002). The immune response to the nascent tumor reacts with neurons to elicit neurological symptoms (Elrington et al. 1991) that correlate with future tumor appearance. These examples for cancer, diabetes and arthritis also address the second issue: is there an immune response among different individuals that appears early in patients with the same disease? The fact that the same autoantigens, or symptoms in the case of PNS, commonly occur indicates that antibodies might also be consistent across patients.

The third issue, and the one we address here, is how to detect the informative antibodies in an efficient and simple way. Most antibody biomarkers discovered to date were the product of arduous research. Protein microarrays have facilitated this process (Roche et al. 2008) by immobilizing most of the proteins from

a pathogen or human onto a glass slide, but these arrays are expensive, exclude non-transcribed antigens, and are pathogen or auto-antibody specific. The ProtoArray™ v5 of Life Technologies currently has ~9000 unique human proteins spotted and these are being used to discover autoantigens associated with a specific disease. However, only autoantigens are discovered and the cost impedes epidemiology-sized studies. A more complicated approach has been to biochemically fractionated cellular proteins, spot the fractions and react the fractions with the case and control sera (Hanash 2003) and while this system does use authentic material, it is limited by having no control over the relative amounts of the proteins dispensed and require cells from the case subjects' tissue.

Screening for antibody reactivity to random peptides has been generally successful using phage or mRNA display of random 8-12 amino acid sequences. Pasqualini and Ruosiahti (1996) panned a phage library against sera from cancer and healthy subjects to find phage that were preferentially bound by the cancer associated antibodies. This method is unbiased as to the nature of the antigen, and the antibody can be captured (if arduously). Given that random sequence peptides can yield mimotopes of almost any type of antigen (Adda et al. 2002), any disease-associated antigen could theoretically be detected. However promising this method appeared, to date this approach has not produced disease biomarkers for a number of reasons. A serious limitation is that the recurrent panning of the phage is subject to many influences besides just binding of the antibody, nor does this protocol lend itself to processing large numbers of samples quickly.

In order to discover and display relevant antibodies contained in the 10^{10} antibody complexity in the blood simultaneously, we explored a universal technology for antibody biomarker discovery that combines simple, rapid and inexpensive assays from microarrays with the enhanced breadth of the ligand repertoire found in phage-based systems. We created microarrays with only 10,000 random sequence peptides but chose a relatively long length (17 amino acids + 3 residue linker) to allow each peptide to encompass much more complexity than typical epitope peptides. The random sequences allow an unbiased display of antibody binding; the length provides many possible epitope positions per peptide and (potentially) allows for some structural complexity. The array format allows the assay to be run without the biological complications of phage display and high-speed piezo printing onto commercially produced substrate allows several thousands of microarrays to be produced inexpensively per month. We have recently demonstrated the potential utility of these arrays by immunosignaturing vaccines, infections and Alzheimer's disease (Legutki et al. 2010; Restrepo et al. 2011). In order to utilize this technology as a clinical diagnostic, we must first characterize the physical and chemical properties of antibody binding to the peptide microarray.

Materials and Methods

Peptide synthesis and microarray construction

The peptide microarray consists of 10,000 20-residue peptides of 17 random sequence amino acids, with a fixed C-terminal linker of Gly-Ser-Cys-COOH, synthesized by Alta Biosciences, Birmingham, UK. The synthesis scale was 2.5uM

(~1mg total at 75% purity) with 2% of the peptides tested at random by mass spectrometry. Dry peptide was brought up in 100% dimethyl formamide until dissolved, then diluted 1:1 with purified water pH 5.5 + 0.5x PBS pH 7.2 to a final concentration of ~1mg/ml for printing. Gold Seal glass microscope slides were obtained from Fisher (Fair Lawn, NJ, cat# 3010) and treated with aminosilane, activated with sulfo-SMCC (Pierce Biotechnology, Rockford, IL) creating a maleimide-activated surface designed to react with the peptide's terminal cysteine. Spotting was done with a Telechem Nanoprint 60 using 48 Telechem series SMP2 style 946 titanium pins which deposit ~500 pL of peptide per spot. The spotting environment is 25°C, 55% humidity. Fluorescent fiducials are applied asymmetrically using Alexa-647 and Alexa-555-labeled bulk peptides. Slides are stored under argon at 4°C until used. Quality control consists of imaging the arrays by laser scanner (Perkin-Elmer ProScanArray HT, Perkin Elmer, Wellesley, MA) at 647nm to image the spot morphology. Print batches have <30% CV (coefficient of variance) average across all peptides. Data extraction uses GenePix Pro 6.0 (Molecular Devices Inc., Sunnyvale, CA), data analysis uses R and GeneSpring 7.2 (Agilent, Santa Clara, CA).

Binding sample to microarrays

Slides were blocked with 1X PBS, 3% BSA, 0.05% Tween 20, 0.014% β-mercaptohexanol for 1hr at 25°C in a darkened humidified chamber, then sera or antibodies were diluted in 3% BSA, 1X PBS, 0.05% Tween 20 pH 7.2 to a 10nM concentration for monoclonal antibodies or a 1:500 dilution for mouse and human sera, and allowed to bind for 1 hour at 37°C at 20 RPM rotation to the microarray

surface. Later slides (Figures 2.8 and 2.9) were processed using the Tecan HS4800 Pro Hybridization Station using custom programs that mirrored these manual steps. Slides were washed 3x 5' with 1X tris-buffered saline (TBS), 0.05 tween 20 pH 7.2 followed by 3 washes with distilled water. The slides were dried by centrifugation and images were recorded using the Agilent 'C' Scanner at 100% laser power (SHG-YAG laser@532nm or HeNe laser@633nm), 70% PMT.

Antibody Detection

Each antibody or IgG fraction was detected by biotinylated secondary antibody followed by streptavidin-conjugated Alexafluor 555 or 647 (see Table 1 for antibodies used). Secondary antibodies were incubated at a concentration of 5nM, streptavidin at a concentration of 1nM. Single-color experiments were performed exclusively, but dye choice depended on availability. Detection wavelength did not affect resolution, dynamic range or reproducibility.

Table 2.1: List of antibodies used in Figure 2.2, their epitope (if known), the isotype and source.

Protein	**Antibody**	**Epitope**	**Isotype**	**Company**	**Secondary**
Tubulin	DM1A	AALEKD (387-392)	IgG1 kappa	Labvision	Invitrogen HL
p53 (Ab1)	PAb240	RHSVV (212-217)	IgG1	LabVision	Invitrogen HL
p53 (Ab8)	DO-7, BP53-12	DLWKLL (21-26)	IgG2b, IgG2a	LabVision	Invitrogen HL
Interleukin2	LNKB-2	KPLEEVLNL (64-72)	IgG1	Santa Cruz Bio	Invitrogen HL
MHC class I	MHC	3D	IgG1	MBL Int'l	Bethyl
H1N1coat protein	H1N1 1, 2 and 3	Unknown	IgG1	US Bio	Invitrogen HL
Transferrin	HTF-14	N-term of transferrin	IgG1	Abcam	Invitrogen HL
Transferrin	11D3	Unknown	IgG1	Abcam	Invitrogen HL
Transferrin	1C10	Unknown	IgG1	Abcam	Invitrogen HL
2E4	polyreactive		IgM	A. Notkins	Novus
B78	autoantibody	GAD65 protein	IgG1	A. Notkins	Novus
B96	autoantibody	GAD65 protein	IgG1	A. Notkins	Novus
Herceptin	HER2-NEU	Unknown	IgG1	Genentech	Novus
8 pooled	1C10, endorphin, IL2, TP, DM1A, p53AB1, p53Ab8, LNKB2				

Antibody blocking with recombinant protein

Blocking for the dnaX experiment was done by pre-incubating anti-dnaX antisera with His-tagged recombinant dnaX protein for two hours at 25°C. The immune complexes were removed from the solution by incubation of antibody-bound dnaX with nickel Sepharose (Amersham, Piscataway, NJ) for two hours followed by centrifugation. The supernatant was quantitated for protein concentration and processed for binding to the microarrays. Negative control was anti-dnaX antisera incubated with an irrelevant His-tagged protein (a fusion of F1-V from *Yersinia pestis*).

Quality Control

Our microarray manufacturing process requires that 3 slides per 136-slide batch are quality checked using pooled human naïve serum. We examine batch-to-batch correlations extending minimally 5 print batches. Typical array-to-array correlations are >0.93 and batch-to-batch correlations >0.89.

Statistical Analysis

Statistical analysis of microarray data was done using GeneSpring 7.3.1 and R by first importing image-processed data from GenePix Pro 6.0 as gpr text files. Preprocessing of raw data: median normalization and $\log_{10}$ transformation. Statistical tests: Student's T-test or 1-way ANOVA with 5% Family Wise Error Rate multiple testing corrections. The 95th percentile minimum detectable fold-change across three technical replicates averaged ~1.3-fold (Stafford and Brun 2007).

Human Subjects

Human subjects were consented and de-identified according to IRB Protocol# 0905004024, Arizona State University. Blood samples of 5ml were taken as noted in the manuscript.

Animal Care

Use of animals was approved by the Animal Care and Ethics Committee of Arizona State University (IACUC#10-1099R). All animals were anaesthetized using

isofluorane or a cocktail of ketamine, xylazine and acepromazine. Animals were sacrificed with CO_2.

Antibodies and Immunizations:

For the dnaX vaccine (DNA polymerase III subunit gamma/tau [Chlamydophila abortus S26/3]; CAH63776), 9 CBA/J mice were immunized via genetic immunization (Stemke-Hale et al. 2005; Tang, DeVit, and Johnston 1992) with a plasmid construct encoding the open reading frame for dnaX with and without a genetic adjuvant of heat-labile enterotoxin (LTA/LTB). Immune serum was collected 60 days after a prime and two genetic boosts.

Results

Our basic premise is that the antibody profile from an individual reflects their health status. If this profile can be displayed on a sufficiently complex array, the particular responses to chronic diseases will be apparent. We manufactured microarrays onto which 10,000 random-sequence 20-mer peptides were printed. Each peptide is (from NH_3 to COOH termini) 17 residues of any amino acid except cysteine followed by GSC as the linker. The GS amino acids offer rotational freedom and the C-terminal cysteine was used to attach the peptide to the surface through a maleimide linkage (Figure 2.1, right) onto activated aminosilane slides, purchased from Schott, Inc. (Mainz, Germany). Since the peptides have no relationship to any natural sequence (Halperin, Stafford, and Johnston 2011), the same array can be used to profile any disease, any species. The sample is diluted,

applied to the array and allowed to bind. The array is washed and detected with a fluorescently-labeled secondary antibody to the appropriate primary antibody isotype. The array is washed, dried and scanned using a conventional microarray scanner. Figure 2.1 left shows the image of a typical slide: on the left, a naïve individual and right, a day-21 post-seasonal flu vaccine recipient. The insets show peptides that bind differentially.

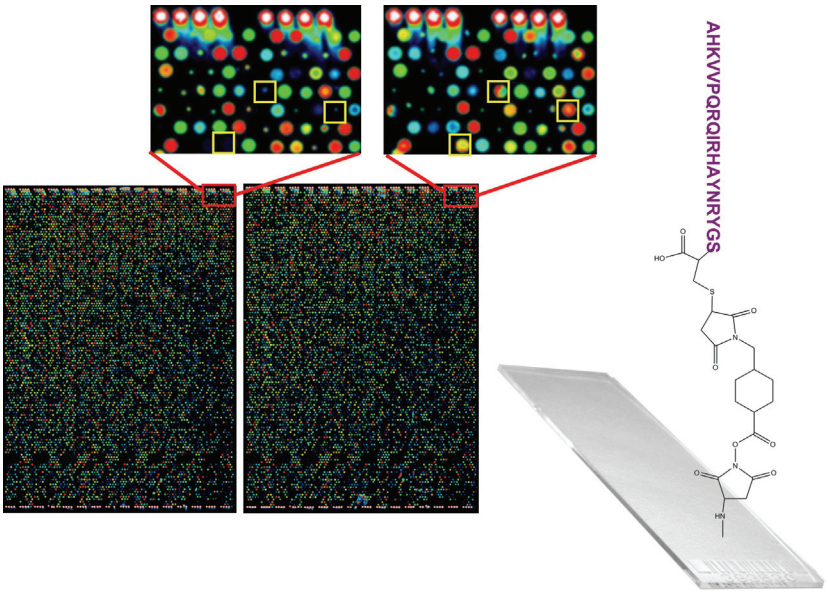


Figure 2.1 Image of the peptide arrays. The microarray is created in a 2-up format, with 10,000 peptides on top and bottom of each slide. In this false-color image, human naïve serum was applied to top the microarray (left), day 21 post-influenza vaccine serum was applied to the bottom (right). The yellow boxes in the small images indicate peptides that show differential binding. Spots are 120um in diameter with intensity values ranging from ~100 to 65,000 relative fluorescence units. Correlation coefficients across technical replicates are typically 0.95 to 0.99. The attachment chemistry is shown on the right with an example peptide attached to the slide through the cysteine to a maleimide linker.

Monoclonal Antibody Profiles

We first asked whether well-characterized antibodies produce discernible profiles on the array and whether those profiles were unique. It has been reported in the literature that monoclonal antibodies do bind to random peptide sequences (Halperin, Stafford, and Johnston 2011; Reineke et al. 2002), however we wished to systematically validate the underlying principals behind this observation. In Figure 2.2 the relative binding of antibodies to a subset of the peptides on the array is portrayed in a heatmap, where blue is low binding and red represents high binding. 272 peptides were selected by ANOVA with a 5% FWER with $p<1x10^{-12}$, representing peptides that were most consistently different across the antibodies listed in the figure legend. Peptides and antibodies are arranged using hierarchical clustering with Euclidean distance as the measure of difference (GeneSpring 7.2.1, Agilent Technologies, Santa Clara, CA). A wide variety of commercial antibodies were used in this experiment including those to phosphorylated or glycosylated proteins, against both conformational and linear epitopes. The important interpretation of this heatmap is that monoclonal antibodies have reproducibly discernible signatures of binding to random peptides. LDA (Linear Discriminate Analysis) using these 272 peptides produced a 0% cross-validation error suggesting that this method could be used to classify different antibodies. This was an important finding and was critical in designing and understanding subsequent experiments. If this finding could be applied to complex mixtures of antibodies, perhaps one signature could be discernible from another when those two antibodies were physically mixed together. We diluted p53Ab1 into p53Ab8 (top of Figure 2.2).

The signatures shown here suggest that the p53Ab1 signature is discernible at a lower concentration than p53Ab8 but there are still a few peptides specific to Ab8 discernible at a near equimolar ratio. The lower heatmap in Figure 2.2 shows the technical replicates of the 10 most differential peptides (X-axis) across these antibodies (Y-axis), at $p<8.23x10^{-23}$. As few as 10 different peptides are able to produce a distinct and reproducible signature that distinguishes this group of antibodies from each other; reproducibility is extremely high.

The actual signals span >3 logs of dynamic range; technical replicates had correlation coefficients 0.92 to 0.99 with an average CV (coefficient of variation) of ~14% and a minimum detectable fold-change of 1.3-fold per 2 replicate arrays at the 95^{th} percentile (Stafford and Brun 2007). We believe the binding to the arrays is largely driven by interaction of the variable region of the antibody and the peptides for two reasons. First, the binding patterns for each antibody was different, even those of the same isotype from the same species. Second, when a directly labeled monoclonal antibody (p53Ab1, IL2 and 11D3 were tested) was competed with 10-fold excess Fc protein, there was no effect on the immunosignature.

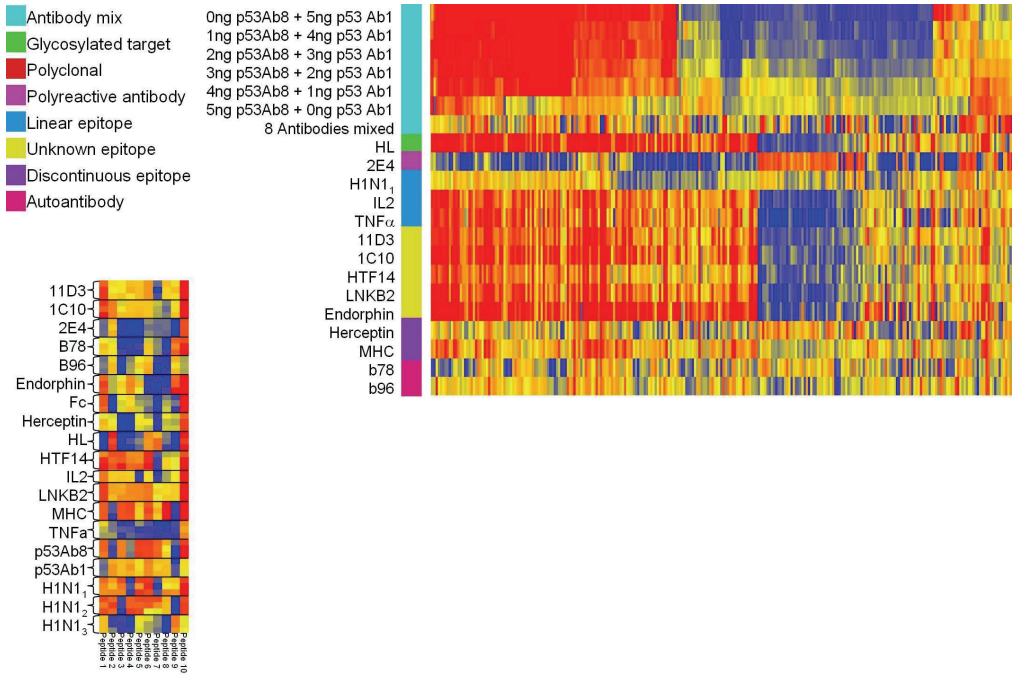


Figure 2.2 Antibodies React with Random Sequence Peptides. Top: A heatmap represents relative binding of antibodies to 267 ANOVA-selected peptides at $p<1x10^{-12}$ (X-axis) relative to their associated antibody (Y-axis). The colored boxes represent the class of epitope for the particular antibody (left). Blue and red in the heatmap indicate low and high binding respectively. The average of three technical replicates is shown per antibody, expect IL2 which had 2. Data is median normalized per array and $\log_{10}$ transformed prior to plotting. Hierarchical clustering is used to group the peptides (X-axis), no grouping was done on the Y-axis. Some proteins have more than one monoclonal antibody represented here (11D3, HTF14 and 1C10 are all against human Transferrin and p53Ab1 and p53Ab8 are both against human TP53). In the large heatmap we examined the outcome of mixing two different antibodies against human TP53: Ab1 and Ab8. The top row shows the p53Ab1 signature; the ratio between Ab1 and Ab8 is reversed until the 6th row which is only Ab8. Ab8 possesses a far less apparent signature than Ab1; there are but a few peptides that recognize Ab1 when any Ab8 is present (~15 peptides to the far right). The next test was whether an equimolar mixture of 8 antibodies (IL2, LNKB2, 11D3, p53A1, H1N1, DM1A, TNFa, and 1C10) would yield a monotonic signature. The '8 Antibodies mixed' row shows reduced signature complexity, but far from monotonic. The poly-reactive antibody 2E4 has low binding overall (Notkins 2004) but binds almost every peptide on the microarray at some level. The significance of these poly-specific antibodies is being investigated (Dimitrov et al. 2010; Halperin, Stafford, and Johnston 2011). Bottom: The small heatmap depicts the three technical replicates per antibody individually plotted using only the 10 most significant peptides at $p<8.23x10^{-23}$. This heatmap indicates the high reproducibility

of the system and the small number of peptides needed to simultaneously discriminate 19 different antibodies with 0% misclassification. Antibodies used: 11D3, HTF14 and 1C10 are against human transferrin; IL2 and LNKB2 are against human Interleukin 2; p53Ab1 and p53Ab8 are against human TP53; b78 and b96 are monoclonal autoantibodies against GAD65; Herceptin is against human HER2/NEU; HL is against a glycosylated target in human cell line HL60; TNF□ is against human TNF-alpha; MHC is against the native human MHC1 complex; $H1N1_{1,2,3}$ are polyclonals against mouse influenza strain PR8; Endorphin is against human endorphin; 2E4 is a poly-reactive antibody(Zhou et al. 2007); Fc is purified constant region from human IgG.

How do the Antibodies Bind the Array?

It may seem surprising that each monoclonal antibody would bind strongly enough to so many different peptides to survive repeated and stringent washings. One would expect that each random peptide would have low affinity to a monoclonal. We have measured the solution phase affinity of particular bound peptides to antibodies using SPR and calorimetry and generally find that the peptide-antibody affinities are in the range of 10-100uM, in line with previous reports (Adda et al. 2002) and sharply contrasted with affinities of 1-10fM for some natural antibody-antigen pairs.

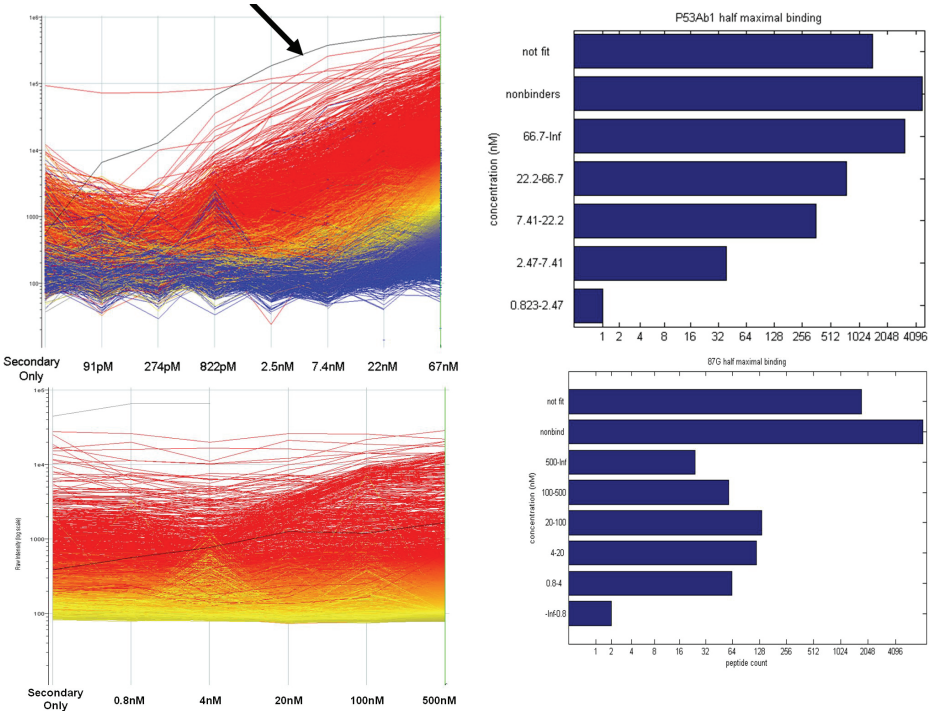


Figure 2.3 Dynamic Range of Antibody Binding.
Top left: Serial dilution of the p53Ab1 monoclonal is shown on the X-axis, relative fluorescence on Y-axis. Each line represents a single peptide colored by its signal at 67nM with red indicating the highest signal, blue the lowest. One peptide (highlighted in black, ETRMIIKLAWETFVDHNGSC) is detected below 100pM (estimated kD). Arrays were $\log_{10}$ transformed. Top right: barchart shows the number of peptides that bind 2 stdev above background at each concentration. 'Not fit' contains peptides that could not be fit with an RSQ>0.8. Peptides that did not bind >2 stdev above background are in the 'not fit' bin. Bottom left: Dilution series of mouse monoclonal anti-HLA-G (clone 87G). Unlike the p53Ab1, only a few peptides show significant binding above below 1nM. An example of a peptide that shows significant binding at 0.8nM is highlighted in black (SREDKDSNDQRKDEQDSGSC). This peptide has an estimated half maximal binding of 3.3nM, suggesting strong apparent affinity. Bottom right: Histogram of half maximal binding concentration for all 10,000 peptides.

From Figure 2.3 we determined that the p53Ab1 antibody appears to have high apparent affinity to many of the peptides on the array, even in the presence of a competing monoclonal. When the p53Ab1 was reacted with the array at various dilutions, one peptide in particular, ETRMIIKLAWETFVDHNGSC (arrow, Figure

2.3, top) demonstrated half maximal binding of 6nM, but it still produced signal detectable an order of magnitude above background at 91pM. Most of the other peptides do not appear to be approaching saturation, so their apparent affinity is likely above 66nM, but over 500 peptides had detectable binding at 822pM. The bottom of Figure 2.3 illustrates a similar experiment where we tested a mouse monoclonal anti-HLA-G (clone 87G). One peptide among a very few shows significant binding at 0.8nM. As opposed to the p53 peptide, the HLA peptide had an estimated half maximal binding of 3.3nM, suggesting that monoclonals may bind random sequence peptides across several orders of magnitude, but we are able to detect even relatively weak binding. We conclude that even though solution affinity of the random peptides for an antibody may be quite low, the apparent affinity for some peptides on the array is very high, presumably due to surface effects (Giraudi et al. 1999).

The most obvious surface effect that might contribute to the amplification of signal on the array would be the high local concentration of the peptides in each spot. We explored the effect of peptide concentration by spotting the peptides on a dendrimeric surface (NSB Postech, Seoul, Korea) where the reactive sites are spaced 9nm apart (NSB27) or 3nm apart (NSB7) (Park 2007). As seen in Figure 2.4, the relative binding on the 3nm surface is on average 30-1000-fold less than on our standard aminosilane surface while the 9nm surface could not support a generally detectable signal. We calculate that in theory the peptides may be as close as ~1nm apart on the microarray based on the density of binding sites of activated aminosilane-coated glass (Kurth and Bein 1993). We conclude that the peptide

density in the spot is contributing to the high relative affinity. For the NSB7 slides, we saw far fewer usable signals than the standard aminosilane slides. When we examined three naïve vs. three PR8 influenza-infected C3H/HEJ mice at day 28, we were unable to statistically distinguish these diseases from each other implying that immunosignaturing needs high signal strength obtained by close packed peptides in order to obtain sufficient discrimination between disease states.

The high density of peptides could lead to high effective affinity by two non-exclusive mechanisms – cooperative binding or avidity (see above). Cooperative binding could arise from two peptides binding one antibody through the interactions with each arm simultaneously. We tested whether bivalent binding was a significant binding mechanism by comparing the binding of an intact monoclonal to its Fab fragment. Overall binding was very similar between the Fab fragment of mouse monoclonal anti-HLA-G clone 87G and the intact IgG (Figure 2.5). Based on this result we conclude that avidity through the high density of peptides in the spot may be sufficient to account for the high relative affinity of the antibodies, but according to our data this effect is dependent to some degree on sequence and to a far lesser degree, on the charge of the peptide during binding.

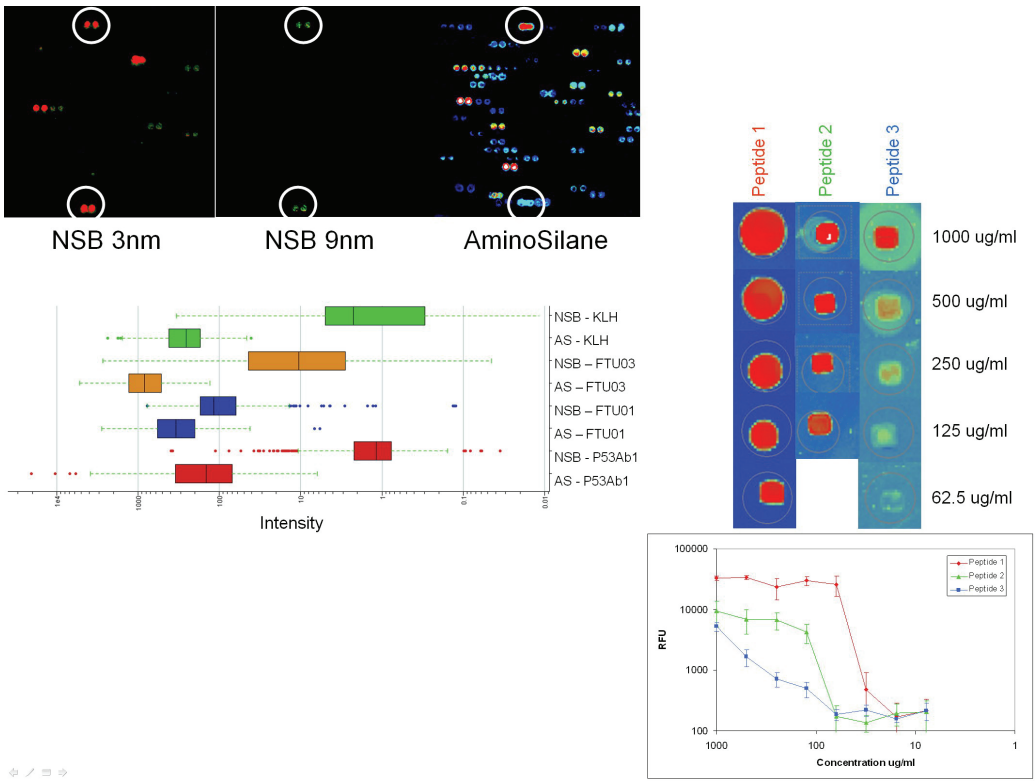


Figure 2.4 Peptide Spacing Impacts the Binding of Antibodies
Peptides were printed on NSB dendrimer slides (NSBPostech, Seoul, Korea) spaced at 3nm (top left), 9nm (middle) and standard aminosilane (right). Colors reflect intensity where white and blue are high binding spots, green indicates mid-level binding and orange to red indicate low binding. The p53Ab1 antibody was allowed to bind to the same 10,000 peptides as on the standard array, but signals notably decreased by 30-1000-fold on the 3nm spacing slide and were almost entirely absent from the 9nm slide. Circles indicate the peptides where signal remained detectable across these 3 different slides. The barchart immediately below indicates peptides that were selected from each of the 3 different experiments (those which bound at least 2 stdev above background). AS = aminosilane, NSB = 3nm spacing. P53Ab1 is a mouse monoclonal against human TP53, FTU01 and FTU03 are peptides from the 10K array which were used as vaccines to immunize BALB/c mice, KLH represents mice immunized with only Keyhole Limpet Hemocyanin adjuvant. Right: several peptides from the 10,000 random peptides were resynthesized to produce a small custom array using dilutions from 1mg/ml to 7.8ug/ml. These small arrays were probed with sera from mice immunized with the same peptide. Shown here are three of these peptides; where there is no image, there was no detectable spot at that concentration. Directly below this image is a log-log plot of the relative fluorescence units measured from three technical replicates. The signals drop off in step with the dilution, but the rate of signal decrease is not constant across all peptides.

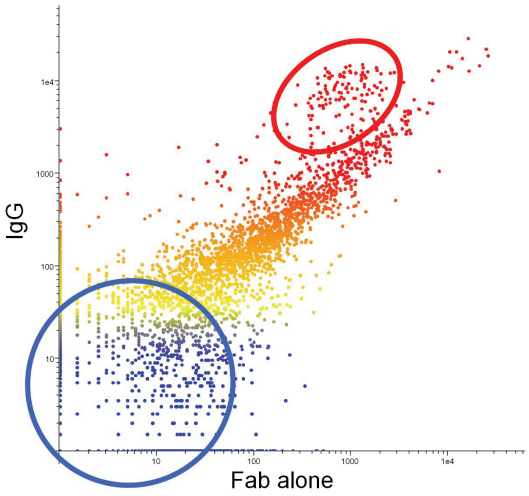


Figure 2.5 Fab Fragment Binds Similarly to Intact IgG. The Fab Fragment and the intact Ig of the same monoclonal (anti-HLA-G clone 87G) were used to probe the peptide array. The signal intensity of the Fab is plotted on the X-axis against the signal of the intact Ig on the Y-axis; the peptides are colored by intensity. The scatterplot shows that most peptides exhibit similar binding to the monovalent and the bivalent forms of the antibody. The red circle highlights peptides with the highest pI, the blue circle contains peptides with the lowest pI values. The differences between the Ig and Fab appear to be driven by the charge of the peptide at pH 7.

Distinguishing Signatures:

A fundamental question underlying this approach is how the mixture of antibodies in serum may interact, or compete, for binding to the random peptides on the array. The observation that each monoclonal antibody we tested binds many different random-sequence peptides implies by simple projection that a collection of antibodies would bind at a generally high level to most peptides on the array. This further suggests that it would be difficult to distinguish a signature of one antibody in a very large collection of different specificities, as in immune serum. If all antibodies recognized the sequence space represented by the random peptides on the

array approximately equally, with 10^{10} specificities in the antibody repertoire, it would seem unlikely that specific antibodies would be recognized at all. An immunosignature of a disease or infection would only be evident if the antibodies produced in response to the disease/infection had higher affinity to the random peptides than the normal immunoglobulins in the sera of healthy people. To test this possibility, we diluted a high affinity commercial monoclonal antibody raised in mice against human TP53 (p53Ab1) into 10X and 100X excess immunolglobulin from healthy volunteers (Figure 2.6). The left panel shows the baseline reproducibility (R=0.97) of two technical replicates. The center panel shows that the p53Ab1 signature is apparent even when diluted by highly complex antibody mixtures, suggesting that antibodies in normal immunoglobulin are not competing for peptide binding sites of the p53 antibody at the relevant concentrations. The panel on the far right shows the contribution of IgG vs. p53 antibody alone. When we mixed 8 monoclonal antibodies together (Figure 2.6) we saw a number of high binding peptides; the naïve human IgG seems to have reached a state where antibodies with strong affinity to specific random-sequence peptides are at a very low concentration. This implies that high affinity antibodies, as would be produced against an infection or chronic disease, would stand out against the background binding of the bulk immunoglobulins in healthy people. This observation is key relative to the immunosignaturing concept and enables the analysis that follows.

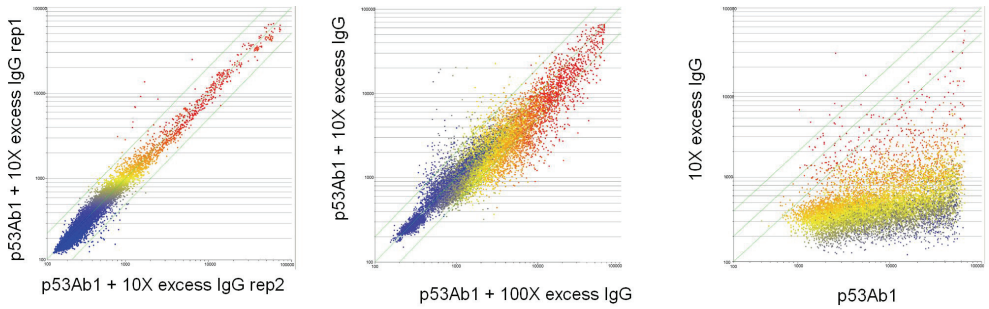


Figure 2.6 Dilution experiment. Left panel shows the baseline reproducibility of the 10,000 peptides on the microarray; correlation coefficient of 0.97 across technical replicates. A mixture of the mouse monoclonal p53Ab1 + 10X excess naïve human IgG was run. The middle panel shows p53Ab1 + 100X excess IgG (X-axis) vs.; p53Ab1 + 10X excess IgG (Y-axis). The correlation coefficient for these two arrays = 0.92. The far right panel shows p53Ab1 monoclonal (X-axis) vs. 10X human IgG alone (Y-axis) correlation coefficient = 0.27.

Analysis of Immunosignatures in a Model System

In order to test the hypothesis that the immunosignature could detect *in vivo* changes in the antibody repertoire, we employed a mouse model. Five mice were bled before and after genetic immunization with a plasmid encoding a protein from *Chlamydia abortus*, dnaX. We had demonstrated earlier that this protein elicits a robust immune response administered as a gene vaccine (Stemke-Hale et al. 2005). In addition, five other mice were immunized with the dnaX plasmid plus a plasmid encoding lethal toxin (LT), a powerful genetic adjuvant (Bowman and Clements 2001). The control mice were mock immunized with plasmid alone, not encoding an antigen. The dnaX-immunized mice on day 14 post immunization showed on average 210 peptides that had significantly ($p<3.31x10^{-9}$) more binding than control mouse serum. A representation of the differences in the arrays is presented in Figure 2.7 (top left). This difference was accentuated when the LT adjuvant was used. We

note that total binding to the array increased upon immunization and even further with immunization with adjuvant, even though the total amount of immunoglobulin was held constant by measure of total IgG. The use of the adjuvant increased binding to peptides that were high binders from the dnaX vaccine alone, as well as a set of new peptides that met the significance cut-off over the controls. Our presumption was that most of this additional binding was driven by antibodies against dnaX, with some against LT alone. In order to test this, the serum from the dnaX immunized mice was adsorbed with beads bearing the dnaX protein and then applied to the array. The control was the same serum absorbed with an irrelevant protein (human Transferrin). As can be seen in Figure 2.7 right, 35 of the 210 dnaX-specific peptides were reduced in intensity by the dnaX adsorption. This indicates that a specific immune signature induced by immunization with dnaX was actually to the dnaX antigen. The peptides that were bound by the dnaX serum but not reduced in intensity by the absorbed serum may have been against the LT adjuvant protein itself. Alternatively, the recombinant dnaX protein may not have presented epitopes that were presented when the protein was made in the mouse cells *in vivo*. For example, antibodies elicited to post translational modifications in the mouse cell would not be presented in the recombinant protein, thus it is possible that most if not all of the signature is to natural dnaX. This result indicates that a specific immune response can be discerned on the random array.

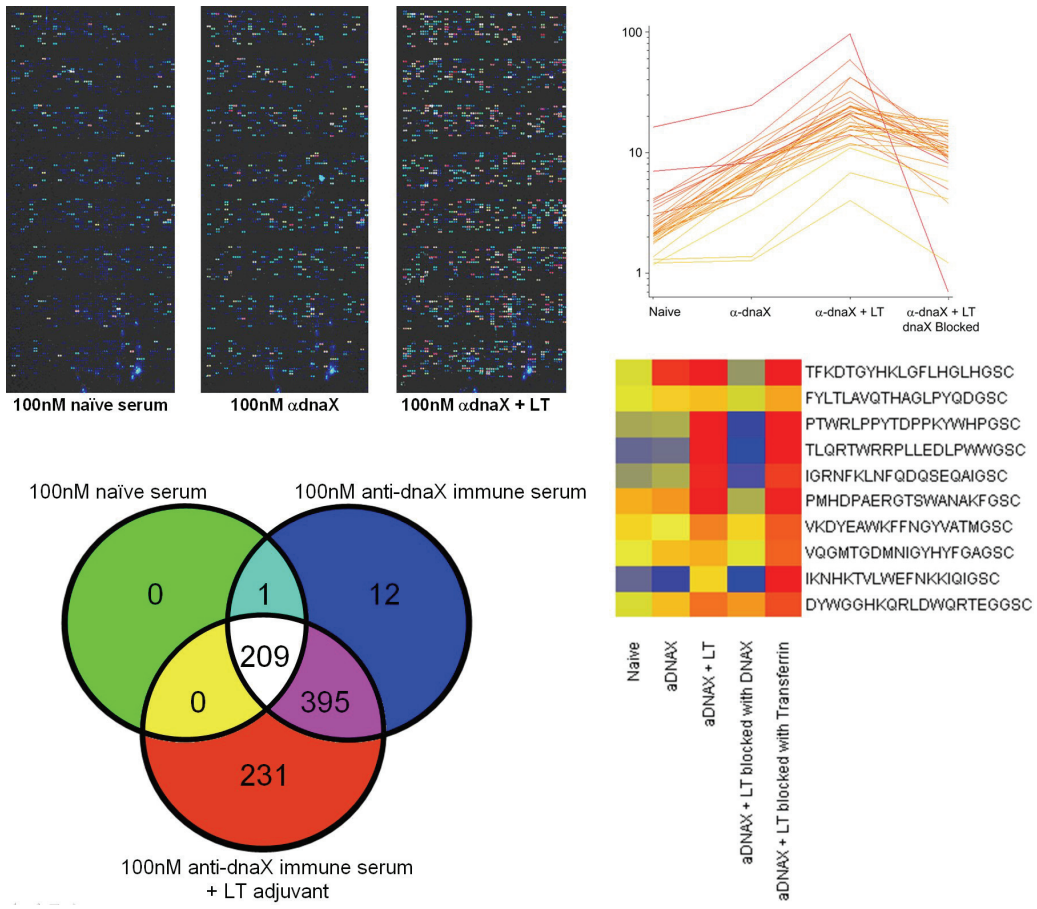


Figure 2.7 Immunological Testing of Random Peptide Binding 9 BALB/c mice were genetically immunized with the coding region to dnaX, a DNA polymerase III subunit found in Chlamydophila abortus S26/3(Stemke-Hale et al. 2005; Tang, DeVit, and Johnston 1992). 60 days following immunization, immune sera was run on the arrays. Top left: images show the peptide arrays as different immune serum is added. As the adjuvant and then the antigen are examined, the total measurable signal on the array increased even though the total amount of IgG remained measurably constant. The Venn diagram immediately below the array images indicates the overlap in the peptides that were 4 stdev above background for each selection. Note the increasing number of peptides selected as the immune response increased. Far right top: line graph showing the peptides that were significantly different between naïve and dnaX-immunized mice at $p<3.31x10^{-9}$. Recombinant dnaX protein produced in *E. coli* an irrelevant human protein, Transferrin, were used to adsorb the immune sera from the dnaX + LT-vaccinated mice. Only the dnaX protein could adsorb the signal.

Immunosignatures in Human Serum

An inbred mouse may have a much simpler repertoire of antibodies than a human. It is possible that this immune complexity in humans would hide the signature of a health-affecting event. To test this possibility we compared the immunosignatures of people with confirmed Valley Fever (elicited by *Coccidiodes immitis*) to the immunosignatures of uninfected control individuals. As seen in Figure 2.8, individuals with Valley Fever have peptides ($p<1.6x10^{-6}$) that are significantly more or less reactive against serum IgG in uninfected controls ('normal donors') or persons who received a seasonal flu vaccine ('day 21 flu vaccine'). These data indicate that in spite of the complexity of the immunoglobulins in humans, it is possible to detect a specific immune response to a health disturbance.

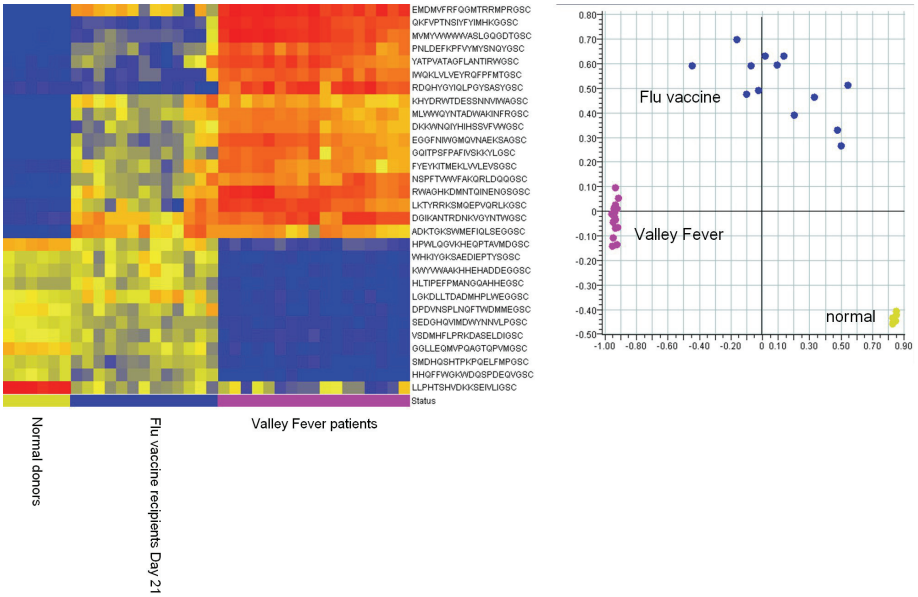


Figure 2.8 Infectious diseases signatured by the peptide microarray Left: Heatmap of 30 ANOVA-selected peptides ($p<1.6x10^{-6}$) classify 6 healthy individuals, 13 day-21 flu vaccine recipients, and 17 Valley Fever patients with 0% misclassification rate using LDA and Leave One Out cross-validation. Right: Scatterplot of the first 2 principal components of the same 30 peptides shows the relative differences between disease states.

Distinguishing a Simulated Multiple Infection

Lastly, we tested the concept raised in the monoclonal experiments, but with more complex polyclonal responses to a vaccine. Could we distinguish two different disease immunosignatures from the same physical sample? This is an important practical consideration since people may have several conditions, such as two infections or chronic or autoimmune disease at the same time. Figure 2.9 demonstrates the ability to distinguish a mixture of two complex 'disease states', simulated here by two different but separate vaccinations in BALB/c mice, and then a physical pooling of equal volumes from each cohort. A double vaccination was not done due to complex interplay within the host, and the desire to rigorously test only the sensitivity parameters of the microarray without imposing additional variances. Either KLH or a random-sequence peptide (PARYANANGRDLITLGIGSC) were used to vaccinate two different groups of 3 mice each. The 6-week immune serum for each was incubated on the 10K microarray, and an additional array was tested with a 50:50 mix. The scatterplots in Figure 2.9 show 30 peptides from KLH ($p<1.04x10^{-8}$ vs. naïve serum) and 30 peptides from PARY-immunized peptide ($p<8.68x10^{-11}$ vs. naïve). Each scatterplot represents the average of the 3 mice, 1 microarray per mouse. The heatmap on the bottom is a visualization of the trend using only the top 30 of the 60 total peptides that by ANOVA discriminate the disease classes ($p<5.24x10^{-18}$).

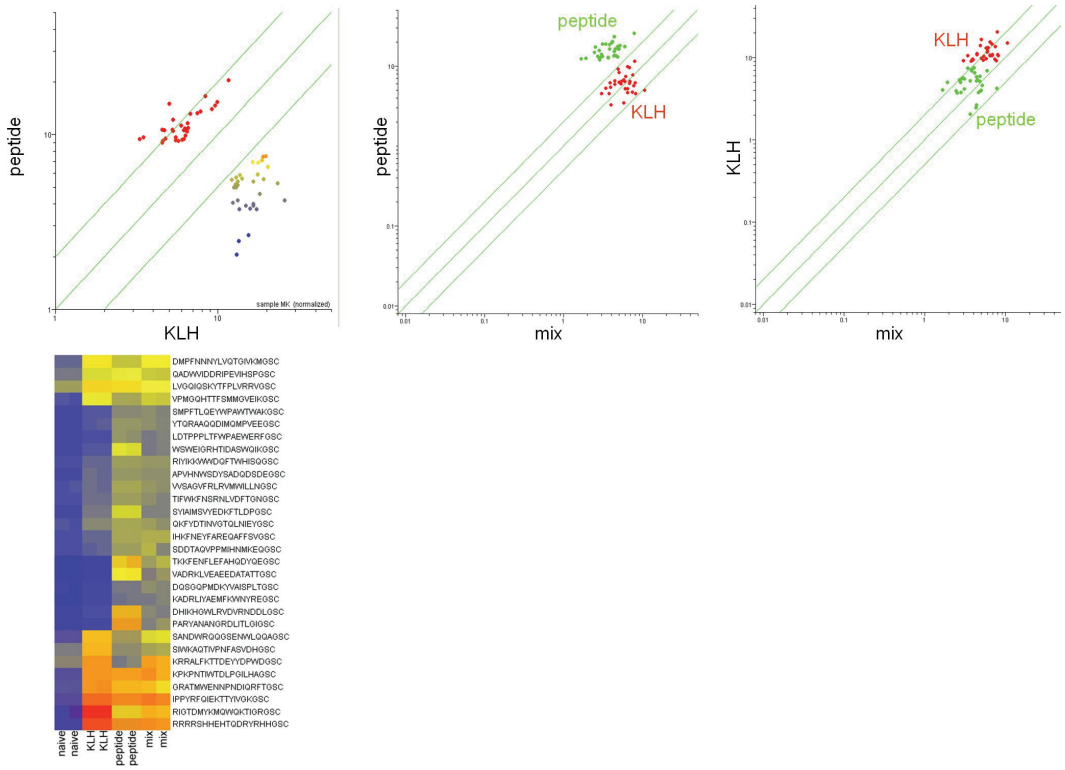


Figure 2.9 Immune sera mixing experiment. Determination of whether different immune sera could be detected when mixed together, simulating conditions where patients have simultaneous infections. Top scatterplots: 3 BALB/c mice were each immunized with either KLH (Keyhole Limpet Hemocyanin) or the peptide PARYANANGRDLITLGIGSC; 6-week post immune serum is used. Peptides that discriminated either KLH-immunized serum (30 peptides at $p<1.04x10^{-8}$ vs. naïve serum) or PARYANANGRDLITLGIGSC-immunized serum (30 peptides at $p<8.68x10^{-11}$) are shown. Far left: The Y-axis shows the PARY-immunized sera, the x-axis shows the KLH-immunized sera (X-axis), peptides are colored by intensity on the array. The intensity differences are quite obvious when contrasting the unadulterated sera. We then physically mixed the two sera and examined additional microarrays. Center: The PARY peptide (Y-axis, green) and mix (X-axis, red) are distinguishable. Far right: the KLH-immunized mouse (Y-axis) and the mix (X-axis) are also distinguishable. The heatmap on the bottom represents a more conventional visualization of the trend: these 30 peptides ($p<5.24x10^{-18}$) were used to plot the values from the two naïve, two KLH, two PARY, and two mix microarrays. The arrays can still distinguish the diseases as distinct using hierarchical clustering. An LDA with Leave One Out cross-validation yields 0% misclassification.

Discussion

We have examined several basic aspects of the immunosignaturing concept using an array of 10,000 relatively long peptides of random sequence. We first showed that all types of monoclonal antibodies tested produced a distinct pattern of binding to these random peptides. This effect has a number of clinical implications, but rather than base a diagnostic on a phenomenon, we investigated the mechanism of binding, providing evidence that the antibody signal we observe is enhanced due to the high peptide density. A notable finding if this technology were to be used as a diagnostic was that the signature of a high affinity antibody was unchanged in the presence of excess immunoglobulin from healthy people. This implies that it may be possible to discern newly developing, high affinity immune responses – responses that evolve presymptomatically in many cases. To test this we compared the serum of mice immunized with a gene vaccine for dnaX to controls and found that there was a clear immunization signal. This general effect has been demonstrated by Merbl *et al.* (2009) in mice with implanted cancer cells, but not on the scale we have demonstrated; the effect was observed but not characterized. A portion of the immunization signature we identified was due to antibodies against the dnaX protein. We demonstrated that signatures could be detected in human sera, showing that people with Valley Fever infections have immunosignatures distinct from non-infected individuals. Finally, we demonstrated that two different disease signatures could theoretically be distinguished in the same person.

It has previously been demonstrated that antibodies bind to peptides of random amino acid sequence on arrays (Halperin, Stafford, and Johnston 2011;

Merbl et al. 2009; Reineke et al. 2002). In general, the feature complexity of arrays reported to date has been less than half the complexity of the peptide microarray used in these studies. The 17aa variable region of 10,000 peptides could encode all possible 3mers and 4mers and 20% of all 5-mers if designed with non-overlapping sequences. We are unable to ascertain whether significant secondary or tertiary structures would exist in these peptides, but calorimetry and *in silico* prediction suggest that the majority of these peptides are not folded or are only transiently non-linear. Even more importantly, these peptides have *no significant similarity to any peptides in on-line sequence databases*. As most antibodies bind regions of 6-11aa it would be unlikely (and confirmed by database searches) that any specific recognition sequence from an actual antigen would be on the array. It is unsurprising then that each antibody would have only weak affinity to a particular peptide, as we have found. Yet, we demonstrated that most of the 18 monoclonal antibody tested bind hundreds of peptides on the array with a dynamic range near 3 logs with high reproducibility. In spite of the weak solution-phase affinity of antibodies to random sequence peptides, the random peptides on our microarray bind in unique and reproducible patterns to most any immunoglobulin molecule.

A number of different classes of monoclonals and affinity purified polyclonal antibodies were tested on the peptide microarray. This included antibodies raised to sugars, proteins modified with phosphates, conformational epitopes and haptens. All tested antibodies produced a distinct and reproducible pattern of binding. Polyreactive antibodies were unusual in that they tended to bind thousands of peptides at a moderate level, while most other monoclonals bound less than 200

different peptides but with relatively high intensity. We investigated whether protein microarrays would work the same way; we tested 11 different monoclonals on a high-quality human protein microarray (Protoarray® from Invitrogen, Carlsbad, CA) and in each case the cognate protein or a near-identical family member had the strongest signal. The peptide microarray described here demonstrates unique and easily discernible signals for antibodies against every antigen tested to date.

It has been reported that aligning random sequence peptides bound by a monoclonal antibody against the protein immunogen could deduce the epitope that elicited the antibody (Burritt et al. 2001). We have also found that this is possible in some circumstances, even with only 10,000 random sequence peptides. By aligning the peptides bound from the dnaX serum using CLUSTAL, it was possible to map discontinuous (2-3 residues) portions of the immunogenic peptides onto the dnaX protein. This method works best when a fairly small protein is used as the immunizing antigen, or if the search space of possible target proteins can be restricted by size, species, or protein family. A monoclonal antibody may bind mimotopes on the array as well as or better than its cognate sequence (Figure 3 in Halperin, Stafford, and Johnston 2011) which complicates this process. This paper shows that the random sequence peptide microarray exhibits a wide range of binding to monoclonal antibodies, but it can also find exact epitopes (Halperin, Stafford, and Johnston 2011). Thus, the immune response to a pathogen or unknown antigens in a chronic disease is unlikely to be precisely discerned by aligning the random sequence peptides bound. The absence of information about the antigen that raised the immune response is a limitation of the immunosignaturing approach. Although

arduous, it *is* feasible to use the high-binding peptides from the microarray to affinity purify antibodies from patient serum which can then be used to identify and isolate the immunogenic antigen from tissue. We accomplished this using a simple murine influenza virus model (Legutki et al. 2010). We also demonstrated that the dnaX protein adsorbed the antibodies that made up the dnaX immunosignature (Figure 2.7). This same approach could be used to test candidate antigens for a given immunosignature. We have found that certain peptides bind to immunoglobulin isotypes specifically. Typically we use pan-isotype secondary antibodies, but we also examined peptides that bound serum from day 21 immunized dnaX mice and found that the IgG1 and IgG2a ratios differed across a fixed range. Thus, one use IgG1:IgG2a ratios as markers for Th1 and Th2 lymphocytes, and other isotypes can be tested on the same array (Legutki et al. 2010) with direct-labeled secondary antibodies labeled with fluorophore. IgM, IgA or IgE have the capacity to be more discerning than IgG in some disease states.

Relative to using these arrays for assaying human serum, we present a preliminary example, but we have found that the approach works well for many types of diseases. We show that people infected with *Coccidiodes immitis* (Valley Fever) have signatures distinct from healthy controls and from people that received the seasonal flu vaccine (Figure 2.8). Any biomarker study would require thousands of cases and controls for validation, but these results suggest that immunosignatures of infected persons are highly consistent even across varied genetic backgrounds and HLA types. This technology may lend itself to very large scale studies since the

arrays are relatively inexpensive to print and can be processed in standard automated systems at ~$1.50 per sample for reagents.

A second favorable feature is that the technology is amenable to utilizing archived samples because of the stability of antibodies and the small amount (~1ul) of serum required. The human samples used in this study were all from stored samples. We tested fresh serum, fresh plasma, frozen serum and frozen plasma from the same volunteer with almost no discernible differences (correlation coefficient > 0.96 at worst). The most high-throughput application would seem to be screening a population with a known disease, selecting the most discriminating peptides, and printing in 24-up format (http://arrayit.com/Products/Microarray_Tools/Multi-Well_Microarrays/multi-well_microarrays.html) enabling a low cost per assay, with robotic washers designed for ELISA plates easily conscripted.

We have noted an unexpected aspect of the immunosignatures. Sera from infected individuals demonstrate generally higher reactivity for some peptides, but also less reactivity relative to normal controls on other peptides (Figure 2.8). Such a difference would not be detectable in standard ELISA assays. In mouse models of infection this phenomenon is due to decreased humoral reactivity over time. In the human samples where only one timepoint is available we cannot readily determine if the lack of some portion of total humoral immune reactivity is a product of the disease response or a precondition for the disease. Either case is testable, and would be of considerable interest to disease specialists.

We envision several potential applications of these arrays. One is the identification of peptides that detect disease-specific biomarker antibodies for clinical

applications. The diagnostic peptides could be used in a printed array format, SPR surfaces, or even solution-phase detection such as calorimetry. Alternatively, the peptides could be used to purify the antibodies that bind to them and then pull down the original antigen which could then be used as a biomarker. It may even be possible to use the microarray described here to continuously monitor healthy individuals for a change in health status in an unbiased manner. We have shown that a person's healthy signature, while often quite different from other healthy signatures, is remarkably self-consistent over time until that person becomes ill or receives a vaccine. After such an immune signal, the immunosignature from multiple individuals becomes extremely homogenous, reflecting that canonical disease signature.

In summary, we present a simple, inexpensive format for profiling the antibody complexity in blood and an investigation of how it works. We anticipate that this format will have broad applicability in research and potentially diagnostics.

References

Adda, C. G., R. F. Anders, L. Tilley, and M. Foley. 2002. Random sequence libraries displayed on phage: Identification of biologically important molecules. *Combinatorial Chemistry and High Throughput Screening* 5, no. 1: 1-14.

Amur, Shashi, Felix W. Frueh, Lawrence J. Lesko, and Shiew-Mei Huang. 2008. Integration and use of biomarkers in drug development, regulation and clinical practice: A us regulatory perspective. *Biomarkers in Medicine* 2, no. 3: 305-311.

Arbuckle, Melissa R., Micah T. McClain, Mark V. Rubertone, R. Hal Scofield, Gregory J. Dennis, Judith A. James, and John B. Harley. 2003. Development of autoantibodies before the clinical onset of systemic lupus erythematosus. *New England Journal of Medicine* 349: 1526-1533.

Bonifacio, E, V Lampasona, L Bernasconi, and AG Ziegler. 2000. Maturation of the humoral autoimmune response to epitopes of gad in preclinical childhood type 1 diabetes. *Diabetes* 49, no. 2: 202-208.

Bowman, Christal C. and John D. Clements. 2001. Differential biological and adjuvant activities of cholera toxin and escherichia coli heat-labile enterotoxin hybrids. *Infect. Immun.* 69, no. 3: 1528-1535.

Burritt, James B., Frank R. DeLeo, Connie L. McDonald, Justin R. Prigge, Mary C. Dinauer, Michio Nakamura, William M. Nauseef, and Algirdas J. Jesaitis. 2001. Phage display epitope mapping of human neutrophil flavocytochromeb 558. *Journal of Biological Chemistry* 276, no. 3: 2053-2061.

Cenci, Simone and Roberto Sitia. 2007. Managing and exploiting stress in the antibody factory. *FEBS Letters* 581, no. 19: 2652-3657.

Cooperman, Jonathan, Robert Neely, David T. Teachey, Stephen Grupp, and John Kim Choi. 2004. Cell division rates of primary human precursor b cells in culture reflect in vivo rates. *Stem Cells* 22: 1111-1120.

Darnell, Robert B. and Jerome B. Posner. 2003. Paraneoplastic syndromes involving the nervous system. *New England Journal of Medicine* 349, no. 16: 1543-1554.

Dimitrov, Jordan D., Cyril Planchais, Jonghoon Kang, Anastas Pashov, Tchavdar L. Vassilev, Srinivas V. Kaveri, and Sebastien Lacroix-Desmazes. 2010. Heterogeneous antigen recognition behavior of induced polyspecific antibodies. *Biochemical and Biophysical Research Communications* 398, no. 2: 266-271.

Elrington, G. M., N. M. Murray, S. G. Spiro, and J. Newson-Davis. 1991. Neurological paraneoplastic syndromes in patients with small cell ling cancer: A prospective survey of 150 patients. *Journal of Neurosurgery and Psychiatry* 54, no. 9: 764-767.

FDA. 2010. Table of valid genomic biomarkers in the context of approved drug labels. http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm accessed Date Accessed)|.

Forster, Irmgard and Klaus Rajewsky. 1990. The bulk of the peripheral b-cell pool in mice is stable and not rapidly renewed from the bone marrow. *Proceedings of the National Academy of Sciences of the United States of America* 87: 4781-4784.

Geijersstam, VeronikaÂ af, Mari Kibur, Zhaohui Wang, Pentti Koskela, EeroÂ Pukka la, John Schiller, Matti Lehtinen, and Joakim Dillner. 1998. Stability over time of serum antibody levels to human papillomavirus type 16. *The Journal of Infectious Diseases* 177, no. 6: 1710-1714.

Giraudi, G., I. Rosso, C. Baggiani, and C. Giovannoli. 1999. Affinity between immobilised monoclonal and polyclonal antibodies and steroid-enzyme tracers increases sharply at high surface density. *Analytica Chimica Acta* 381, no. 2-3: 133-146.

Gultekin, S. Humayun, Myrna R. Rosenfeld, Raymond Voltz, Joseph Eichen, Jerome B. Posner, and Josep Dalmau. 2000. Paraneoplastic limbic encephalitis: Neurological symptoms, immunological findings and tumour association in 50 patients. *Brain* 123, no. 7: 1481-1494.

Halperin, Rebecca F., Phillip Stafford, and Stephen Albert Johnston. 2011. Exploring antibody recognition of sequence space through random-sequence peptide microarrays. *Molecular & Cellular Proteomics*.

Hampton, Tracy. 2003. Autoantibodies predict lupus. *Journal of the American Medical Association* 290: 3186.

Hanash, Sam. 2003. Disease proteomics. *Nature* 422, no. 6928: 226-232.

Hao, Zhenyue and Klaus Rajewsky. 2001. Homeostasis of peripheral b cells in the absence of b cell influx from the bone marrow. *The Journal of Experimental Medicine* 194, no. 8: 1151-1164.

Kurian, Sunil M., Raymond Heilman, Tony S. Mondala, Aleksey Nakorchevsky, Johannes A. Hewel, Daniel Campbell, Elizabeth H. Robison, Lin Wang, Wen Lin, Lillian Gaber, Kim Solez, Hamid Shidban, Robert Mendez, Randolph L. Schaffer, Jonathan S. Fisher, Stuart M. Flechner, Steve R. Head, Steve Horvath, John R. Yates, III, Christopher L. Marsh, and Daniel R. Salomon. 2009. Biomarkers for early and late stage chronic allograft nephropathy by proteogenomic profiling of peripheral blood. *PLoS ONE* 4, no. 7: e6212.

Kurth, Dirk G. and Thomas Bein. 1993. Surface reactions on thin layers of silane coupling agents. *Langmuir* 9: 2965-2973.

Legutki, Joseph Barten, D. Mitchell Magee, Phillip Stafford, and Stephen Albert Johnston. 2010. A general method for characterization of humoral immunity induced by a vaccine or infection. *Vaccine* 28, no. 28: 4529-4537.

Leslie, R. D. G., M. A. Atkinson, and Abner L. Notkins. 1999. Autoantigens ia-2 and gad in type i (insulin-dependent) diabetes. *Diabetologia* 42: 3-14.

Merbl, Yifat, Royi Itzchak, Tal Vider-Shalit, Yoram Louzoun, Francisco J. Quintana, Ezra Vadai, Lea Eisenbach, and Irun R. Cohen. 2009. A systems immunology approach to the host-tumor interaction: Large-scale patterns of natural autoantibodies distinguish healthy and tumor-bearing mice. *PLoS ONE* 4, no. 6: e6053.

Notkins, Abner Louis. 2004. Polyreactivity of antibody molecules. *Trends in immunology* 25, no. 4: 174-179.

Park, Joon Won. 2007. Nsb amine slides. no. http://www.nsbpostech.com/2007/order/order3_detail.html?g_no=34 accessed Date Accessed)|.

Pasqualini, Renata and Erkki Ruosiahti. 1996. Organ targeting *in vivo* using phage display peptide libraries. *Nature* 380: 364-366.

Reineke, U., C. Ivascu, M. Schlief, C. Landgraf, S. Gericke, G. Zahn, H. Herzel, R. Volkmer-Engert, and J. Schneider-Mergener. 2002. Identification of distinct antibody epitopes and mimotopes from a peptide array of 5520 randomly generated sequences. *J Immunol Methods* 267, no. 1: 37-51.

Restrepo, Lucas, Phillip Stafford, D. Mitchell Magee, and Stephen A. Johnston. 2011. Application of immunosignatures to the assessment of alzheimer's disease. *Annals of Neurology* doi: 10.1002/ana.22405.

Roche, Stéphane, Yves Dauvilliers, Laurent Tiers, Carine Couderc, Marie-Thérèse Piva, Monique Provansal, Audrey Gabelle, and Sylvain Lehmann. 2008. Autoantibody profiling on high-density protein microarrays for biomarker discovery in the cerebrospinal fluid. *Journal of Immunological Methods* 338, no. 1-2: 75-78.

Scofield, R. Hal. 2004. Autoantibodies as predictors of disease. *Lancet* 363: 1544-1546.

Stafford, Phillip and Marcel Brun. 2007. Three methods for optimization of cross-laboratory and cross-platform microarray expression data. *Nucleic Acids Research* 35, no. 10: e72.

Stemke-Hale, Katherine, Bernhard Kaltenboeck, Fred J. DeGraves, Kathryn F. Sykes, Jin Huang, Chun-hui Bu, and Stephen Albert Johnston. 2005. Screening the whole genome of a pathogen in vivo for individual protective antigens. *Vaccine* 23, no. 23: 3016-3025.

Stockert, Elisabeth, Elke Jäger, Yao-Tseng Chen, Matthew J. Scanlan, Ivan Gout, Julia Karbach, Michael Arand, Alexander Knuth, and Lloyd J. Old. 1998. A survey of the humoral immune response of cancer patients to a panel of human tumor antigens. *Journal of Experimental Medicine* 187, no. 8: 1349-1354.

Sulzer, Bernhard, J. Leo van Hemmen, Avidan U. Neumann, and Ulrich Behn. 1993. Memory in idiotypic networks due to competition between proliferation and differentiation. *Bulletin of Mathematical Biology* 55, no. 6: 1133-1182.

Tang, De-chu, Michael DeVit, and Stephen A. Johnston. 1992. Genetic immunization is a simple method for eliciting an immune response. *Nature* 356, no. 6365: 152-154.

Thurlings, RM, K Vos, DM Gerlag, and PP Tak. 2006. The humoral response in rheumatoid arthritis and the effect of b-cell depleting therapy. *Ned Tijdschr Geneeskd* 150, no. 30: 1657-1661.

Voltz, Raymond. 2002. Paraneoplastic neurological syndromes: An update on diagnosis, pathogenesis, and therapy. *The Lancet Neurology* 1, no. 5: 294-305.

Zhou, Zhao-Hua, Yahong Zhang, Ya-Fang Hu, Larry M. Wahl, John O. Cisar, and Abner Louis Notkins. 2007. The broad antibacterial activity of the natural antibody repertoire is due to polyreactive antibodies. *Cell Host & Microbe* 1, no. 1: 51-61.

CHAPTER 3

A POLYREACTIVE ANTIBODY RECOGNIZES A LARGE PROPORTION OF BOTH RANDOM AND PROTEIN DERIVED PEPTIDE SEQUENCES

Abstract

Some antibodies found in naïve all individuals recognize a diverse set of antigens with low affinity. These antibodies, known as polyreactive or natural antibodies typically make up 10-20% of the naïve antibody repertoire. Little is known about the epitopes that these antibodies recognize. Classical monoclonal antibodies raised against a particular antigen are typically assumed to have very specific and high affinity interactions with their cognate target. We and others have previously shown that specific antibodies may cross-react with a large number of random-sequence peptides on a microarray. Here we examine how a polyreactive monoclonal antibody and a specific monoclonal antibody compare in recognizing both random and protein derived sequences. The polyreactive antibody was found to recognize several epitopes of two unrelated protein sequences. As expected, the specific antibody only recognized one region of the protein against which it was raised with no cross-reactivity to peptides of an unrelated protein. The specific and polyreactive antibodies also had quite different distributions of random-sequence peptide binding. Several of the random sequence peptides that bound to each antibody were selected for further study. The random peptides selected to bind the specific antibody bound those peptides almost as well it bound its epitope. The polyreactive antibody bound both the protein derived and random sequence peptides

with similar apparent affinity, and was weaker than any of the specific antibody-peptide interactions. The data presented here demonstrates that the concept that polyreactive antibodies recognize more targets but with lower affinity than specific antibodies also applies to random and protein derived peptides.

Introduction

Antibodies are best known for being able to recognize their target antigen with high specificity and sensitivity. However, there is a class of antibodies that bind to diverse sets of antigens, but with poor affinity. These antibodies, known as polyreactive or natural antibodies make up about 10-20% of the naïve antibody repertoire (Chen et al. 1996; Chen et al. 1998; Chen, Wheeler, and Notkins 1995). Little is known about the epitopes recognized by these antibodies. Highly specific antibodies chosen for their selectivity to a single antigen during the maturation process of hybridomas have also been found to cross-react with some unrelated targets, including random-sequence peptides (Halperin, Stafford, and Johnston 2011; Pinilla et al. 1999). Here we will explore how binding of a polyreactive antibody compares to specific antibodies in recognizing protein sequence derived peptide and random-sequence peptides.

The importance of polyreactive antibodies is implied by their evolutionary conservation (Marchalonis et al. 2001) and their abundance early in life (Chen et al. 1998; Madi et al. 2009; Merbl et al. 2007). Their functions *in vivo* involve recognizing both self antigens in the clearing of apoptotic cells (Fu et al. 2007; Litvack et al. 2010; Silverman 2011) and foreign antigens in their non-specific anti-microbial activity

(Zhou et al. 2007). They may also play a role in the pathogenesis of autoimmune diseases. For example, in Lupus patients, it appears that anti-DNA antibodies may have originated from natural IgM polyreactive antibodies that have class switched to IgG (Zhang et al. 2009). A suppressive role of autoreactive natural antibodies in other autoimmune diseases has also been proposed (Shimomura et al. 2008; Silverman 2011). A better understanding of the epitopes that a polyreactive antibody may recognize may lead to better understanding of there function in health and disease.

Polyreactive antibodies are characterized by their ability to bind diverse targets with weak affinity, usually on the order of 10^{-3} to 10^{-7} M (Notkins 2004). Conformational flexibility enables both multi-specificity and decreases the affinity of the interactions. Polyreactive antibodies typically have more flexible paratopes than specific antibodies, and will adopt different conformations when binding different antigens (Mohan et al. 2009; Thorpe and Brooks 2007; Zhou, Tzioufas, and Notkins 2007). These flexible paratopes incur a significant entropic penalty upon binding, which decreases the affinity and slows the on-rate. However, this same penalty allows the antibody to 'probe' and bind to an enormously larger set of protein domains than a typical monoclonal antibody. Since peptides also tend be unstructured, it is not clear whether polyreactive antibodies would be able to recognize flexible peptides or if they are only able to recognize structure regions of proteins.

Antibodies raised against specific antigens have also been found to cross-react with unrelated targets. For example, antibodies raised against viruses have been

found to cross-react with normal tissue (Fujinami et al. 1983; Srinivasappa et al. 1986). It has also been found that peptides may be selected from libraries that mimic conformational and carbohydrate epitopes (Hoess et al. 1994; Kieber-Emmons 1998). While antibodies that recognize a linear epitope usually pull out peptides that have sequence similarity to that epitope, sometimes they also select mimotope peptides with no homology to the original antigen (Stephen, Helminen, and Lane 1995). Previously, it appeared that mimotopes were somewhat rare, since large phage display libraries were used and only a small number of sequences were recovered. However, library selection methods only enrich for the best binders and may be influenced by other factors such as affecting phage growth and viability (Rodi and Makowski 1999). Peptide arrays enable reasonable accurate estimations of relative affinity for smaller libraries (Tapia et al. 2007). Recently, we have shown that monoclonal antibodies raised to specific antigens have measurable binding to a large percentage of random-sequence peptides when they are immobilized on a surface (Halperin, Stafford, and Johnston 2011). Since "specific" antibodies are surprisingly capable of such polyreactivity, seeing how an antibody classically categorized as polyreactive recognizes these peptides should give insight into the recognition capabilities of polyreactive and specific antibodies.

Methods

Hybridoma supernatant from the polyreactive antibody 2E4 was purified using Mannan Binding Protein Columns (Pierce, Rockford) according to the

manufacturer's instructions. Anti-PBEF clone (E-10) was purchased from Santa Cruz Biotechnology.

9500 peptide sequences were generated using a random number generator, with all amino acids but cysteine uniformly distributed at 17 positions with an additional 500 peptides constrained to 45% bias of aromatic groups. An N-terminal "CSG" linker was added to each peptide sequence to allow attachment to the slide through the sulfyhdryl. The C-terminus was amindated to remove the negative charge. The peptides were synthesized by Sigma SIAL (St. Louis, MO). Peptides were resuspended in water, then diluted to an average of 0.5mg/ml in 20mM HEPES, 5mM TCEP, 1mM EDTA, 10% Acetonitrile. Aminosilane coated glass slides (Schott , Jena, Germany) were activated with sulfo-SMCC. Peptides were printed by piezo at AMI (Tempe, AZ) with the entire peptide library replicated on the top and bottom of each slide. Biotinylated peptides were printed at the top and bottom border of each subarray.

Arrays were probed with the anti-PBEF and the polyreactive antibodies, each in triplicate. First, slides were prewashed to remove unattached peptide with a 33% isopropanol, 7.33% acetonitrile, 0.55% TFA solution in nanopure water. All subsequent steps were preformed on the Tecan HS4800 (Switzerland). Slides were blocked with 0.014% mercaptohexanol in 3% BSA PBST for 1hr at 25C. Then 100nM of the antibody in 3% BSA PBST was incubated on the slide for 1hr at 37C, followed by 5nM biotinylated anti-Mouse IgM (Vector, Burlingame, CA in 3% BSA PBST for 1hr at 37C, and finally 5nM Alexa-647 Streptavadin (Invitrogen, Carlsbad, CA) in 3% BSA PBST for 1hr at 37C. A 30 second TBST wash was performed after

each incubation step and a final 30 second water wash was performed before drying the slides with nitrogen for 5min. Slides were scanned in a 'C' Scanner (Agilent, Santa Clara) on the red channel using the 0.1 xdr setting to get scans at 10PMT and 100PMT.

Images were aligned using GenePix Pro (Molecular Devices, Sunnyvale) and the median intensity of each spot was used as the signal in subsequent analysis. Data from the high and low PMT scans were combined as described by García de la Nava, van Hijum, and Trelles (2004). Briefly, peptides whose signal was greater than 200 RFU and less than 50,000 RFU on both scans were used to fit a line. The equation of the line was used to transform the signal intensities from the low scan to the scale of the high scan. A weighted average of the signal intensities of the two scans was calculated, using the percent above background minus the percent saturated as the weights. Data was imported into GeneSpring 7.0 (Agilent) for statistical analysis and peptide selection.

Peptides tiling the sequences of a region of PBEF and AKT1 were designed. Seventeen amino acid segments were taken starting every five amino acids, and the CSG linker was added to the N-terminus to enable directed attachment to the slide. Selected peptides from the 10K library were also included. After synthesis, peptides were normalized to the same concentration, and diluted from 0.8, 0.4, 0.2, 0.1, 0.05, and 0.025mg/ml in 20mM HEPES, 5mM TCEP, 1mM EDTA. Aminosilane coated glass slides were activated with sulfo-SMCC. Peptides were printed using a quill style pin printer. Slides were processed as described above, but varying the primary antibody concentration and the incubation time.

Images were aligned as above, and data was imported into Matlab (Mathworks, Natick, MA). Binding curves were fit to the equation I = imax*(h/(1-h))+b, where I is the signal intensity, imax is the maximum signal intensity, h is the half maximal binding concentration and b is the background signal, using a non-linear least squares regression.

Results

The polyreactive monoclonal IgM antibody 2E4 was used to probe the CIMv2 microarray. A specific antibody of the same isotype (anti-PBEF E-10) was used to probe the array for comparison. Though the anti-PBEF had a high background binding to the slide surface, there were clearly peptide spots well above background. This high background is not unusual among other specific antibodies we have tested. The polyreactive antibody bound to 72% percent of the peptides on the array above background compared to only 52% percent that the specific antibody recognized (figure 3.1). The polyreactive antibody has a greater interquartile range and fewer outliers compared to the anti-PBEF and other specific monoclonals (data not shown).

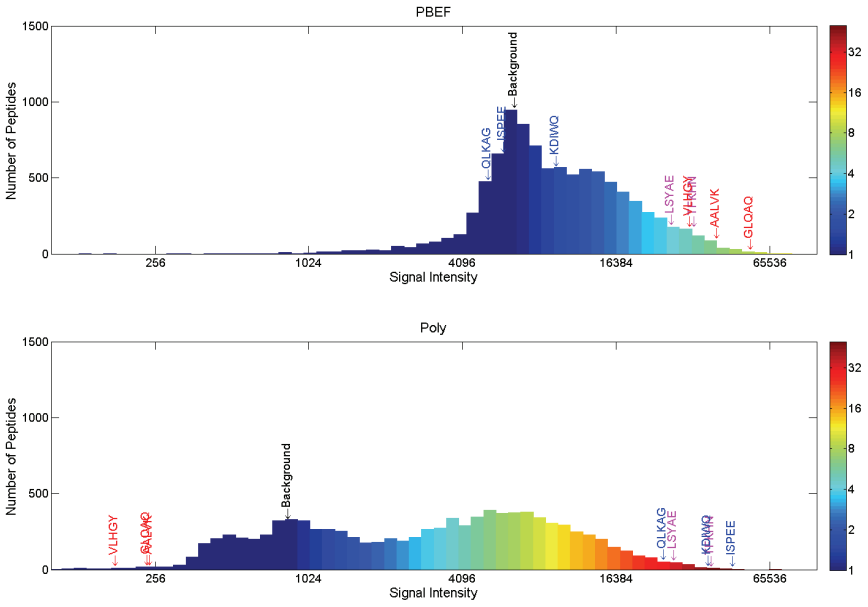


Figure 3.1 Comparison of binding distribution of specific IgM and polyreactive IgM on CIM10K random sequence microarray. Histograms of fluorescence intensities for anti-PBEF and Polyreactive 2E4 are shown. Bars are colored by the ratio of that signal intensity to the background on that slide, which represents binding to the slide surface. Peptides selected for further study are annotated, with those specific for 2E4 in blue, those specific for anti-PBEF in red, and those that are recognized by both antibodies in pink.

Selected peptides that bound to each antibody were re-synthesized for further study. A t-test was used to identify peptides recognized significantly differently by the two antibodies. 594 peptides passed an FDR cut-off of 0.015. Of these peptides the three peptides with the largest fold difference between the two antibodies were selected for each antibody. Peptides recognized by both antibodies were selected by first identifying peptides that bound each antibody significantly greater than the secondary at a FDR of 0.3 (212 peptides), then identify two peptides with the highest geometric mean signal. Together these 12 peptides (table 1) were selected for synthesis.

Table 3.1 List of peptides used in custom array. The peptides shown in blue were selected as 2E4 specific, those shown in red were selected as PBEF specific, and those shown in pink were recognized similarly by both antibodies. The log ratio indicates the log ratio of the signal intensity for anti-PBEF to that of 2E4. The mean indicates the mean signal intensity across both antibodies.

Sequence	Identifier	Log Ratio	Mean	Hydro-pathicity	Net Charge at pH7
CSGKDIWQKH-QDFYMATGHW	KDIWQ	-1.95	18648	-1.01	0.14
CSGISPEEWW-WEDSMPKYQK	ISPEE	-2.94	16374	-1.32	-1.99
CSGQLKAGYP-EYMSNNFPCN	QLKAG	-2.26	11237	-0.72	0.00
CSGGLQAQSG-CVIAILGKRC	GLQAQ	6.42	5863	0.67	2.00
CSGAALVKLF-HLPTSRCQSP	AALVK	6.36	4426	0.26	2.05
CSGVLHGYRK-AIVGLIKKHV	VLHGY	6.08	3876	0.38	4.18
CSGTFKHNMW-VPQYWWATST	TFKHN	-0.21	35152	-0.53	0.96
CSGLSYAEPYF-IPLKTQNHV	LSYAE	-0.03	26943	-0.14	0.05

Peptides were designed from protein sequences to compare to the random-sequence peptides. Ten peptides tiling the region of PBEF recognized by the anti-PBEF IgM E-10 were synthesized along with 21 peptides tiling a region of AKT1 as a negative control. Peptides were designed so that they overlap by 12 amino acids. These protein-sequence derived peptides along with the 12 random-sequence peptides were checked by MALDI that the correct mass was present. These 39 peptides were spotted at six concentrations and the microarrays were probed with anti-PBEF and the polyreactive antibody 2E4 at seven concentrations.

The quality of the microarray array data was assessed and the effect of peptide concentration analyzed. Any replicate measurements with a coefficient of variation greater than 0.5 were discarded. There was no significant difference

between the mean signal intensities or replicate coefficients of variation between peptide spotting concentrations, so these were treated as replicate measurements.

Epitopes were identified based on the peptides that the antibodies recognized. Three overlapping PBEF sequence derived peptides were found to react strongly with the anti-PBEF (E-10). These peptides have a seven amino acid region in common which likely corresponds to the PBEF (E-10) epitope (Figure 3.2). The polyreactive antibody 2E10 binds 43% of the protein sequence derived peptides. Because so many of the peptides are recognized, it is difficult to infer if there are shorter epitope sequences that are driving the binding.

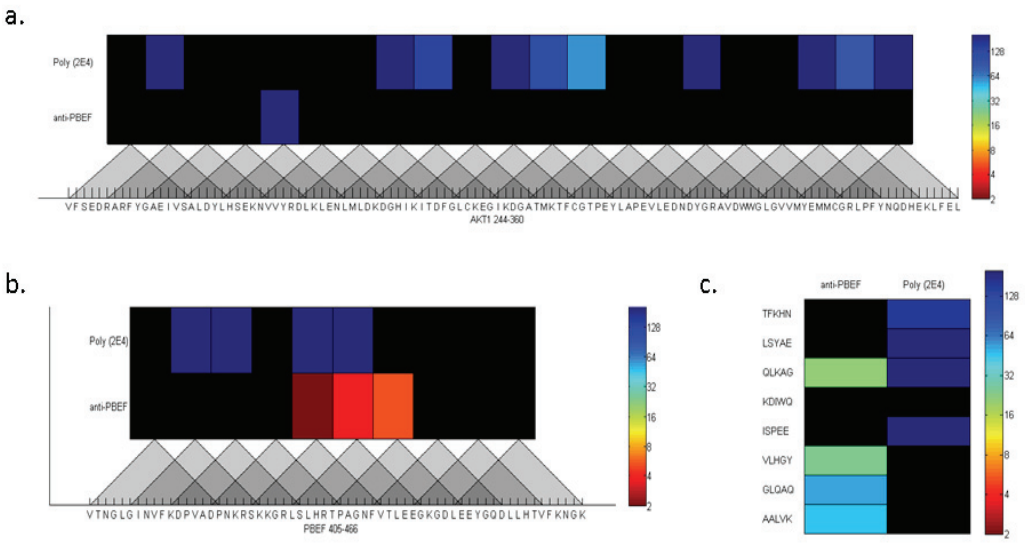


Figure 3.2 Comparison of half maximal binding concentrations for a specific vs. a polyreactive IgM on protein sequence derived and random sequence peptides. Peptides tiling portions of the AKT1 (a) or PBEF (b) were printed along with peptides selected from the 10K random-sequence array (c). Arrays were probed with the polyreactive antibody 2E4 or anti-PBEF at concentrations from 50nM to 0.781nM. The half maximal binding concentration in nM is indicated by the color. Black indicates that the binding curve could not be fit with and RSQ of at least 0.75

Half maximal binding concentration was determined to estimate relative affinity. Curves were fit for 8 peptides for anti-PBEF and 23 peptides for 2E4. The line graphs of the binding curves (Appendix B) show that anti-PBEF recognizes three consecutive PBEF tiling peptides in the 5-20 nM range. It also recognizes four of the five peptides selected from the CIM10Kv2 array in the 20-60 nM range, and one of the peptides selected to bind only to 2E4 at around 40 nM. The polyreactive antibody 2E4 recognized about half of the protein-sequence derived peptides, as well as three out of the five random-sequence peptides selected to bind to it, and one of the peptides selected to bind to PBEF, all at >100nM.

Discussion

Polyreactive antibodies recognize many random-sequence peptides and protein sequence derived peptides, but with low apparent affinity. Compared to the specific monoclonal antibody, the polyreactive antibody had a broad distribution of binding intensities and recognized more peptides above the background threshold. The polyreactive antibody also recognized a similar proportion of the protein sequence derived peptides. Using the concentration at which half maximal binding is observed as a measure of relative affinity, the polyreactive antibody's interactions with any of the peptides were significantly weaker than the specific antibody's interactions with either its epitope or the selected random sequence peptides.

The ability of the specific antibody too cross-react with unrelated targets has important implications for autoimmune disease. A common hypothesis for the origin of self-reactive immunity is that a pathogen has an epitope that mimics a self

antigen and raises an immune response that cross-reacts with the self antigen (Blank, Barzilai, and Shoenfeld 2007; Oldstone 1998; Tsuchiya and Williams 1992). Studies in animal models have suggested that molecular mimicry can play an important role in initiation of autoimmune disease, but self-reactivity alone is not sufficient to cause disease (Christen et al. 2010; Davies 1997; Sfriso et al. 2010). The results from this and previous studies on antibodies binding to random-sequence peptides suggest that even antibodies raised to specific targets are able to recognize a diversity of other sequences. This suggests that antibodies having chance reactivity with self may not be as rare as previously thought.

The cross-reactivity of antibodies can be taken advantage of for diagnostic purposes. Mimotope peptides discovered from a library selection approach such as phage display have also shown diagnostic potential (Casey et al. 2009) but a much faster method was described in Legutki et. al. (2010). By screening sera directly on a random-sequence peptide microarray, diagnostic patterns may be discovered. This approach, known as immunosignaturing, has the potential to diagnosis diseases that involve a humoral immune response. Its promise has been demonstrated with influenza and Alzheimer's disease (Legutki et al. 2010; Restrepo et al. 2011) as well other chronic and autoimmune disease (Williams in prep, Stafford in prep). Natural antibodies have also been shown to have potential as biomarkers (Merbl et al. 2009; Quintana et al. 2004).

The potential for polyreactive antibodies to bind to multiple sites on the same protein may have important implications for how they perform their function *in vivo*. Because polyreactive antibodies have such low affinity, it is difficult to

imagine how they may be important biologically. However, antibody affinity measures describe the interaction of a single binding arm of the Fab with a single antigen, while ignoring that antibodies are adept at multivalent binding. Having multiple binding interactions approaches a multiplicative effect on the apparent affinity. If an antibody is specific to a target, then it could bind multivalently only if that target is spaced appropriately, such as on the surface of a virus or bacteria where repeating subunits are the norm. The ability to bind multiple targets makes it much more likely to be able to find binding sites at the appropriate spacing for multivalent binding. For example, Moquet et al. recently showed that an anti-HIV spike protein antibody also has low affinity for the viral membrane. Being able to bind to the spike protein with one arm and the membrane with the other greatly enhances its affinity for viral particles (Mouquet et al.2010). The data presented here show that not only can polyreactive antibodies bind multiple targets, but they could potentially bind multiple sites on the same target, greatly enhancing its binding possibilities.

Reference

Blank, M., O. Barzilai, and Y. Shoenfeld. 2007. Molecular mimicry and auto-immunity. *Clinical Reviews in Allergy & Immunology* 32, no. 1: 111-8.

Casey, J. L., A. M. Coley, K. Parisi, and M. Foley. 2009. Peptide mimics selected from immune sera using phage display technology can replace native antigens in the diagnosis of epstein-barr virus infection. *Protein Engineering, Design & Selection* 22, no. 2: 85-91.

Chen, Z. J., F. Shimizu, J. Wheeler, and A. L. Notkins. 1996. Polyreactive antigen-binding b cells in the peripheral circulation are igd+ and b7. *European Journal of Immunology* 26, no. 12: 2916-23.

Chen, Z. J., C. J. Wheeler, W. Shi, A. J. Wu, C. H. Yarboro, M. Gallagher, and A. L. Notkins. 1998. Polyreactive antigen-binding b cells are the predominant cell type in the newborn b cell repertoire. *European Journal of Immunology* 28, no. 3: 989-94.

Chen, Z. J., J. Wheeler, and A. L. Notkins. 1995. Antigen-binding b cells and polyreactive antibodies. *European Journal of Immunology* 25, no. 2: 579-86.

Christen, U., E. Hintermann, M. Holdener, and M. G. von Herrath. 2010. Viral triggers for autoimmunity: Is the 'glass of molecular mimicry' half full or half empty? *Journal of Autoimmunity* 34, no. 1: 38-44.

Davies, J. M. 1997. Molecular mimicry: Can epitope mimicry induce autoimmune disease? *Immunology & Cell Biology* 75, no. 2: 113-26.

Fu, M., P. S. Fan, W. Li, C. X. Li, Y. Xing, J. G. An, G. Wang, X. L. Fan, T. W. Gao, Y. F. Liu, and S. Ikeda. 2007. Identification of poly-reactive natural igm antibody that recognizes late apoptotic cells and promotes phagocytosis of the cells. *Apoptosis* 12, no. 2: 355-62.

Fujinami, R. S., M. B. Oldstone, Z. Wroblewska, M. E. Frankel, and H. Koprowski. 1983. Molecular mimicry in virus infection: Crossreaction of measles virus phosphoprotein or of herpes simplex virus protein with human intermediate filaments. *Proceedings of the National Academy of Sciences of the United States of America* 80, no. 8: 2346-50.

Garcia de la Nava, J., S. van Hijum, and O. Trelles. 2004. Saturation and quantization reduction in microarray experiments using two scans at different sensitivities. *Stat Appl Genet Mol Biol* 3, no. 8: Article11.

Halperin, Rebecca F., Phillip Stafford, and Stephen Albert Johnston. 2011. Exploring antibody recognition of sequence space through random-sequence peptide microarrays. *Molecular & Cellular Proteomics* 10, no. 3.

Hoess, R. H., A. J. Mack, H. Walton, and T. M. Reilly. 1994. Identification of a structural epitope by using a peptide library displayed on filamentous bacteriophage. *Journal of Immunology* 153, no. 2: 724-9.

Kieber-Emmons, T. 1998. Peptide mimotopes of carbohydrate antigens. *Immunologic Research* 17, no. 1-2: 95-108.

Legutki, J. B., D. M. Magee, P. Stafford, and S. A. Johnston. 2010. A general method for characterization of humoral immunity induced by a vaccine or infection. *Vaccine* 28, no. 28: 4529-37.

Litvack, M. L., P. Djiadeu, S. D. Renganathan, S. Sy, M. Post, and N. Palaniyar. 2010. Natural igm and innate immune collectin sp-d bind to late apoptotic cells and enhance their clearance by alveolar macrophages in vivo. *Molecular Immunology* 48, no. 1-3: 37-47.

Madi, A., I. Hecht, S. Bransburg-Zabary, Y. Merbl, A. Pick, M. Zucker-Toledano, F. J. Quintana, A. I. Tauber, I. R. Cohen, and E. Ben-Jacob. 2009. Organization of the autoantibody repertoire in healthy newborns and adults revealed by system level informatics of antigen microarray data. *Proceedings of the National Academy of Sciences of the United States of America* 106, no. 34: 14484-9.

Marchalonis, J. J., M. K. Adelman, I. F. Robey, S. F. Schluter, and A. B. Edmundson. 2001. Exquisite specificity and peptide epitope recognition promiscuity, properties shared by antibodies from sharks to humans. *Journal of Molecular Recognition* 14, no. 2: 110-21.

Merbl, Y., R. Itzchak, T. Vider-Shalit, Y. Louzoun, F. J. Quintana, E. Vadai, L. Eisenbach, and I. R. Cohen. 2009. A systems immunology approach to the host-tumor interaction: Large-scale patterns of natural autoantibodies distinguish healthy and tumor-bearing mice. *PLoS ONE [Electronic Resource]* 4, no. 6: e6053.

Merbl, Y., M. Zucker-Toledano, F. J. Quintana, and I. R. Cohen. 2007. Newborn humans manifest autoantibodies to defined self molecules detected by antigen microarray informatics. *Journal of Clinical Investigation* 117, no. 3: 712-8.

Mohan, S., K. Kourentzi, K. A. Schick, C. Uehara, C. A. Lipschultz, M. Acchione, M. E. Desantis, S. J. Smith-Gill, and R. C. Willson. 2009. Association energetics of cross-reactive and specific antibodies. *Biochemistry* 48, no. 6: 1390-8.

Mouquet, H., J. F. Scheid, M. J. Zoller, M. Krogsgaard, R. G. Ott, S. Shukair, M. N. Artyomov, J. Pietzsch, M. Connors, F. Pereyra, B. D. Walker, D. D. Ho, D. B. Wilson, M. S. Seaman, H. N. Eisen, A. K. Chakraborty, T. J. Hope, J. V. Ravetch, H. Wardemann, and M. C. Nussenzweig. 2010. Polyreactivity increases the apparent affinity of anti-hiv antibodies by heteroligation. *Nature.* 467, no. 7315: 591-5.

Notkins, A. L. 2004. Polyreactivity of antibody molecules. *Trends in Immunology* 25, no. 4: 174-9.

Oldstone, M. B. 1998. Molecular mimicry and immune-mediated diseases. *FASEB Journal* 12, no. 13: 1255-65.

Pinilla, C., R. Martin, B. Gran, J. R. Appel, C. Boggiano, D. B. Wilson, and R. A. Houghten. 1999. Exploring immunological specificity using synthetic peptide combinatorial libraries. *Curr Opin Immunol.* 11, no. 2: 193-202.

Quintana, F. J., P. H. Hagedorn, G. Elizur, Y. Merbl, E. Domany, and I. R. Cohen. 2004. Functional immunomics: Microarray analysis of igg autoantibody repertoires predicts the future response of mice to induced diabetes. *Proceedings of the National Academy of Sciences of the United States of America* 101 Suppl 2: 14615-21.

Restrepo, Lucas, Phillip Stafford, D. Mitch Magee, and Stephen Albert Johnston. 2011. Application of immunosignatures to the assessment of alzheimer's disease. *Annals of Neurology*: n/a-n/a.

Rodi, D. J. and L. Makowski. 1999. Phage-display technology--finding a needle in a vast molecular haystack. *Current Opinion in Biotechnology* 10, no. 1: 87-93.

Sfriso, P., A. Ghirardello, C. Botsios, M. Tonon, M. Zen, N. Bassi, F. Bassetto, and A. Doria. 2010. Infections and autoimmunity: The multifaceted relationship. *Journal of Leukocyte Biology* 87, no. 3: 385-95.

Shimomura, Y., E. Mizoguchi, K. Sugimoto, R. Kibe, Y. Benno, A. Mizoguchi, and A. K. Bhan. 2008. Regulatory role of b-1 b cells in chronic colitis. *International Immunology* 20, no. 6: 729-37.

Silverman, Gregg J. 2011. Regulatory natural autoantibodies to apoptotic cells: Pallbearers and protectors. *Arthritis & Rheumatism* 63, no. 3: 597-602.

Srinivasappa, J., J. Saegusa, B. S. Prabhakar, M. K. Gentry, M. J. Buchmeier, T. J. Wiktor, H. Koprowski, M. B. Oldstone, and A. L. Notkins. 1986. Molecular mimicry: Frequency of reactivity of monoclonal antiviral antibodies with normal tissues. *Journal of Virology* 57, no. 1: 397-401.

Stephen, C. W., P. Helminen, and D. P. Lane. 1995. Characterisation of epitopes on human p53 using phage-displayed peptide libraries: Insights into antibody-peptide interactions. *Journal of Molecular Biology* 248, no. 1: 58-78.

Tapia, V., J. Bongartz, M. Schutkowski, N. Bruni, A. Weiser, B. Ay, R. Volkmer, and M. Or-Guil. 2007. Affinity profiling using the peptide microarray technology: A case study. *Analytical Biochemistry* 363, no. 1: 108-18.

Thorpe, I. F. and C. L. Brooks, 3rd. 2007. Molecular evolution of affinity and flexibility in the immune system. *Proceedings of the National Academy of Sciences of the United States of America* 104, no. 21: 8821-6.

Tsuchiya, N. and R. C. Williams, Jr. 1992. Molecular mimicry--hypothesis or reality? *Western Journal of Medicine* 157, no. 2: 133-8.

Zhang, J., A. M. Jacobi, J. Wang, R. Berlin, B. T. Volpe, and B. Diamond. 2009. Polyreactive autoantibodies in systemic lupus erythematosus have pathogenic potential. *J Autoimmun.* 33, no. 3-4: 270-4. Epub 2009 Apr 26.

Zhou, Z. H. , A. G. Tzioufas, and A. L. Notkins. 2007. Properties and function of polyreactive antibodies and polyreactive antigen-binding b cells. *J Autoimmun.* 29, no. 4: 219-28. Epub 2007 Sep 20.

Zhou, Z. H., J. Zhang, Y. F. Hu, L. M. Wahl, J. O. Cisar, and A. L. Notkins. 2007. The broad antibacterial activity of the natural antibody repertoire is due to polyreactive antibodies. *Cell Host Microbe.* 1, no. 1: 51-61.

CHAPTER 4

THE EFFECTS OF PEPTIDE PHYSIOCHEMICAL PROPERTIES AND SEQUENCE COMPLEXITY ON ANTIBODY BINDING

Introduction

Antibodies are of critical importance for understanding response to infectious disease and vaccine development (Amanna and Slifka 2011), in the diagnosis and pathology of autoimmune disease (Eggert, Zettl, and Neeck 2010), and are increasingly of interest in cancer (Anderson and LaBaer 2005), Alzheimer (Colasanti et al. 2010), and atherosclerosis (Hansson and Hermansson 2011). A better of understanding of the forces driving antibody recognition would have implications for all of these fields. Peptides often serve as surrogates for antibody targets in the laboratory. They are frequently used to map linear epitopes (Carter and Loomis-Price 2004). Antibodies to a wide range of other types of targets are often found to bind to peptides, usually derived from a random peptide library (Meloen, Puijk, and Slootstra 2000). These targets include haptens (Kalaycioglu, Russell, and Howard), glycans (Kieber-Emmons 1998), and tertiary and quaternary protein structures (Denisova, Denisov, and Bramson 2010). In most peptide library approaches, only a small number of peptides identified to bind to a given antibody are characterized. Here I will use several random sequence peptide microarrays to analyze trends in antibody recognition of peptides.

Previous studies on antibody peptide interactions have focused on the small number of antibody-peptide pairs for which crystal structures are available. A recent

survey of these structures identified a number of trends in the peptides, including an enrichment of coil and turn secondary structures, and a bias in amino acids toward His, Asp, and Pro (Chen, Van Regenmortel, and Pellequer 2009). The same study also found that residues in the bound peptide would not be accessible in the native antigen in most cases (Chen, Van Regenmortel, and Pellequer 2009). Other studies have looked in detail at the mechanisms of molecular mimicry in mimotope binding. In some cases, the peptide truly does mimic the interactions of the cognate epitope, but in other cases the mimotope peptide contacts with the paratope than those made by the cognate epitope (Dudak, Boyaci, and Orner 2011). In a particularly interesting example of the multifarious nature of mimotope interactions, two monoclonal antibodies to identical epitopes select non cross-reactive mimotopes (Xu et al. 2004). While these structural studies highlight the versatility of recognition by the paratope, they do not offer any principle to predict antibody binding.

Attempts predict likely B-cell epitopes from protein sequences or structures have met with limited success. Some sequence properties that have been identified to be correlated with B-cell epitopes, including hydrophilicity (Parker, Guo, and Hodges 1986), flexibility (Karplus and Schulz 1985), beta-turn propensity (Pellequer, Westhof, and Van Regenmortel 1993), and solvent accessibility (Emini et al. 1985). The frequency of amino acid occurrence in epitopes was used to develop an antigenicity scale (Kolaskar and Tongaonkar 1990). However, predictions based on any amino acid scale have yielded prediction accuracy only slightly better than chance when evaluated on a large enough dataset (Blythe and Flower 2005). As more epitope data have become available for training, various machine learning

approaches have been developed that have improved accuracy over the propensity scales alone (Blythe and Flower 2005). When a crystal structure of the antigen is available, other methods that take into account geometric shape may have improved prediction over sequence based methods alone (El-Manzalawy and Honavar 2010). However, to date the best prediction methods are only achieving AUC values of less than 0.7 (Liang et al. 2010; Zhang et al. 2011). Improvements are expected as more high quality epitope information becomes available to train on and better prediction methods are developed, but there are likely some inherit limitation to predicting any antibody epitope from the common features.

While identifying the epitope of an antibody can be very useful, particularly for vaccine design, the native antigen may not be the most efficient method for detecting a particular antibody. While technologies exist for producing proteins with mammalian post-translational modifications, they are still quite labor intensive and costly (Walsh 2010). Glycans are a particularly important post-translational modification that antibodies target, and they may also be chemically synthesized, though their synthesis poses a continuing challenge for chemists (Kajihara et al. 2010). Conformational epitopes may not be detected in standard solid phase assays, so careful attention must be paid to assay design to be able to detect these important epitopes (Butler 2000). Alternatively, peptides may serve as molecular mimics for both conformational and glycan epitopes (Meloen, Puijk, and Slootstra 2000). Peptide synthesis is much more cost effective and scalable than glycan synthesis or any protein production method (Vlieghe et al. 2010). Additionally, peptides are amenable to use in standard solid phase assays, including microarray formats

(Uttamchandani and Yao 2008). While most mimotope peptides have previously been discovered using phage display techniques, peptide microarrays can also be an efficient tool for mimotope discovery (Reineke et al. 2002). Peptide microarrays are an efficient method to obtain semi-quantitative binding data on a large number of peptide simultaneously (Tapia et al. 2007).

Several technologies exist for creating peptide microarrays. One of the earliest developed was SPOT synthesis. In this technique, drops of activated amino acids are deposited at predetermined locations on the surface, and in this manner the peptides are synthesized on the surface. While many studies have utilized spot arrays to screen peptides for a variety of activities, they are limited in the density of peptides that can be synthesized and have substantial batch to batch variability (Volkmer 2009). Other peptide arrays, such as the ones employed in this study involve having peptides synthesized using conventional peptide synthesis techniques and spotting them on an activated surface. This enables an improvement in miniaturization and reproducibility over the spot synthesis methods, but still does not approach the current density that is routine in DNA microarrays. Another *in situ* technique that has better miniaturization compared to SPOT arrays is the light directed method originally described by Fodor et. al. and now commercially produced by LCSciences (Fodor et al. 1991). More recently, a technique similar to a laser printer has been developed to make peptide microarrays, called pepperprint (Breitling et al. 2009). These arrays can have around a tenfold higher peptide density than conventional spotted arrays, but in our hands, they lack sensitivity in detecting antibody binding (unpublished data). Here I will present data from a chip synthesized using a

photolithography technique that has 100,000 peptides on a 11mm by 11mm silicon wafer.

Methods

The CIM10Kv1 production was first described in Morales et al (2009). Briefly, 10K random 17mer sequences containing equal probability of the 20 amino acids except cysteine were generated. A C-terminal linker of Gly-Ser-Cys was added to each sequence to make 20mer peptides. These peptides were synthesized by Alta Bioscience (Birmingham, UK), diluted in 25% DMF 75% PBS, and spotted on a Telechem Nanoprint quill type contact printer onto sulfo-SMCC activated aminosilane coated glass slides. Antibodies were used to probe these arrays a described previously (Halperin, Stafford, and Johnston 2011). Briefly, slides were processed using Agilent chambers (Agilent Technologies, Santa Clara CA). Slides were blocked with 0.014% mercaptohexanol in 3% BSA PBST for 1hr at 25C. Then 100nM of the antibody in 3% BSA PBST was incubated on the slide for 1hr at 37C, followed by appropriate biotinylated secondary antibodies in 3% BSA PBST for 1hr at 37C, and finally 5nM Alexa-647 Streptavadin (Invitrogen, Carlsbad, CA) in 3% BSA PBST for 1hr at 37C. Three TBST washes followed by three water washes were performed between each step with manual agitation. Slides were scanned on the ScanArray (Perkin Elmer, Waltham, MA). Sources of antibodies are listed in table 1.

Table 4.1 Monoclonals summary. Antibodies used in this chapter are listed by the name used here, and clone names are shown in parenthesis. If the epitope has been mapped it is listed as linear; the others are unknown.

Antibody Name	**Target**	**Target Type**	**Species/Isotype**	**Source**
11D3	Transferrin	Protein	Mouse IgG1	Abcam
1C10	Transferrin	Protein	Mouse IgG1	Abcam
TP Cocci	coccidioidomycosis TP Antigen	Glycan	?	D. M. Magee
AbcamHA (16B12)	Influenza HA tag	peptide	Mouse IgG1	Abcam
Endorphin (B31.15)	Beta-endorphin	peptide	Mouse IgG1	Abcam
HTF14	Transferrin	Protein	Mouse IgG1	Abcam
Herceptin	HER2/neu	Protein	Human IgG1	D. Lake
IL2	IL-2	Protein	Mouse IgG1	Peprotech
LNKB2	IL-2	Protein (linear)	Mouse IgG1	Santa Cruz
MHC	MHC-beta2 microglobulin complex	Protein (quartenary structure)	?	D. Lake
P53Ab1 (pAb240)	P53	Protein (linear)	Mouse IgG1	Labvision
P53Ab8 (BP53-12)	P53	Protein (linear)	Mouse IgG2a	Labvision
TNF	TNF-alpha	Protein	Mouse IgG1	Peprotech
DM1A	Tubulin-alpha	Protein (linear)	Mouse IgG1	Labvision
b78	GAD65	Protein	Human IgG1	ATS
b96	GAD65	Protein	Human IgG1	ATS
cMyc (9E10)	cMyc tag	Peptide	Mouse IgG1	AbD Serotec
V5	V5 Tag	Peptide	Mouse IgG1	AbD Serotec
LeuEnk	Leu-Enkephalin	Peptide	Mouse IgG1	AbD Serotec
PBEF (E-10)	PBEF	Protein	Mouse IgM	Santa Cruz
2E4	Polyreactive	Diverse	Mouse IgM	A. Notkins

The CIM10Kv2 was also designed with random 17mers with the same 19 amino acids at equal frequencies. An N-terminal Cys-Ser-Gly linker was added and the C-terminus was modified to an amine group. The peptides were synthesized by Sigma SIAL (St. Louis, MO). Peptides were resuspended in water, then diluted to an average of 0.5mg/ml in 20mM HEPES, 5mM TCEP, 1mM EDTA, 10% Acetonitrile. Aminosilane coated glass slides (Schott , Jena, Germany) were activated

with sulfo-SMCC. Peptides were printed by piezo at AMI (Tempe, AZ) with the entire peptide library replicated on the top and bottom of each slide. Biotinylated peptides were printed at the top and bottom border of each subarray. After printing, slides were prewashed to remove unattached peptide with a 33% isopropanol, 7.33% acetonitrile, 0.55% TFA solution in nanopure water. The antibody probing assay was performed similarly to the assay on the CIM10Kv1 except that the process was largely automated using the Tecan HS4800 (Switzerland). Slides were blocked with 0.014% mercaptohexanol in 3% BSA PBST for 1hr at 25C. Then 100nM of the antibody in 3% BSA PBST was incubated on the slide for 1hr at 37C, followed by appropriate biotinylated secondary antibodies in 3% BSA PBST for 1hr at 37C, and finally 5nM Alexa-647 Streptavadin (Invitrogen, Carlsbad, CA) in 3% BSA PBST for 1hr at 37C. A 30 second TBST wash was performed after each incubation step and a final 30 second water wash was performed before drying the slides with nitrogen for 5min. Slides were scanned in a 'C' Scanner (Agilent, Santa Clara) on the red channel using the 0.1 xdr setting to get scans at 10PMT and 100PMT.

The BioMicroChips were purchased from HealthTel (?) Chips were processed in a 12 well tissue culture plate. All steps were performed at room temperature and the plate was set on a rotating platform for mixing. The chip was pre-incubated with 3% BSA PBST for 1hr before probing with 10nM monoclonal in 3% BSA PBST for 1hr, followed by 5nM Alexa-647 labeled anti-Mouse IgG in 3% BSA PBST for 1hr. Six two minute washes in 1% BSA TBST were performed after each incubation, and a final three two minute washes in deionized water were performed before drying with 40PSI nitrogen nozzle. The chips were scanned in the

Innoscan 900 (Innopsys, Carbonne, France) scanner using custom made slide adapters at 50PMT low laser power on the red channel.

Images from all three platforms were aligned in Genepix Pro 6.0 (Molecular Devices, Sunnyvale, CA) using appropriate GAL files. Streaks, particulates, and other artifacts were visually identified and flagged. Flagged spots were treated as missing data in subsequent analysis. The spot medians were used as the signal intensities. Data was imported into Matlab R2010b (Mathworks, Natick, MA) for further analysis. The contribution of the secondary antibody was subtracted from raw signal intensity based on secondary only arrays. The global background intensity for each slide was estimated based on the average signal intensity in between the spots, and the signal intensities were normalized by taking the ratio to the background intensity. The normalized signal of any spot at or below the average background was set to one and the data was log transformed. Normalized signals were averaged across replicate arrays. Specificity was calculated based on finding the pearson correlation between the signal intensities and a profile corresponding to having a signal of one in that antibody and zero in the others.

Sequence properties were calculated for all peptides using Matlab. The GSC (CIM10Kv1) or CSG (CIM10Kv2) linker sequences were not included in calculating sequence properties. The isoelectric point was calculated using the function from the matlab bioinformatics package. The average residue volume (Fauchere et al. 1988), beta turn propensity (Chou and Fasman 1978), flexibility (Vihinen, Torkkila, and Riikonen 1994), and hydrophilicity (Parker, Guo, and Hodges 1986) scales were obtained from the AAindex database (Kawashima et al. 2008). The accessibility

(Emini et al. 1985) and antigenicity (Kolaskar and Tongaonkar 1990) scales were obtained from the Immune Epitope DataBase antibody epitope prediction page (Zhang et al. 2008).

Results

Three unique peptide libraries were created in microarray format. The key differences between the microarrays are summarized in table 2. The first microarray called CIM10Kv1.0 has 10,000 random-sequence peptides with seventeen randomized positions containing equal frequencies of all amino acids except Cys and a constant C-terminal Gly-Ser-Cys linker. The peptides were contact printed on aminosilane coated microscope slides that had been activated with a sulfo-SMCC linker to covalently link the peptide through the sulfhydryl group of the Cys. The CIM10Kv2 was created in a very similar many to the first version with a few key differences. The peptides were designed with an N-terminal Cys-Gly-Ser so that the peptides would be immobilized in the opposite orientation and the C-terminus was modified to an amine group to preserve the charge at the free end of the peptide. The spotting buffer was changed to a more aqueous buffer to in order to reduce the printing of hydrophobic peptide synthesis byproducts. Piezo printing was adopted because it was able to print two replicate arrays on one physical slide and was able to print more slides in a shorter time period. Reproducibility was also improved from contact printing (data not shown).

Table 2. Peptide microarray platform comparison. The two 10K microarrays are fairly similar, differing only in the peptide orientation, spotting buffer, and printing method. The BioMicroChip are made using a completely different technology, allowing for many more features in a smaller array. The first version of the BioMicroChip has only 12mer peptides made from eight different amino acids. [1]X indicates randomized positions. On the 10K arrays, the peptides are immobilized through the Cys (C-terminal on version 1 and N-terminal on version 2). The BioMicroChip has a polyethelene glycol (Peg) linker. [2]Hydrophobic amino acids are shown in black, hydrophilic neutral amino acids are shown in blue, negatively charged amino acids are shown in red, positively charged amino acids are shown in green, and aromatic amino acids are underlined.

Microarray	**CIM10Kv1.0**	**CIM10Kv2.0**	**BioMicroChip v1.0**
Peptide Design[1]	XXXXXXXXXX XXXXXXGSC	CSGXXXXXXX XXXXXXXXX	XXXXXXXXXX X-(Peg)
Number of peptides	10,000 Peptides	10,000 Peptides	100,000 Peptides
Amino acids in randomized positions[2]	GAVLIMFWP STYNQDEKRH	GAVLIMFWP STYNQDEKRH	GFWPEKHY
Spotting Buffer	DMF/PBS	HEPES/TCEP	N/A
Printing Method	Contact Printed	Piezo Printed	Synthesized in Situ
Array Substrate	Aminosilane coated glass slide	Aminosilane coated glass slide	Silicon Wafer
Physical Dimensions	75mm x 25mm	75mm x 25mm	11mm x 11mm

The BioMicroChip represents an entirely different type of peptide microarray technology than the CIM10K arrays. The peptides were synthesized *in situ*, which removes any potential effects of differences in peptide immobilization efficiency. The stepwise yield of the synthesis was approximately X% so the features should have approximately X% full length peptide. The synthesis method also enables a much greater miniaturization, with ten times as many peptides contained in about 13% of the area. These chips contain shorter peptides generated from fewer amino acids than the 10K libraries.

Eight different properties were calculated from the peptide sequences. The distributions of those sequence properties were examined for each of the peptide

library (Figure 4.1). Note that four of the properties (beta turn, accessibility, flexibility, and hydrophilicity) are moderately correlated with each other (Pearson>0.6). The two 10K library had very similar distributions of all properties. The 100K library has dramatically lower sequence complexity compared to the 10K libraries as a result of only containing eight amino acids. In fact, the 95% complexity score of the 100K library was lower than the 5% complexity score on the 10K library.

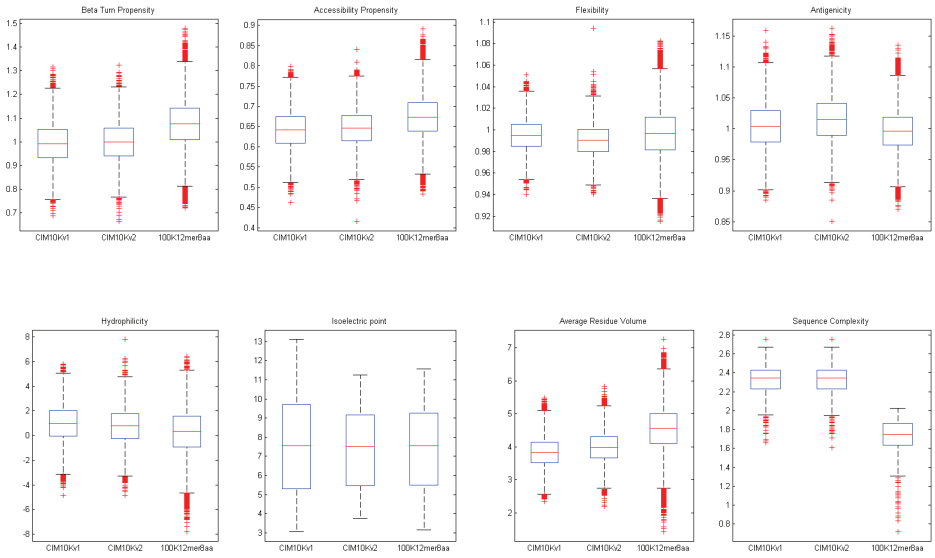


Figure 4.1 Platform comparisons of properties distributions. The distributions of eight different sequence properties across the three random peptide libraries are shown as boxplots above.

Monoclonal antibodies were found to bind to a significant fraction of all three peptide libraries. False color images of a representative array are shown in Figure 4.2 for the CIM10Kv2 and BioMicroChip. Representative images of the CIM10Kv1 were published previously (Halperin, Stafford, and Johnston 2011). The number of peptides bound varied greatly between monoclonals with the percent

above background ranging from 0.1% to 86% on the 10K platforms. The BioMicroChip did not have any non-printed area to be able to estimate the signal intensity of the background so the number of spots above background could not be estimated. However, the chips demonstrated a large dynamic range spanning three logs of fluorescence intensity.

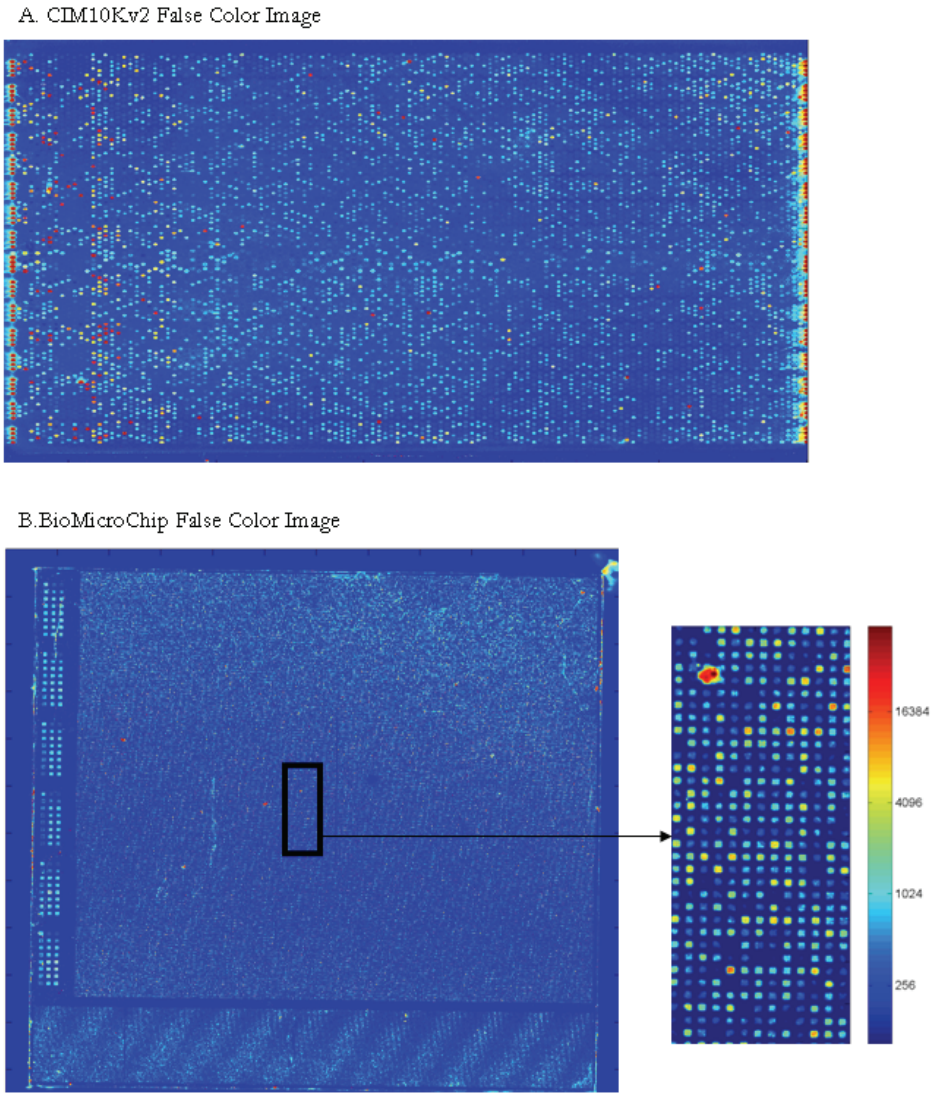


Figure 4.2 False color microarray images. False color images generated in Matlab of the P53Ab1 antibody probing the CIM10Kv2 (A) or BioMicroChip (B).

The correlation of the signal intensity was examined between each monoclonal antibody and each peptide property (Figure 4.3A). On the CIMv10Kv1, the largest correlations are seen with isoelectric point, those these vary greatly between antibody (mean 0.03, range: -0.47 to 0.38). Binding is generally negatively correlated with flexibility (mean -0.08, range: -0.29 to 0.18), accessibility, (mean -0.07, range -0.19 to 0.02) and hydrophilicity (mean -0.07, range: -0.25 to 0.18). Slight positive correlations were observed for sequence complexity (mean 0.05, range -0.01 to 0.20). These trends are quite different when examined on the CIMv10Kv2. Isoelectric point still showed the most variability among antibodies, but here the correlation was generally positive (mean 0.17, range: -0.1 to 0.46). Correlations with beta turn propensity (mean 0.15, range -0.02 to 0.33), flexibility (mean 0.14, range -0.03 to 0.30), accessibility (mean 0.13, range -0.05 to 0.32), and hydrophilicity (mean 0.13, range -0.05 to 0.30) were generally positive. While only three monoclonals were run on the 100K, all of them had moderate positive correlations with sequence complexity (mean 0.31, range 0.27 to 0.37), and moderately negative correlations with flexibility (mean -0.19, range -0.14 to -0.22), accessibility (mean -0.16, range -0.09 to -0.20), and hydrophilicity (mean -0.13, range -0.08 to -0.14). Remarkably, the only monoclonal that was run on all three platforms (P53Ab1) seems to prefer different properties depending on the microarray format that was used. Correlations between peptide properties and specificity were also examined. No property consistently correlated with peptide specificity on any platform.

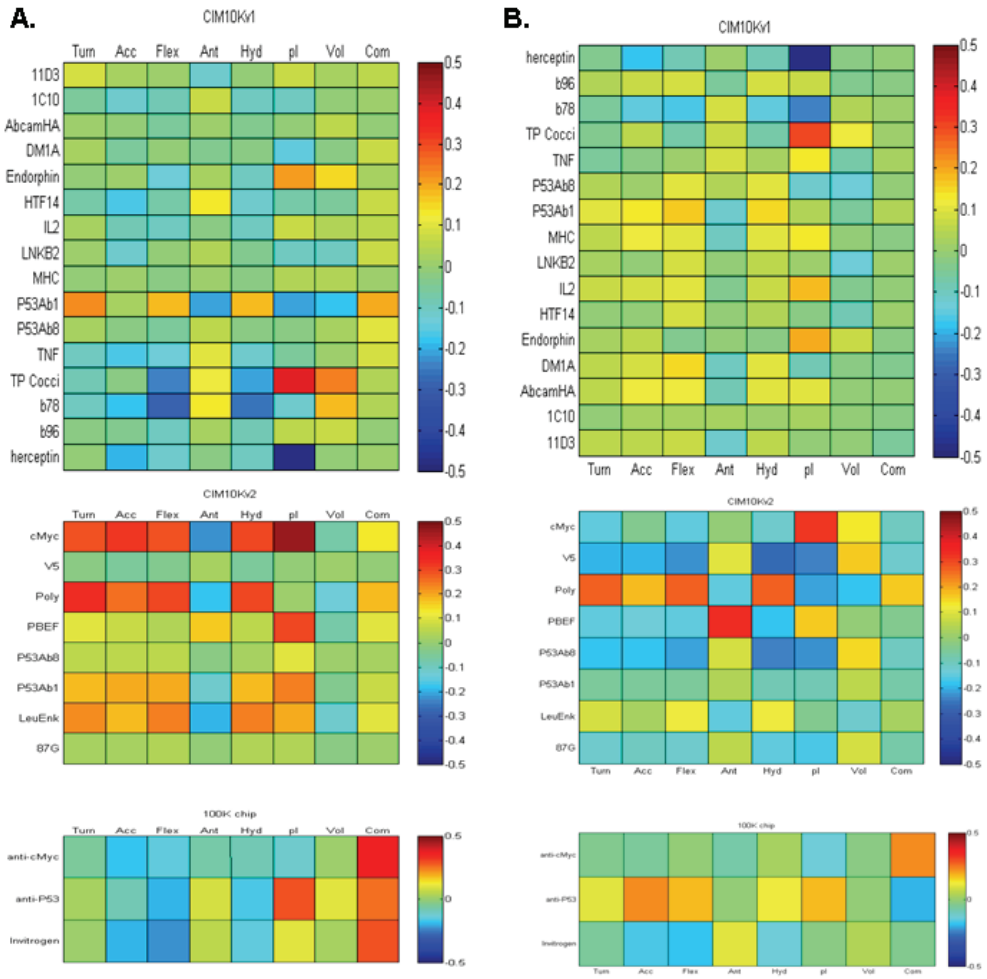


Figure 4.3 Correlations with properties for all mAb's and platforms. Pearson correlations between signal intensities (A) or specificity (B) of monoclonal antibody binding and sequence properties are shown as a heatmap.

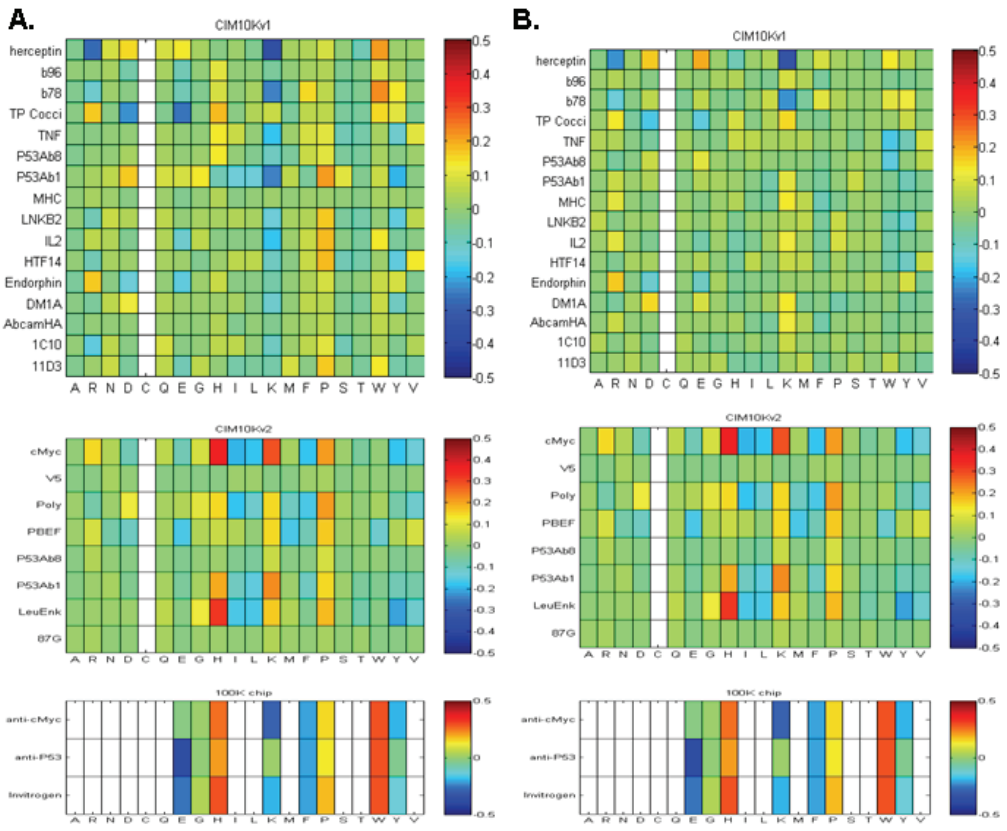


Figure 4.4 Correlations of amino acids counts for all mAb's and platforms. The pearson correlation between signal intensities (A) or specificity (B) of monoclonal antibodies and amino acid counts of peptides are shown as a heatmap.

The relationship between the average signal intensities of all of the monoclonals run on a given platform was plotted against each peptide property. For the CIMv1 and CIMv2 the relationships were fairly linear or flat (data not shown). More interesting relationships were observed between the properties and the average monoclonal signals on the 100K chip (Figure 4.5). Boxplots revealed an inverted U shape curve of the average signal intensities plotted against six of the properties: accessibility, flexibility, hydrophilicity, isoelectric point, and residue volume. In other

words, signal intensities were higher for peptides that had intermediate levels of these properties rather than high or low values. This preference for intermediate peptides correlates with the preference for high complexity peptides as peptides with lower complexity tend to have more extreme values of any property.

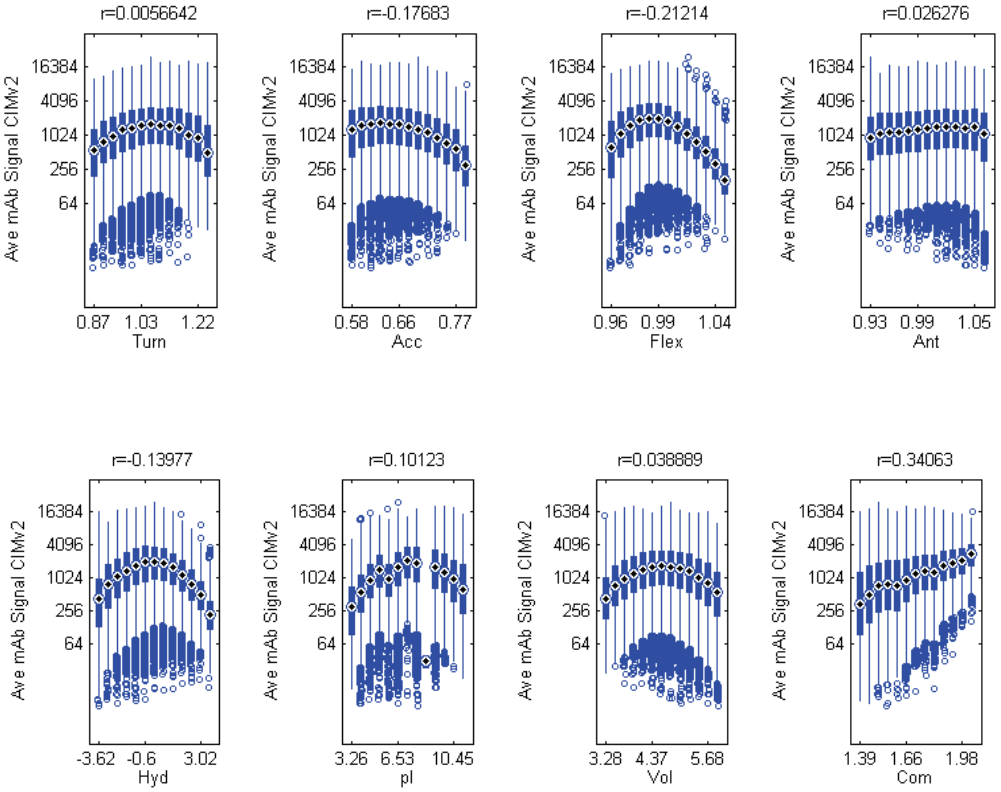


Figure 4.5 Boxplots of Properties vs. Signal Intensities on 100K Chip. The average of the three monoclonal antibodies binding signal is plotted against each sequence property as a boxplot. Pearson correlations between the average signal and the property are shown above each graph.

It seems likely that the antibodies bind to a subsequence of any peptide rather than the whole peptide. 5mers are likely the shortest subsequence that an antibody recognizes, but less than 5% of all possible 5mers with 19 aa occur on

either 10K array. Do to the larger library and more limited diversity, each unique 5mer of 8aa diversity occurs a median of 25 times on the 100K array (Figure 4.6). The geometric rank of the signal intensity for the three monoclonals was found for each 5mer. A set of 5mers stood out has having low geometric mean ranks for the P53Ab1 antibody. No 5mer were identified for the other two antibodies (data not shown). A scatterplot of the signal intensities of all peptides for two of the antibodies confirms that most of the peptides containing these 5mers bind more to the P53Ab1 than the invitrogen antibody. These 5mers have very similar sequences to the best 5mer identified (PGGYK), but are not at all related to the epitope sequence of P53Ab1 (RHSVV). Note that only one of the amino acids contained in the epitope was present on the 100K chip.

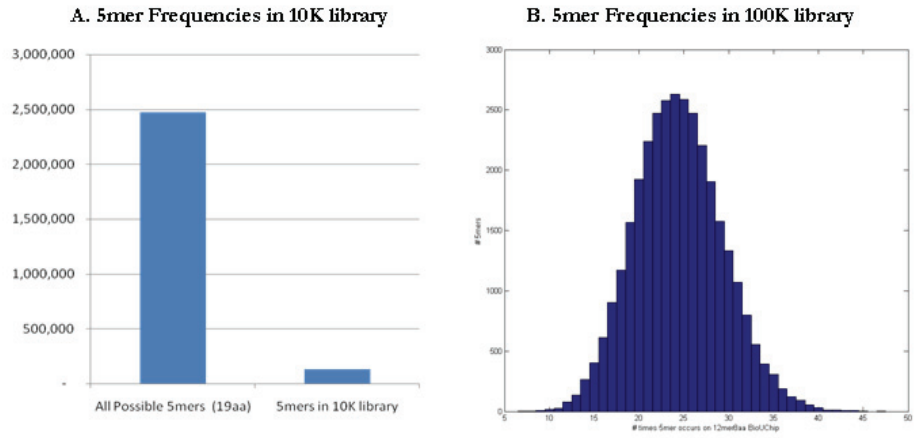


Figure 4.6 N-mer coverage comparisons on different platforms. The number of occurrences of times each unique 5mer occurs on the two different peptide library designs. The 10K libraries only have about 5% of all possible 5mers, and rarely have any repeated. The 100K library has every possible 5mer represented at least seven times and a median of 25 times.

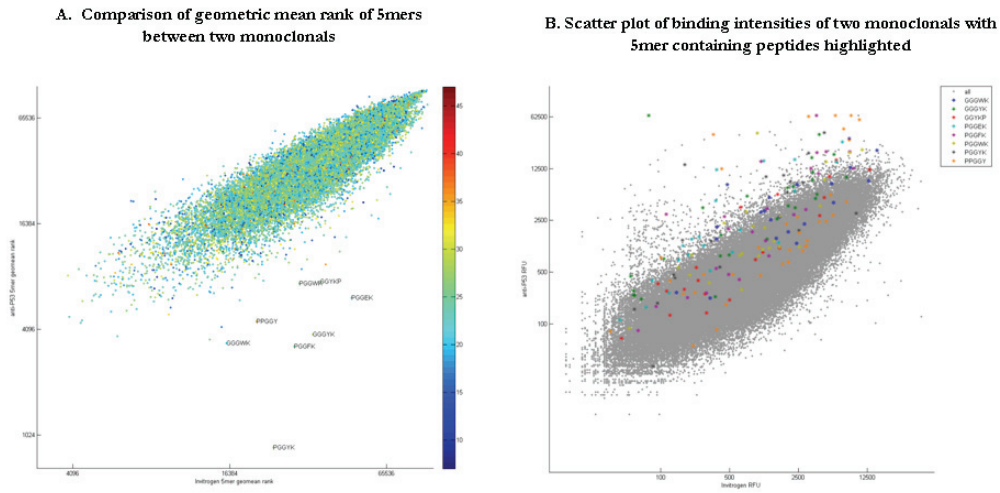


Figure 4.7 5mer analyses on 100K. All peptides were ranked by signal intensity for two monoclonal antibodies. The geometric mean rank of all the peptides containing each 5mer where found and are plotted as a scatter plot (A). The 5mers identified to have better ranks for the P53Ab1 are shown. A scatter plot of the signal intensities for each peptide for the same monoclonals (B) has the peptides containing those 5mers highlighted.

Discussion

Three different random peptide libraries assayed on three different peptide microarray platforms were described. The three peptide libraries had similar distributions of most sequence properties examined, except sequence complexity was drastically reduced on the 100K library compared to the 10K libraries. Monoclonal antibodies were found to bind a variable number of peptides on all three libraries. Some platform specific patterns of correlations with sequence properties were noted. Of particular interest was that on the 100K platform binding correlated most strongly with sequence complexity. The correlations between properties and antibody specificity tended to vary more between antibodies than with the platform. The more dense coverage of sequence space on the 100K platform enabled analysis

by 5mer subsequences. A set of 5mers specific for one of out of the three antibodies was identified.

The two 10K libraries and the 100K library have very different amino acid compositions, with the 10K's utilizing 19 amino acids and the 100K only utilizing eight. Despite the limited diversity, the 100K library had a similar coverage of most physiochemical properties. The eight amino acids used in the 100K included all of the aromatic or ring containing amino acids and thus had a distribution of average residue volume shifted up from the 10K libraries. The much more dramatic difference between the 10K and 100K libraries was the lower sequence complexity in the 100K.

The differences in the correlations between antibody binding on the different platforms could be due to a number of factors. The differences in the printing technologies used could have lead to differences in the peptide presentations that lead to differences in antibody binding. The peptide orientation appears to play an important role in how antibodies recognize different properties, as the CIM10Kv1 and 100K chip both are attached at the C-terminus and show similar correlations with peptide properties. However, often the question is not which peptides bind the best, but which are able to differentiate between two conditions. When examining antibody specificity, there does not appear to be any particular physiochemical property or set of properties that specific peptides have in common.

Sequence complexity does appear to be important in identifying specific peptides. On the 10K arrays, an average of 440 highly specific peptides was identified for each antibody and 23 out of the 24 antibodies tested had at least one

highly specific peptide (Appendix C, Table 1). Since replicate chips were not available for the monoclonals on the 100K, it was not possible to identify specific peptides in the same manner as on the 10K. However the repetition of the 5mers on the 100K enabled the confident identification of 5mers that were preferentially bound by an antibody. Surprisingly only one of the three antibodies was able to recognize specific 5mer sequences. We also saw a similar issue with differentiating infected and naïve sera on the 100K platform. While previous studies have shown that the 10K library can robustly distinguish mouse influenza immune sera from naïve mouse sera (Legutki et al. 2010 and unpublished data on the 10Kv2), we were not able to identify peptides that were significantly different between these sera on the 100K chip (Figure F-5).

Previous studies have found that the aromatic residues are particularly important in paratopes, and have designed Fab libraries that contain only two to four different amino acids in the CDR loops. They have been able to identify high affinity and specificity scFvs to a variety of ligands using these limited libraries (Koide and Sidhu 2009). Instead of having a limited diversity library of paratopes, here we have limited diversity library of potential epitopes. While we were able to identify many antibody binders to all three antibodies tested, we were only able to identify high specificity binders to one of the antibodies. We conjecture that the paratope library relies on structural diversity to confer specificity, but since peptides are likely unstructured, sequence diversity is required to identify specific binders.

Immunosignaturing, a technique utilizing random-sequence peptide microarrays to profile the humoral immune response, has shown great promise as

diagnostic technology (Legutki et al. 2010). While the peptides shown to be differentially bound by disease sera and control sera may be used a as biomarkers of the disease, they do not readily identify what antigen originally elicited those antibodies. Previous work on the 10K libraries has shown that the sequence similarity between peptides recognized by a monoclonal and its epitope would not have been sufficient to predict the antigen had it been unknown (Halperin, Stafford, and Johnston 2011). However, the 100K platform has a much denser sampling of sequence space, as illustrated by the repetition of 5mer sequences. While we were not able to identify an epitope sequence from this library because most of the amino acids in the epitopes of the antibodies tested were not on the chip, we were able to identify other 5mers that were preferentially bound by one of the antibodies. When peptides containing these 5mers were resynthesized and compared directly to a peptide containing the epitope sequence, much greater binding was observed for the epitope peptide (Figure F-11). Therefore, we would expect that if peptides containing the epitope sequence were present on the 100K array, they would likely stand out even more than the 5mers we identified. Although a 5mer sequence would not be sufficient to uniquely identify a protein out of a database of all known proteins, the 5mer along with other biological information such as organism, expected molecular weight, or sub-cellular localization could help narrow down to a reasonable number of candidate antigens to test experimentally. When new versions of the 100K chip become available that have more amino acid diversity they are likely to have much more power in predicting epitopes from immunosignaturing experiments.

Reference

Amanna, I. J. and M. K. Slifka. Contributions of humoral and cellular immunity to vaccine-induced protection in humans. 2011. Virology 411, no. 2: 206-15.

Anderson, Karen S. and Joshua LaBaer. 2005. The sentinel within: Exploiting the immune system for cancer biomarkers. Journal of Proteome Research 4, no. 4: 1123-1133.

Blythe, M. J. and D. R. Flower. 2005. Benchmarking b cell epitope prediction: Underperformance of existing methods. Protein Science 14, no. 1: 246-8.

Breitling, F., T. Felgenhauer, A. Nesterov, V. Lindenstruth, V. Stadler, and F. R. Bischoff. 2009. Particle-based synthesis of peptide arrays. Chembiochem 10, no. 5: 803-8.

Butler, J. E. 2000. Solid supports in enzyme-linked immunosorbent assay and other solid-phase immunoassays. Methods (Duluth) 22, no. 1: 4-23.

Carter, J. M. and L. Loomis-Price. 2004. B cell epitope mapping using synthetic peptides. Current Protocols in Immunology Chapter 9: Unit 9 4.

Chen, S. W., M. H. Van Regenmortel, and J. L. Pellequer. 2009. Structure-activity relationships in peptide-antibody complexes: Implications for epitope prediction and development of synthetic peptide vaccines. Curr Med Chem 16, no. 8: 953-64.

Chou, P. Y. and G. D. Fasman. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. Advances in Enzymology & Related Areas of Molecular Biology 47: 45-148.

Colasanti, T., C. Barbati, G. Rosano, W. Malorni, and E. Ortona. 2010. Autoantibodies in patients with alzheimer's disease: Pathogenetic role and potential use as biomarkers of disease progression. Autoimmunity Reviews 9, no. 12: 807-11.

Denisova, G. F., D. A. Denisov, and J. L. Bramson. 2010. Applying bioinformatics for antibody epitope prediction using affinity-selected mimotopes - relevance for vaccine design. Immunome Res 6 Suppl 2: S6.

Dudak, F. C., I. H. Boyaci, and B. P. Orner. 2011. The discovery of small-molecule mimicking peptides through phage display. Molecules 16, no. 1: 774-89.

Eggert, M., U. K. Zettl, and G. Neeck. 2010. Autoantibodies in autoimmune diseases. Current Pharmaceutical Design 16, no. 14: 1634-43.

El-Manzalawy, Y. and V. Honavar. 2010. Recent advances in b-cell epitope prediction methods. Immunome Res 6 Suppl 2: S2.

Emini, E. A., J. V. Hughes, D. S. Perlow, and J. Boger. 1985. Induction of hepatitis a virus-neutralizing antibody by a virus-specific synthetic peptide. Journal of Virology 55, no. 3: 836-9.

Fauchere, J. L., M. Charton, L. B. Kier, A. Verloop, and V. Pliska. 1988. Amino acid side chain parameters for correlation studies in biology and pharmacology. International Journal of Peptide & Protein Research 32, no. 4: 269-78.

Fodor, S. P., J. L. Read, M. C. Pirrung, L. Stryer, A. T. Lu, and D. Solas. 1991. Light-directed, spatially addressable parallel chemical synthesis. Science 251, no. 4995: 767-73.

Halperin, R. F., P. Stafford, and S. A. Johnston. 2011. Exploring antibody recognition of sequence space through random-sequence peptide microarrays. Molecular & Cellular Proteomics 10, no. 3: M110 000786.

Hansson, G. K. and A. Hermansson. 2011. The immune system in atherosclerosis. Nature Immunology 12, no. 3: 204-12.

Kajihara, Y., N. Yamamoto, R. Okamoto, K. Hirano, and T. Murase. Chemical synthesis of homogeneous glycopeptides and glycoproteins. 2010. Chemical Record: an Official Publication of the Chemical Society of Japan 10, no. 2: 80-100.

Kalaycioglu, A. T., P. H. Russell, and C. R. Howard. Peptide mimics of hapten dnp: The effect of affinity of anti-dnp monoclonal antibodies for the selection of phage-displayed mimotopes. Journal of Immunological Methods 366, no. 1-2: 36-42.

Karplus, P. A. and G. E. Schulz. 1985. Prediction of chain flexibility in proteins. Naturwissenschaften 72, no. 4: 212-213.

Kawashima, S., P. Pokarowski, M. Pokarowska, A. Kolinski, T. Katayama, and M. Kanehisa. 2008. Aaindex: Amino acid index database, progress report 2008. Nucleic Acids Research 36, no. Database issue: D202-5.

Kieber-Emmons, T. 1998. Peptide mimotopes of carbohydrate antigens. Immunologic Research 17, no. 1-2: 95-108.

Koide, S. and S. S. Sidhu. 2009. The importance of being tyrosine: Lessons in molecular recognition from minimalist synthetic binding proteins. ACS Chemical Biology [Electronic Resource] 4, no. 5: 325-34.

Kolaskar, A. S. and P. C. Tongaonkar. 1990. A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Letters 276, no. 1-2: 172-4.

Legutki, J. B., D. M. Magee, P. Stafford, and S. A. Johnston. 2010. A general method for characterization of humoral immunity induced by a vaccine or infection. Vaccine 28, no. 28: 4529-37.

Liang, S., D. Zheng, D. M. Standley, B. Yao, M. Zacharias, and C. Zhang. 2010. Epsvr and epmeta: Prediction of antigenic epitopes using support vector regression and multiple server results. BMC Bioinformatics 11: 381.

Morales Betanzos, C., M. J. Gonzalez-Moa, K. W. Boltz, B. D. Vander Werf, S. A. Johnston, and S. A. Svarovsky. 2009. Bacterial glycoprofiling by using random sequence peptide microarrays. *Chembiochem* 10, no. 5: 877-88.

Meloen, R. H., W. C. Puijk, and J. W. Slootstra. 2000. Mimotopes: Realization of an unlikely concept. J Mol Recognit 13, no. 6: 352-9.

Parker, J. M., D. Guo, and R. S. Hodges. 1986. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: Correlation of predicted surface residues with antigenicity and x-ray-derived accessible sites. Biochemistry 25, no. 19: 5425-32.

Pellequer, J. L., E. Westhof, and M. H. Van Regenmortel. 1993. Correlation between the location of antigenic sites and the prediction of turns in proteins. Immunology Letters 36, no. 1: 83-99.

Reineke, U., C. Ivascu, M. Schlief, C. Landgraf, S. Gericke, G. Zahn, H. Herzel, R. Volkmer-Engert, and J. Schneider-Mergener. 2002. Identification of distinct antibody epitopes and mimotopes from a peptide array of 5520 randomly generated sequences. J Immunol Methods 267, no. 1: 37-51.

Tapia, V., J. Bongartz, M. Schutkowski, N. Bruni, A. Weiser, B. Ay, R. Volkmer, and M. Or-Guil. 2007. Affinity profiling using the peptide microarray technology: A case study. Analytical Biochemistry 363, no. 1: 108-18.

Uttamchandani, M. and S. Q. Yao. 2008. Peptide microarrays: Next generation biochips for detection, diagnostics and high-throughput screening. Curr Pharm Des 14, no. 24: 2428-38.

Vihinen, M., E. Torkkila, and P. Riikonen. 1994. Accuracy of protein flexibility predictions. Proteins 19, no. 2: 141-9.

Vlieghe, P., V. Lisowski, J. Martinez, and M. Khrestchatisky. 2010. Synthetic therapeutic peptides: Science and market. Drug Discovery Today 15, no. 1-2: 40-56.

Volkmer, R. 2009. Synthesis and application of peptide arrays: Quo vadis spot technology. Chembiochem 10, no. 9: 1431-42.

Walsh, G. 2010. Post-translational modifications of protein biopharmaceuticals. Drug Discovery Today 15, no. 17-18: 773-80.

Xu, Y., P. A. Ramsland, J. M. Davies, M. Scealy, K. S. Nandakumar, R. Holmdahl, and M. J. Rowley. 2004. Two monoclonal antibodies to precisely the same epitope of type ii collagen select non-crossreactive phage clones by phage display: Implications for autoimmunity and molecular mimicry. Molecular Immunology 41, no. 4: 411-9.

Zhang, Q., P. Wang, Y. Kim, P. Haste-Andersen, J. Beaver, P. E. Bourne, H. H. Bui, S. Buus, S. Frankild, J. Greenbaum, O. Lund, C. Lundegaard, M. Nielsen, J. Ponomarenko, A. Sette, Z. Zhu, and B. Peters. 2008. Immune epitope database analysis resource (iedb-ar). Nucleic Acids Research 36, no. Web Server issue: W513-8.

Zhang, Wen, Yi Xiong, Meng Zhao, Hua Zou, Xinghuo Ye, and Juan Liu. Prediction of conformational b-cell epitopes from 3d structures by random forest with a distance-based feature. BMC Bioinformatics 12, no. 1: 341.

CHAPTER 5

EXPLORING ANTIBODY RECOGNITION OF SEQUENCE SPACE THROUGH RANDOM-SEQUENCE PEPTIDE MICROARRAYS

Summary

A universal platform for efficiently mapping antibody epitopes would be of great use for many applications, ranging from antibody therapeutic development to vaccine design. Here we tested the feasibility of using a random peptide microarray to map antibody epitopes. Although peptide microarrays are physically constrained to ~10^4 peptides per array, compared to 10^8 permitted in library panning approaches such as phage display, they enable a much more high though-put and direct measure of binding. Long (20mer) random sequence peptides were chosen for this study to look at an unbiased sampling of sequence space. This sampling of sequence space is sparse, as an exact epitope sequence is unlikely to appear. Commercial monoclonal antibodies with known linear epitopes or polyclonal antibodies raised against engineered 20mer peptides were used to evaluate this array as an epitope mapping platform. Remarkably, peptides with the most sequence similarity to known epitopes were only slightly more likely to be recognized by the antibody than other random peptides. We explored the ability of two methods singly and in combination to predict the actual epitope from the random sequence peptides bound. Though the epitopes were not directly evident, subtle motifs were found among the top binding peptides for each antibody. These motifs did have some predictive ability in

searching for the known epitopes among a set of decoy sequences. The second approach using a windowing alignment strategy, was able to score known epitopes of monoclonal antibodies well within the test dataset, but did not perform as well on polyclonals. Random peptide microarrays of even limited diversity may serve as a useful tool to prioritize candidates for epitope mapping or antigen identification.

Introduction.

Antibodies play an important role in protecting against infectious disease and contribute to pathology in autoimmune disease. Understanding antibody-antigen interactions is important for elucidating disease etiology, as well as facilitating vaccine design and diagnostic test development. In addition to their role in the immune system, antibodies are also very useful as affinity reagents for detection and purification as well as clinical diagnostic tools and pharmaceuticals. Epitope mapping is often an important step in determining if an antibody is suitable for a particular application, sorting among antibodies or determining how it performs its function. Many methods exist for identifying the epitope of an antibody, including crystallography, peptide tiling, and phage display (Fack et al. 1997; Reineke 2009). In this study, we will examine whether a faster, less expensive array based approach could be applied to the epitope mapping problem

A crystal structure of the antibody bound to the target is generally considered the gold standard of epitope mapping because it gives the most detailed information about the binding mechanism and will work for both conformational and linear epitopes. To identify a linear epitope, the peptide tiling method is often preferred

because it is simple and straightforward. However, the expense of synthesizing tiling peptides for every protein target may be prohibitive. To avoid the costly synthesis step, a library approach such as phage display may be employed. Peptides with random sequences can be displayed on the surface of phage, and those that bind best to the antibody can be selected and amplified. In the case of a linear epitope, the sequences recovered generally have sequences that very closely or exactly match the epitope sequence (Bongartz, Bruni, and Or-Guil 2009; Cortese et al. 1994; Gershoni et al. 2007; Irving, Pan, and Scott 2001; Wang and Yu 2004; Yip and Ward 1999). However, several rounds of selection, as well as sequencing of many selected clones makes this process expensive and time consuming. Furthermore, phage display has an inherent bias in selecting peptides that facilitate growth which reduces the effective size of the library. A faster method that allows a more direct measure of binding would be ideal.

Peptide arrays provide an alternative for screening a library of peptides for binding activity. The challenge of the array based approach is that the size of the peptide library feasible is several orders of magnitude smaller than those typically used in phage display. We have developed a random-sequence peptide microarray and are exploring its usefulness in a number of applications. The peptide library consists of 10,340 random sequence peptides that have seventeen randomized positions and a three amino acid linker. The library represents a very sparse sampling of sequence space, as only five percent of all possible 5mer sequences are represented. Despite the small library size, the random-sequence peptide microarray was successfully used to identify protein and glycan binding peptides, and most

pertinent to this study, to profile humoral immune responses (Boltz et al. 2009; Legutki et al. 2010; Morales Betanzos et al. 2009; Williams et al. 2009). This technique known as immunosignaturing is a novel method to detect changes in antibody reactivity has been described by our group elsewhere (Legutki et al. 2010). The peptides need only be mimotopes for immunosignaturing to serve its main purpose as a diagnostic platform. However, it is obvious to ask how the peptide sequences may relate to the antigen that raised the detected antibody response. It would be very useful if the peptide sequences identified in the immunosignaturing experiments could be used to identify the immunogenic epitopes in a pathogen or an autoantigen. Some preliminary use of random peptide arrays for epitope mapping was done by Reinke et al., but required subsequent rounds of mutational analysis of the peptides to hone in on the epitope sequence (Reineke et al. 2002). However, it may be possible to use a more sophisticated bioinformatics approach to infer the epitope from these loosely related sequences.

In order to evaluate how well the random-sequence peptide microarray could work for predicting the antibody epitopes, ten antibodies with known epitopes were selected. Five well characterized monoclonal antibodies that recognize linear epitopes were selected. Two recognize distinct P53 epitopes, which had been identified through peptide tiling and phage display experiments (Stephen, Helminen, and Lane 1995). The antibody against tubulin was epitope mapped using protease digestions, and the anti-IL2 has been co-crystallized with its target (Afonin et al. 2001; Breitling and Little 1986). The anti-HA monoclonal was raised against the peptide commonly used as an epitope tag. Five polyclonal antibodies raised against

random peptides from the library were included to evaluate the more difficult task of predicting an epitope from a polyclonal immune response. Using anti-peptide sera allows us to have a polyclonal response, but still know where the epitopes must be. We chose peptides from the library to facilitate comparison of the relative binding levels of the cognate peptide and the array of mimotopes. This known epitope antibody set will allow us to evaluate our epitope prediction strategy (Fig. 1).

Motif finding algorithms are able to find subtle patterns in sets of unaligned sequences. These algorithms may be classified in two main categories: deterministic and optimizing. Deterministic algorithms will exhaustively search a sequence set for motifs fitting a well defined set of criteria. Some popular implementations of deterministic motif finding algorithms are TEIRESIAS or PRATT (Jonassen 1997; Rigoutsos and Floratos 1998). The optimizing algorithms represent the motif probabilistically and try to maximize a scoring function. The optimization can be preformed stochastically such as using Gibbs motif sampling or by expectation maximization as implemented in MEME (Bailey et al. 2006). An optimization approach seems most appropriate for this problem because we do not know what criteria the motif should fulfill. Since it is possible that gapped motifs may be useful here, the GLAM2 implementation of the Gibbs motif sampling algorithm will be used here because it allows for gaps (Frith et al. 2008).

An alternative to finding a motif among the peptides would be to compare the peptides one at a time to the antigen sequence(s). A similar sequence analysis problem was addressed by a group using peptides selected by phage display to bind to small molecules to identify analogous binding sites in real proteins. The algorithm

implemented in the RELIC MATCH program compares each peptide sequence to the target protein sequence in five amino acid windows, and scores each window for similarity (Mandava et al. 2004). The scores for all of the peptides are added up across the protein sequence to predict potential small molecule binding site. A similar approach may be useful for predicting antibody recognition sites from dissimilar peptide sequences selected in a peptide microarray experiment.

Typically, epitope mapping is performed in order to identify the specific part of a protein target that is recognized by the antibody. A recent study has demonstrated the feasibility of using a similar approach to identify an unknown protein target of an antibody (Bastas et al. 2008). Peptides from a phage display library can be selected against an antibody. Motifs in the peptide sequences can be used to search a database of potential antibody targets. The authors concluded that a motif of at least seven amino acids or two shorter motifs in combination could be used to reasonably identify a protein target among a database of candidates. This approach could be powerful in identifying the antigenic proteins in a pathogen, targets in an autoimmune disease, or even discovering the cause of an unknown infection and the potential to translate this approach to a microarray platform will be explored here. Here we determine to what extent a bioinformatics approach may be used to predict the epitope directly from the 10K peptide array data.

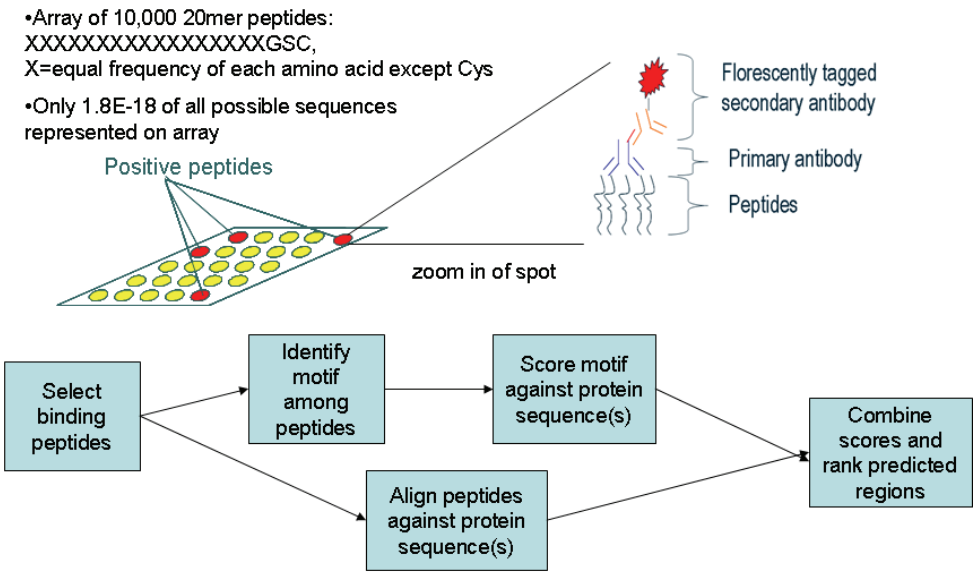


Figure 5.1 Experimental design schematic.

Methods. Random sequence peptide arrays were produced as described in Morales Betanzos et al. (2009). Briefly, 10K random 17mer sequences containing equal probability of the 20 amino acids except cysteine were generated. A C-terminal linker of Gly-Ser-Cys was added to each sequence to make 20mer peptides. These peptides were synthesized by Alta Bioscience (Birmingham, UK), diluted in 25% DMF 75% PBS, and spotted on a Telechem Nanoprint quill type contact printer onto sulfo-SMCC activated aminosilane coated glass slides.

Table 5.1 Antibodies used in this study. mAb1 is a monoclonal antibody raised against a peptide. mAb2 through mAb5 are monoclonal antibodies with epitopes previously determined in the literature. All of the polyclonal antibodies were raised against peptides selected from the random sequence peptide array.

Name	Immunogen	Clonality	Isotype	Clone Name	Sequence	pI	Hydro-pathicity
mAb1	HA peptide	Monoclonal	IgG1	16B12	YPYDVPDYA	3.56	-0.9
mAb2	Human IL2	Monoclonal	IGg1	LNKB2	KPLEEVLNL	4.53	-0.044
mAb3	Human p53	Monoclonal	IgG1	PAb240	RHSVV	9.76	-0.02
mAb4	Human p53	Monoclonal	IgG2b, IgG2a	DO-7, BP53-12	SDLWKL	5.55	-0.25
mAb5	Human Tubulin-alpha	Monoclonal	IgG1kappa	DM1A	AALEKD	4.67	-0.583
pAb1	KLH-peptide	Polyclonal	IgG detected	NA	MDQDDGEGV-IGHFHPILGSC	4.21	-0.185
pAb2	KLH-peptide	Polyclonal	IgG detected	NA	EFWDKEWHTR-ADWPVWDGSC	4.43	-1.245
pAb3	KLH-peptide	Polyclonal	IgG detected	NA	TIPAHNIFWI-LYFSIGTGSC	6.4	0.87
pAb4	KLH-peptide	Polyclonal	IgG detected	NA	PAMKHREPH-WVIPGIIWGSC	8.64	-0.13
pAb5	KLH-peptide	Polyclonal	IgG detected	NA	EFSNPTAQVF-PDFWMSDGSC	3.49	-0.315

Antibodies with known epitopes were purchased from commercial suppliers. mAb1, mAb2, mAb3 and mAb4 were purchased from Labvision (Fremont, CA) and mAb5 was purchased from Abcam (Cambridge, MA). The five polyclonal antibodies were produced as follows. Mice were immunized with keyhole limpet hemocyanin (KLH) conjugated peptides and sera was obtained at day 35. All animal work was conducted following an animal use protocol which was approved by the Arizona State University Institutional Animal Care and Use Committee. Antigen specific antibodies were absorbed from sera by binding to KLH immobilized on nitrocellulose membrane. A one by six centimeter nitrocellulose membrane was placed in a 15 ml conical tube with 1.0 mg/ml KLH in 2.0 ml PBS. The membrane was washed three times in TBST and incubated with 1.0% BSA in TBST for at least

one hour or until used. After washing three times in TBST, the membrane was placed into 2.0 ml of sera diluted 1:500 in 3% BSA, 0.05% Tween, PBS buffer. Reactivity of sera for KLH was tested in an ELISA. Sera were considered absorbed when no reactivity for KLH was detected at the 1:500 dilution.

The monoclonal antibodies or polyclonal sera were used to probe the peptide microarrays. After slides were passivated with 0.014% Mercaptohexanol, antibody was diluted to 100nM or sera were diluted 1:500 in 3% BSA, 0.05% Tween, PBS. Antibodies were incubated with slides for 1 hour at 37C in Agilent Chambers with rotation. Slides were washed three times with TBS, 0.05% Tween and three times with diH_2O. The incubation and wash procedure was repeated with a biotinalyted secondary antibody (Bethyl Laboratories, Inc. Montgomery, TX), then with Alexa-555 labeled Streptavadin (Invitrogen, Carlsbad, CA). Negative control arrays with no primary antibody or naïve mouse sera were also run for comparison. At least three replicate arrays were run for each antibody.

Slides were aligned using GenePix Pro 6.0 (Molecular Devices) and median spot intensities were averaged across replicate slides. Negative control signals were subtracted from antibody signals to remove the contribution of the secondary binding. The top 500 peptides in fluorescent intensity were selected for each antibody. The number of times each peptide occurs in one of the top 500 peptides lists was tabulated. Peptides appearing in five or more lists were eliminated as they are likely Fc binders or other nonspecific interactions. For the polyclonal datasets, the immunizing peptides were also excluded. The peptide lists described here will be referred to as the *binders* lists in subsequent analysis.

Peptides from the array were compared to the epitope sequences to identify those with sequence similarity. The epitope was expressed as a GLAM2 motif and was used in GLAM2SCAN to search against the peptides from the array inserted in strings of cysteines, with an alphabet of equal amino acid frequencies. Peptides were sorted by the highest scoring match and lists of the best matching peptides were created and these lists are referred to as the *aligners* lists. These lists were compared with lists of peptides that most strongly bind to each peptide and the proportion of overlap was examined.

Test datasets were generated for the monoclonal antibodies by randomly selecting sequences from human Swissprot and then randomly selecting a window of that sequence the same length as the epitope sequence. Two hundred negative examples were generated for each monoclonal. One thousand random peptides were generated as the negative examples for the polyclonal antibodies with equal frequencies of the nineteen amino acids (cysteine was not included as in the arrays). All of these sequences were inserted within a string of seventeen cysteines on each side to allow peptides to be aligned overhanging the test sequences.

Motifs were generated from the *binders* peptide lists using GLAM2, with a starting width of five amino acids, 1,000,000 iterations without improvement, 10 runs, and an alphabet of equal proportions of the 20 amino acids. GLAM2SCAN was used to search the corresponding test sets for sequences matching the motif with the alphabet set as the default protein alphabet for the monoclonal antibodies or equal amino acid frequencies for the polyclonal antibodies. GLAM2SCAN output is

the score for each place the motif matches in the test sequence set. The test sequences were ranked by the highest score match within each sequence.

The RELIC Fastaskan program was used to align the binding peptides to the test dataset. The *binders* peptide lists were uploaded as the affinity selected peptides and the corresponding test dataset was uploaded as the FASTA file. Random peptides were not subtracted. Fastaskan compares each five amino acid window of the test sequence with the selected peptide sequence and summing scores of the alignments above a threshold. It outputs a score for each test sequence corresponding to the window of maximum similarity between the peptides and that sequence.

Table 5.2 Comparison of RELIC and GLAM2 approaches

RELIC	**GLAM2**
Compares peptide sequences and database sequences pair-wise	Looks for a motif within all of the peptide sequences, then uses motif to search database
Uses a five amino acid window	Starts with a five amino acid window, but can adjust window size
Scores include amino acid similarity	Finds patterns using identity only

For both the GLAM2SCAN and the RELIC analysis, the rank of the true epitopes was compared to the test sequences using ROC analysis. A Matlab script to calculate the true positive and false positive rate for each score cutoff was obtained from http://theoval.cmp.uea.ac.uk/~gcc/matlab/roc/ and modified to smooth tied scores. The area under the ROC curve was also calculated using a Matlab script from the same website. The area under the curve will be used to predict the

probability of finding an epitope in a database of a given size. We will assume positive and negative examples will be selected from a database of a fixed size without replacement weighted by the probability that a positive is chosen over a negative as estimated by the area under the curve.

Results. The peptide array consists of randomly generated 17aa sequences with a 3aa, C-terminal linker. The length was chosen for two reasons. Practically, commercial sources of peptide synthesis limit the length to 20aa for large syntheses. Second, we have found that peptides longer than 20aa tend to assume secondary structure and are less soluble (saj, unpublished data). We chose to print 10K peptides because this was the maximum number that could be printed in duplicate on one standard slide. Standard glass slides were used in order to facilitate the processing of the slides on standard equipment and to reduce the cost.

In order to evaluate the peptide microarray platform, examples of antibodies against a known set of variable types of epitopes were chosen. Five monoclonal antibodies with known linear epitopes, and five examples of anti-peptide polyclonal mouse sera raised against peptides selected from the array were used as the test set. Together these epitopes cover a wide range of lengths and physiochemical properties (Table 1). These antibodies will allow us generate a dataset to test how well different sequence analysis approaches are able to predict these antibody epitopes. Clear binding above background was observed as illustrated in the representative slide images shown (Fig. 2). The monoclonal antibodies were found to bind to a median of 64.1% (range 37.6% -74.9%) of the random peptides above the slide surface

background and secondary only controls. Polyclonal sera showed similar peptide reactivity with a median of 63.6% (range 54.0% - 68.6%). The rank of the peptide used for immunization ranked within the top 100 peptides by signal intensity in four out of the five examples (Fig. 3). Replicate slides had an average Pearson correlation of 0.785 for monoclonals and 0.764 for polyclonals. The heatmap (Fig. 4) shows that each antibody has a distinct binding pattern on the array. While there is some overlap between the peptides bound by each antibody, about 22% of the top 500 peptides recognized by each antibody are not recognized by the other nine antibodies tested (Fig. 5). The uniqueness of the peptides recognized by each antibody implies that the peptide sequences may contain information about antibody specificity.

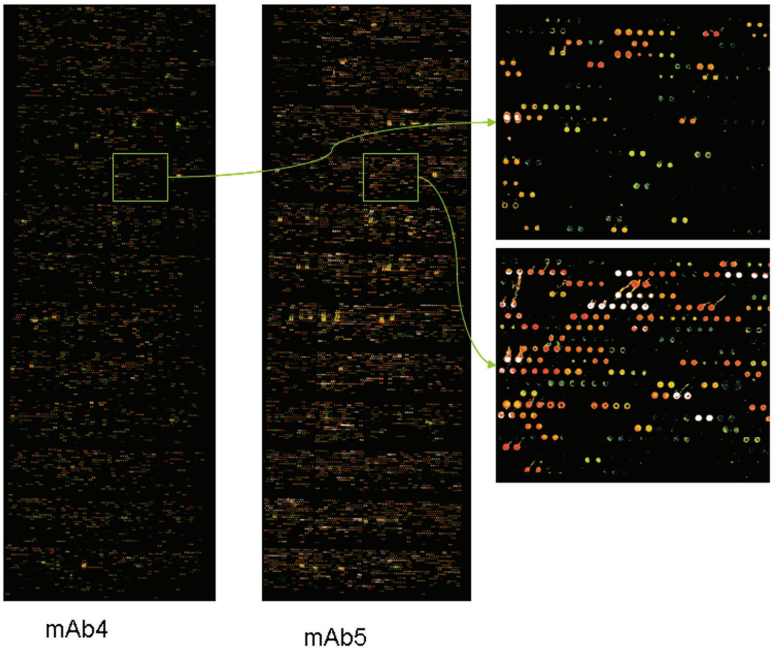


Figure 5.2 Array images. Two representative array images are shown in pseudocolor, with zoom in on the same block on each array to the right.

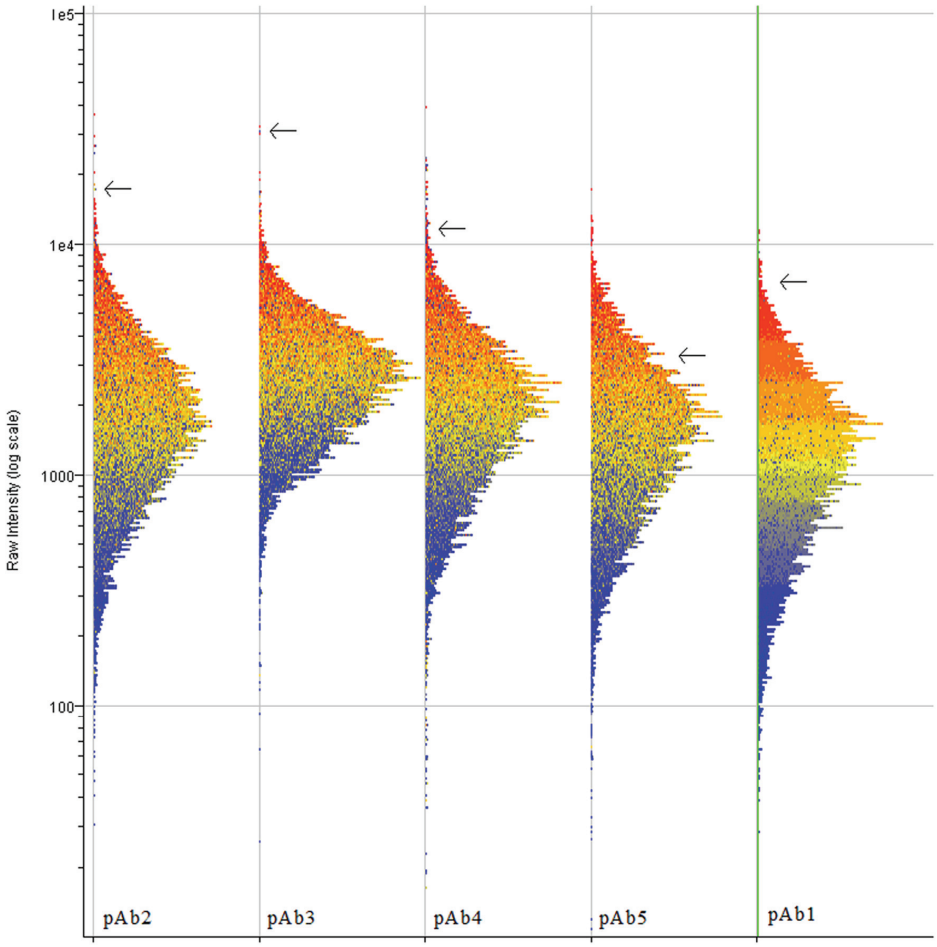


Figure 5.3 Histograms of peptide binding. The distributions of signal intensities for the polyclonal antibodies are shown as histograms with the signal intensities on the y-axis and the relative length of the bar proportional to the number of peptides with that signal intensity. The arrows indicate the location of the peptide that the polyclonal antibody was raised against and the number indicates the rank.

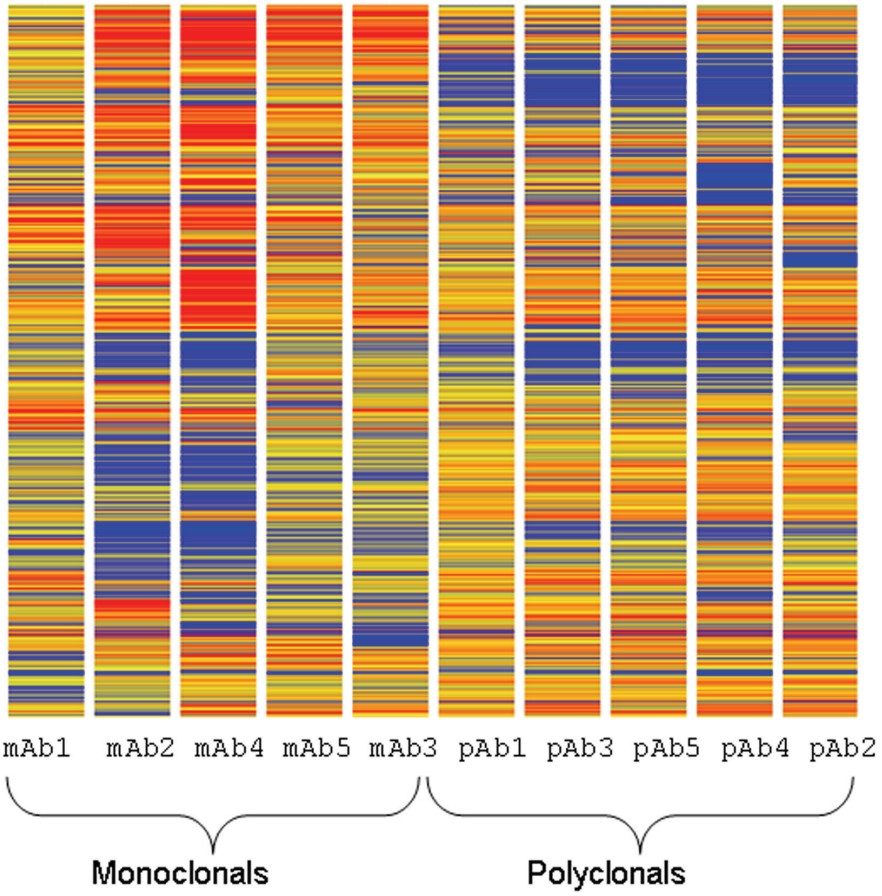


Figure 5.4 Heatmap of arrays. Data was normalized to the median on each array and clustered using GeneSpring. Averages of replicates are shown. Red represents high signals, yellow intermediate signals and blue low signals.

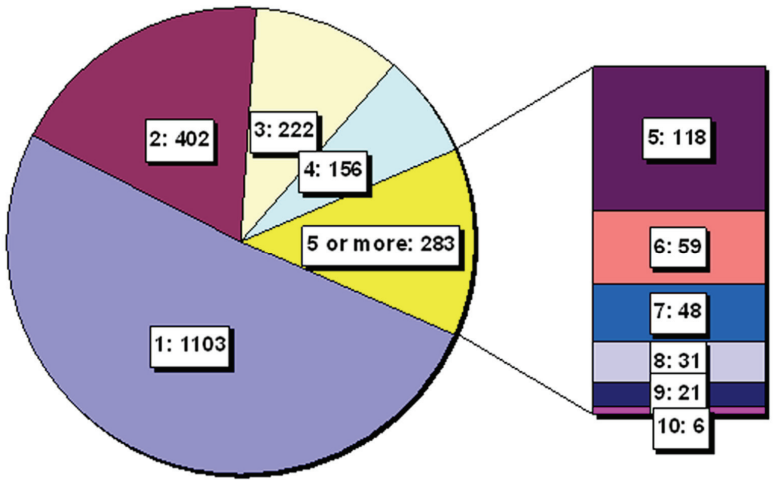


Figure 5.5 Pie chart of peptide binding showing the overlap between the top five hundred peptides for each of the ten antibodies. Of the 10,340 peptides, 2166 are in the top 500 peptides for at least one of the ten antibodies. 1103 peptides are unique to one of the antibodies and only 6 peptides are recognized by all ten antibodies.

Each peptide sequence was scored for similarity against each protein sequence. Most of the peptides bound by the antibodies did not show strong sequence similarity to the epitope (Table 3). However, there was some enrichment for sequence similar peptides among the binders. Most of the peptides bound are mimotopes rather than having any obvious similarity to the epitope.

Table 5.3 Sequence similarity vs. Binders. The *binders* column indicates the number of peptides selected from the microarray experiments and the *aligners* are the number of peptides that had sequence similarity with the known epitope as described in the Methods section. The number of peptides in common between the binders and aligners lists is in the *both* column. *Expected* column lists the number of peptides expected to be in common if lists were drawn at random from the peptide library. *Ratio* is the ratio of *both* to *expected*. The percentage of the *binders* list that is in *both* is in the last column. The p-value describes the probability that the *binder* and *aligner* list overlap as much as in the *both* list by chance based on the Fisher's Exact test.

	binders	***aligner***	***both***	***expected***	***ratio***	***% of binders align***	***P-value***
mAb1	350	181	**6**	5.88	1.02	1.70%	0.551
mAb2	391	379	**21**	13.75	1.53	5.40%	0.062
mAb3	379	188	**36**	6.61	5.45	9.50%	<0.001
mAb4	354	96	**2**	3.15	0.63	0.60%	0.334
mAb5	369	365	**6**	12.5	0.48	1.60%	0.016
pAb1	258	722	**26**	17.28	1.5	10.10%	0.053
pAb2	258	755	**26**	18.07	1.44	10.10%	0.068
pAb3	263	710	**18**	17.32	1.04	6.80%	0.493
pAb4	274	742	**22**	18.86	1.17	8.00%	0.424
pAb5	267	699	**21**	17.31	1.21	7.90%	0.301
average	316.3	483.7	**18.4**	13.07	1.55	6.20%	0.232

In order to assess the predictive power of these sequences the alignment of the peptides to the epitopes was compared to their alignment with a set of negative examples. The RELIC alignment program was able to align binding peptides to all of the monoclonal epitopes and 62.7% of the negative examples. The true epitopes had an average score of 14.3 while the negative examples had an average score of 5.9. The ROC analysis found an area under the curve of 0.87 indicating that a true epitope has an 87% chance of having a higher score than randomly selected negative example. All of the polyclonals also had positive peptide alignment scores as well as

86.5% of the positive examples. The true epitopes had an average score of 14.7 while the negative examples had a score of 15.2. The ROC analysis (Fig. 6) indicates that a positive example has a 46% chance of having a higher score than a negative example based on the area under the curve. The monoclonal epitopes were predicted well by this method, while the polyclonal predictions were similar to chance.

An algorithm capable of detecting subtle patterns may be able to garnish predictive power from these peptide sequences. Convergent motifs were identified for all of the antibodies using GLAM2. The motifs for the monoclonal antibodies ranged from three to five amino acids in width. The polyclonal motifs were four to five amino acids wide. The monoclonal motifs matched the epitope sequences with an average score of 3.5, while the negative examples had an average score of -3.7. Polyclonal motifs matched the immunizing peptide with an average score of 3.8 while the negative examples had an average score of 3.7. The ROC analysis demonstrates that the monoclonals epitopes have an 89.8% chance of being scored higher than the corresponding negative examples in the motif analysis while the polyclonals have a 67.9% chance of scoring higher than the negatives. The motif finding approach demonstrated predictive power on both datasets.

In order to test if combining the two approaches may improve the predictive ability, the scores from the RELIC analysis and the GLAM2 analysis were each scaled to have a minimum score of zero and a maximum score of one and averaged. The ROC analysis was performed on the averaged scores (Figure 5.6 and 5.7). The area under the curve was 0.92 for the monoclonals and 0.69 for the polyclonals.

Based on the probability estimated from the ROC analysis, there is about a 70% chance of finding a monoclonal epitope in the top ten windows out of a one hundred amino acid protein. There would be a 21% chance of correctly identifying a polyclonal epitope in a small virus among the top 100 hits out of a possible 1000 amino acid database, which is a two-fold enrichment.

While a decoy sequence set was used to estimate the accuracy of these prediction methods, the analysis was also run against the antibody target protein sequences to illustrate how the method would work for predicting an epitope of a known target. P53 was used as the example since both mAb3 and mAb4 are directed against different epitopes of P53, and represent the easiest and hardest epitopes to predict respectively of the five monoclonal examples. Figure 5.8a. and b. illustrate how well the motif finding and alignment approaches may work when binding peptides show some enrichment for sequence similarity to the epitope, as demonstrated by mAb3. When compared against the whole P53 sequence, both approaches predict the true epitope region with the highest score (Figure 5.8c). In contrast, the true epitope region was not predicted among the top regions by either method for mAb4 (Figure 5.8d.)

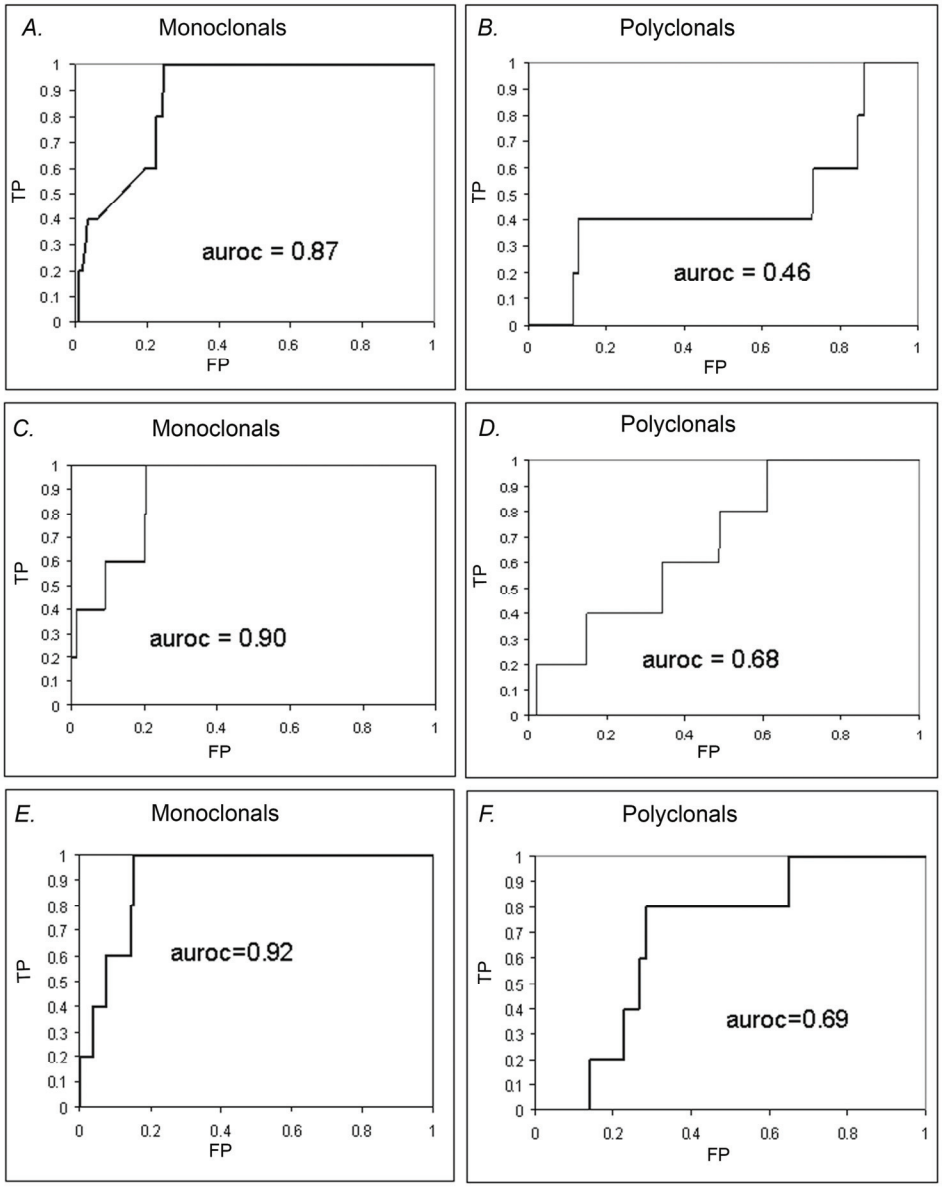


Figure 5.6 Epitope prediction accuracy. ROC curves of epitope predictions among decoy dataset. The true positive (TP) rate is plotted on the y-axis and the false positive (FP) rate is plotted on the x-axis. The area under the ROC curve (auroc) indicates the probability that a true epitope would rank higher than a decoy sequence. a. RELIC alignment predictions monoclonal epitopes. b. RELIC alignment predictions of polyclonal epitopes. c. GLAM2 motif predictions of monoclonal epitopes. d. GLAM2 motif predictions of polyclonal epitopes. e. Combined predictions of monoclonal epitopes. f. Combined predictions of polyclonal epitopes.

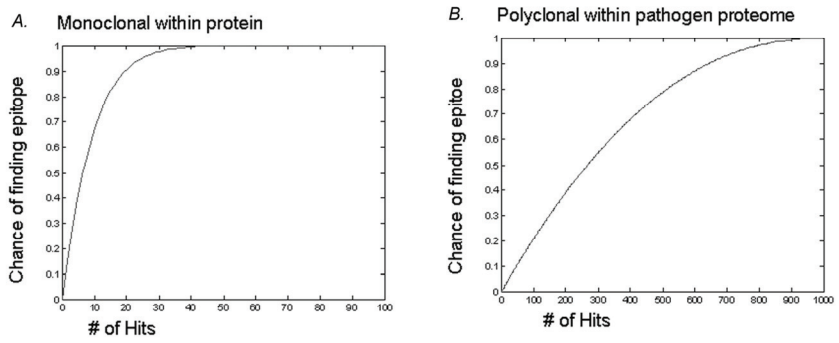


Figure 5.7. Chance of finding true epitope in a database. Using the probability calculating from the ROC analysis, the chance of finding an epitope in a database was calculated. The x-axis shows the number of hits that would be examined, and the y-axis shows chance the true epitope would be found among those hits. a. Chance of finding a monoclonal epitope within a dataset of 100 peptides. b. Chance of finding a polyclonal epitope with a set of 1000 peptides.

Discussion.

We have shown that a diverse set of antibodies will each bind a high percentage of the 10K random sequence peptides on our arrays. Each antibody bound a unique set of peptides with little overlap between sets. The list of peptides that bind best to each antibody is only slightly enriched for peptides with sequence similarity to a known epitope or immunogen. However, motif finding and alignment approaches were able to score monoclonal epitopes well among a set of decoy sequences. Predicting the immunogen in a polyclonal response was more difficult, with only the motif finding approach having predictive power.

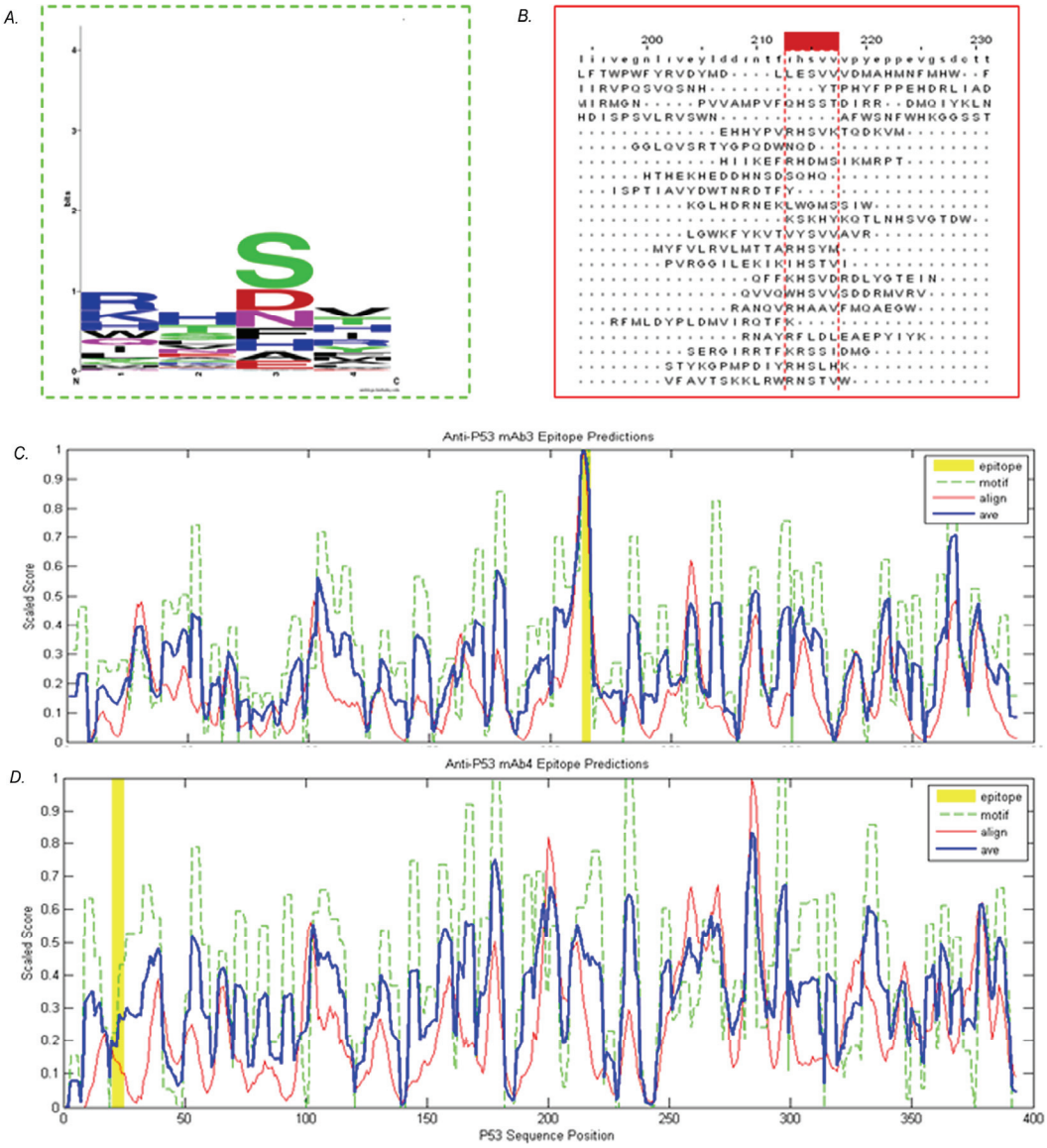


Figure 5.8 Example analyses of epitope predictions on P53. The anti-P53 monoclonals mAb3 and mAb4 had the best and worst enrichment respectively for sequence similar peptides among the binders. a. Motif identified by Glam2 analysis of mAb3 binders peptides is similar to mAb3 epitope sequence RHSVV. b. RELIC MATCH alignment of mAb3 binder peptides to epitope region. c. & d. Scores for glam2 motif (- -), RELIC alignment (—), and averaged scores (—) along the p53 sequence for mAb3 (c) and mAb4 (d). The respective epitope locations are highlighted in yellow.

Antibodies are conventionally characterized into two groups: polyspecific antibodies that recognize many different antigens at low affinity and monospecific antibodies that recognize one or few antigens at high affinity (Michaud et al. 2003; Zhou, Tzioufas, and Notkins 2007). Polyspecific antibodies typically have flexible paratopes which allows them to interact with antigens of a variety of shapes and the conformational entropy lost upon binding prohibits high affinity interactions. In contrast monospecific antibodies tend to be closer to a lock and key binding model, which implies that only antigens with very similar shapes should bind (James, Roversi, and Tawfik 2003; Mariuzza 2006). Here we found that a set of commonly used and well characterized monoclonal antibodies is capable of binding to many peptides of unrelated sequences when presented on the surface of a microarray. We also observed that polyclonal sera raised to one peptide was able to bind to a median of 20 peptides with higher intensities than the one the sera was raised against, and hundreds at intensities within a fold change of the cognate peptide (Fig. 3).

Weak motifs can still be found using a Gibbs motif sampler algorithm as implemented in GLAM2. The GLAM2 implementation was chosen because it allows for the insertion of variable lengths gaps in the motif. However, none of the motifs found had any insertions or deletions indicating that exact spacing is probably important in binding. Though the patterns identified were subtle, they were repeatedly found over multiple runs. These motifs were used to search a test dataset and were able to score true epitopes higher than decoy sequences. Another approach aligning one sequence at a time using RELIC also showed some predictive power, especially with the monoclonal examples. This indicates that a few sequences

with a highly similar window can be informative. The example analysis for the two P53 antibodies illustrate how this data may be used in practice to predict an epitope. The two extremes are depicted where the true epitope is clearly predicted correctly and when it is missed. It is interesting to note that when the motif and alignment methods clearly agree, the correct region is predicted suggesting that the concordance may be reason for confidence in the prediction.

We have identified several sources of error inherent to our microarray such as the presence of impurities and truncations from the peptide synthesis and are currently developing a new version of the microarray to mitigate these problems. Optimizing the parameters of the motif finding or alignment may be helpful, but would require a larger training dataset to avoid overfitting to these specific examples. The polyclonal is likely a more difficult problem because there may be multiple epitopes within the peptide sequence, but the algorithms are looking for one peak. It may be possible to modify the algorithm to pick up on multiple epitopes in one protein. The biggest limitation of this approach is most likely the sparse sampling of sequence space dictated by the peptide microarray technology employed here. However, much higher density peptide arrays may be produced by in situ synthesis methods. Currently there are oligonucleotide arrays commercially available that have 10^6 features, and it should be possible to achieve similar densities with peptide arrays (Gao et al. 2004). A peptide library of that size could be designed to cover all possible 5mer sequences, which would enable exact matches to short epitopes to be present.

The prediction accuracy observed in the monoclonal test dataset would be sufficient to narrow down possible linear protein epitopes to a few peptides and have a good chance of correct identification. A much more challenging but interesting question is predicting the immunogenic antigen(s) from polyclonal sera, as would be required based on an immunosignaturing experiment. It will be important for investigators to know the amount of uncertainty present in inferring epitopes from peptide sequences as more of these studies are performed. When identification of the antigen raising the immune response is desired, a number of approaches are possible depending on the biological information available. For example, if one had an immunosignature to a pathogen and wanted to identify the immunogenic epitopes within that pathogen, one could use the prediction methods described here in conjunction with other information available such as subcellular localization to prioritize proteins for further testing. In a more difficult scenario, one may suspect a pathogen may play a role in a chronic disease and would want to use the immunosignature to identify the pathogen. A more concrete idea of the sequence preference of the sera would be needed in order to search a database of all pathogens than could be inferred directly from the random peptide array results. Alternatively, if technology allowed ready production of new sets of peptides, one could build another array with peptides around the space of the original binders to hone in on the relevant sequence.

References

Afonin, P. V., A. V. Fokin, I. N. Tsygannik, I. Y. Mikhailova, L. V. Onoprienko, Mikhaleva, II, V. T. Ivanov, T. Y. Mareeva, V. A. Nesmeyanov, N. Li, W. A. Pangborn, W. L. Duax, and V. Z. Pletnev. 2001. Crystal structure of an anti-interleukin-2 monoclonal antibody fab complexed with an antigenic nonapeptide. *Protein Science* 10, no. 8: 1514-21.

Bailey, T. L., N. Williams, C. Misleh, and W. W. Li. 2006. Meme: Discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research* 34, no. Web Server issue: W369-73.

Bastas, G., S. R. Sompuram, B. Pierce, K. Vani, and S. A. Bogen. 2008. Bioinformatic requirements for protein database searching using predicted epitopes from disease-associated antibodies. *Mol Cell Proteomics* 7, no. 2: 247-56.

Boltz, K. W., M. J. Gonzalez-Moa, P. Stafford, S. A. Johnston, and S. A. Svarovsky. 2009. Peptide microarrays for carbohydrate recognition. *Analyst* 134, no. 4: 650-2.

Bongartz, J., N. Bruni, and M. Or-Guil. 2009. Epitope mapping using randomly generated peptide libraries. *Methods Mol Biol* 524: 237-46.

Breitling, F. and M. Little. 1986. Carboxy-terminal regions on the surface of tubulin and microtubules. Epitope locations of yol1/34, dm1a and dm1b. *Journal of Molecular Biology* 189, no. 2: 367-70.

Cortese, R., F. Felici, G. Galfre, A. Luzzago, P. Monaci, and A. Nicosia. 1994. Epitope discovery using peptide libraries displayed on phage. *Trends in Biotechnology* 12, no. 7: 262-7.

Fack, F., B. Hugle-Dorr, D. Song, I. Queitsch, G. Petersen, and E. K. Bautz. 1997. Epitope mapping by phage display: Random versus gene-fragment libraries. *J Immunol Methods* 206, no. 1-2: 43-52.

Frith, M. C., N. F. Saunders, B. Kobe, and T. L. Bailey. 2008. Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Computational Biology* 4, no. 4: e1000071.

Gao, X., J. P. Pellois, Y. Na, Y. Kim, E. Gulari, and X. Zhou. 2004. High density peptide microarrays. In situ synthesis and applications. *Molecular Diversity* 8, no. 3: 177-87.

Gershoni, J. M., A. Roitburd-Berman, D. D. Siman-Tov, N. Tarnovitski Freund, and Y. Weiss. 2007. Epitope mapping: The first step in developing epitope-based vaccines. *BioDrugs* 21, no. 3: 145-56.

Irving, M. B., O. Pan, and J. K. Scott. 2001. Random-peptide libraries and antigen-fragment libraries for epitope mapping and the development of vaccines and diagnostics. *Current Opinion in Chemical Biology* 5, no. 3: 314-24.

James, L. C., P. Roversi, and D. S. Tawfik. 2003. Antibody multispecificity mediated by conformational diversity. *Science* 299, no. 5611: 1362-7.

Jonassen, I. 1997. Efficient discovery of conserved patterns using a pattern graph. *Computer Applications in the Biosciences* 13, no. 5: 509-22.

Legutki, Joseph Barten, D. Mitchell Magee, Phillip Stafford, and Stephen Albert Johnston. 2010. A general method for characterization of humoral immunity induced by a vaccine or infection. *Vaccine* 28, no. 28: 4529-4537.

Mandava, S., L. Makowski, S. Devarapalli, J. Uzubell, and D. J. Rodi. 2004. Relic--a bioinformatics server for combinatorial peptide analysis and identification of protein-ligand interaction sites. *Proteomics* 4, no. 5: 1439-60.

Mariuzza, R. A. 2006. Multiple paths to multispecificity. *Immunity* 24, no. 4: 359-61.

Michaud, G. A., M. Salcius, F. Zhou, R. Bangham, J. Bonin, H. Guo, M. Snyder, P. F. Predki, and B. I. Schweitzer. 2003. Analyzing antibody specificity with whole proteome microarrays. *Nat Biotechnol* 21, no. 12: 1509-12.

Morales Betanzos, C., M. J. Gonzalez-Moa, K. W. Boltz, B. D. Vander Werf, S. A. Johnston, and S. A. Svarovsky. 2009. Bacterial glycoprofiling by using random sequence peptide microarrays. *Chembiochem* 10, no. 5: 877-88.

Reineke, U. 2009. Antibody epitope mapping using de novo generated synthetic peptide libraries. *Methods in Molecular Biology* 524: 203-11.

Reineke, U., C. Ivascu, M. Schlief, C. Landgraf, S. Gericke, G. Zahn, H. Herzel, R. Volkmer-Engert, and J. Schneider-Mergener. 2002. Identification of distinct antibody epitopes and mimotopes from a peptide array of 5520 randomly generated sequences. *J Immunol Methods* 267, no. 1: 37-51.

Rigoutsos, I. and A. Floratos. 1998. Combinatorial pattern discovery in biological sequences: The teiresias algorithm. *Bioinformatics* 14, no. 1: 55-67.

Stephen, C. W., P. Helminen, and D. P. Lane. 1995. Characterisation of epitopes on human p53 using phage-displayed peptide libraries: Insights into antibody-peptide interactions. *Journal of Molecular Biology* 248, no. 1: 58-78.

Wang, L. F. and M. Yu. 2004. Epitope identification and discovery using phage display libraries: Applications in vaccine development and diagnostics. *Current Drug Targets* 5, no. 1: 1-15.

Williams, B. A., C. W. Diehnelt, P. Belcher, M. Greving, N. W. Woodbury, S. A. Johnston, and J. C. Chaput. 2009. Creating protein affinity reagents by combining peptide ligands on synthetic DNA scaffolds. *Journal of the American Chemical Society* 131, no. 47: 17233-41.

Yip, Y. L. and R. L. Ward. 1999. Epitope discovery using monoclonal antibodies and phage peptide libraries. *Combinatorial Chemistry & High Throughput Screening* 2, no. 3: 125-38.

Zhou, Z. H., A. G. Tzioufas, and A. L. Notkins. 2007. Properties and function of polyreactive antibodies and polyreactive antigen-binding b cells. *Journal of Autoimmunity* 29, no. 4: 219-28.

CHAPTER 6

GUITOPE: AN APPLICATION FOR MAPPING RANDOM-SEQUENCE PEPTIDES TO PROTEIN SEQUENCES

Abstract

Background

Random-sequence peptide libraries are a commonly used tool to identify novel ligands for binding antibodies, other proteins, and small molecules. It is often of interest to compare the selected peptide sequences to the natural protein binding partners to infer the exact binding site or the importance of particular residues. The ability to search a set of sequences for similarity to a set of peptides may sometimes enable the prediction of an antibody epitope or a novel binding partner. We have developed a software application designed specifically for this task.

Results

GuiTope provides a graphical user interface for aligning peptide sequences to protein sequences. All alignment parameters are accessible to the user including the ability to specify the underlying amino acid frequency in the peptide library, which often differs significantly from the frequencies assumed by popular alignment programs. It also includes a novel feature to align dipeptide inversions, which we have found improves the accuracy of antibody epitope prediction from peptide microarray data and shows utility in analyzing phage display datasets. Finally,

GuiTope can randomly select peptides from a given library to estimate a null distribution of scores to calculate statistical significance.

Conclusions

GuiTope provides a convenient method for comparing selected peptide sequences to protein sequences, including flexible alignment parameters, novel alignment features, ability to search a database, and statistical significance of results. The latest version of the software available as an executable (for PC) at www.immunosignature.com/software and ongoing updates and source code will be available at sourceforge.net.

Background

Random-sequence peptide library screening approaches are an increasingly popular and powerful tool for identifying ligands for antibodies and other proteins as well as carbohydrates, pharmaceuticals, and other small molecules. Peptide library methods generally fall into two categories: molecular display approaches such as phage display, and immobilized arrays such as SPOT. Display approaches can typically accommodate much larger libraries, but information is typically obtained only on the clones that survive several rounds of panning which is heavily biased by sequences that facilitate growth (Derda et al. 2011). In contrast, array based approaches may be used to screen smaller libraries with higher throughput than display approaches and semi-quantitative binding information obtained on all of the peptides in the library. New technologies both on the display side and the array

approach promise to overcome these limitations (Breitling et al. 2009; Takakusagi et al. 2008; Ullman, Frigotto, and Cooley 2011). The decreasing cost of both sequencing and peptide synthesis also promise to accelerate research in this area. Furthermore, new applications for random-sequence peptide libraries such as profiling the humoral immune response (Legutki et al. 2010) promise to increase interest in connecting random-sequence peptides to protein sequences. Therefore, an increase in the demand for appropriate algorithms and software to facilitate the data analysis would also be expected.

While the peptides discovered in these library screening experiments serve as useful ligands in and of themselves, comparison of the sequences to natural protein sequences can reveal novel biological insight. Panning phage display libraries against monoclonal antibodies often selects for peptides that closely match the antibody epitope making the sequence analysis rather straightforward (Stephen and Lane 1992). If a strong enough motif is uncovered among the peptide sequences, it may even be used to search a database to predict an antibody target (Bastas et al. 2008). Though current array technology does not allow sufficient coverage of sequence space to contain sequence closely resembling natural protein sequences by chance, we have shown that experiments of this type still have utility for predicting monoclonal epitopes (Halperin, Stafford, and Johnston 2011). Other groups have shown that peptides selected to bind to other types of proteins have utility in understanding and predicting binding to natural binding partners (Cao and Mao 2009; Carter et al. 2006; Nie et al. 2008). Even small molecule binding peptides have

enabled biological insight on natural protein small molecule binders (Rodi et al. 1999; Takakusagi et al. 2008).

Analysis of the peptide sequences obtained from any selection experiment has unique challenges, primarily because the structural requirements for the selected peptides may be met by many diverse sequences. One approach is to identify a motif among the peptide sequences and use the consensus sequence or a probabilistic representation of the motif to compare to the sequence(s) of interest (Zhao and Lee 1997). While the motif approach may be powerful in many cases, the peptides of interest may not always have a common pattern because different amino acids may match in the same region of the sequence or different peptides should align to different parts of the sequence(s). Another approach would be to align every discovered peptide sequence to the protein sequence targets and add up the alignments scores at each position. The RELIC MATCH program (not currently available or supported) used this approach with some success (Cao and Mao 2009; Carter et al. 2006; Mandava et al. 2004; Takakusagi et al. 2008). This program also had several limitations in regard to transparency, flexibility, statistical analysis, and the ability to search multiple sequences. Here we will present an open source application that gives the user access to all parameters, can empirically estimate the statistical significance of the results, and enables the analysis of many sequences at once.

Methods

Algorithm Overview

The user inputs protein sequence(s) to search, a set of selected peptides, and (optionally) a representative or complete list of peptides from the library. A scoring matrix may be generated by the program as described below or entered by the user. The maximal local alignment between each selected peptide and protein sequence is found. If the alignment score is greater than the user defined score threshold, the score at each protein residue position is added to the protein residue scores. If the moving average window size is set to greater than one, after all peptides have been aligned to a given protein, the moving average across the protein residue positions is calculated and the residue scores provided correspond to the score at the start of the window. The same number of peptides as in the selected list are randomly selected from the library if a library set was entered, and aligned to the protein(s) as described above for the selected peptides; this process is repeated for the specified number of sampling iterations. If the subtract library scores box is checked, the average scores at each residue position from the randomly selected peptides from the library are subtracted from the residue scores. The selected peptide scores across each protein sequence are graphed, as well as the maximum and average scores from the random sampling iterations. The user may use the sort button to order the proteins by their maximal residue scores. The text output tab may be used to view a summary table of the maximum alignment scores for each protein or a table of all of the alignments identified for the number of proteins specified.

Scoring Matrix

GuiTope generates a log-odds-like scoring matrix based on a given measure of amino acid distances and amino acid frequencies. The distance matrix is taken to be inversely proportional to the frequencies of an amino acid pair appearing in a true alignment after a pseudocount of 10% of the average distance is added to the distance matrix to avoid dividing by zero. The rows and columns are iteratively scaled to sum to the expected amino acid frequencies. This matrix is then divided by the product of protein and peptide amino acid frequencies at each position and $\log_{10}$ transformed.

Alignment Algorithm and Inversion Scoring

The maximal gapless local alignment of each peptide with each protein is calculated using the Smith–Waterman algorithm. If the inversion weight is set to greater than 0, it will identify places where the protein sequence at position i is the same as the peptide at position j +1 AND the protein at position i+1 is the same as the peptide at position j. The residue scores for these inversions will be the product of the inversion weight and the average of the identity scores for the amino acids at the protein positions i and i+1.

Statistical Analysis

For each sampling iteration and each protein sequence, a set of peptides, with the same number of peptides as the selected peptide list, is randomly selected from the library and the residue scores are calculated. After these residue scores are

found for all iterations, the maximum and average residue scores are calculated for each position. If the 'subtract library scores' option is selected, the average library scores are subtracted from the residue scores from each iteration. The maximum scores from each protein iteration by iteration are ranked. For each protein, the maximum residue score from the selected peptides is compared to the ranked scores. The percentage of library scores that are higher than the selected peptide score is reported as the significance.

Evaluation Datasets

A set of peptide lists identified from a random-sequence peptide microarray selected to bind to monoclonal antibodies with known epitopes was previously described (Halperin, Stafford, and Johnston 2011). This dataset was used to optimize the alignment parameters. The polyclonal anti-peptide dataset from the same publication was used to evaluate the algorithm. Additionally, another set of monoclonal antibodies with known epitopes was used to probe a completely different set of 10,000 peptides on a microarray. The two anti-P53 antibodies from the first monoclonal antibody dataset were repeated on both the first and second version of the 10,000 peptide microarray. Additionally, an anti-cMyc clone 9E10 (AbD SeroTec, Raleigh), anti-Leu-Enkaphalin clone 1193/220 (AbD SeroTec, Raleigh), anti-PBEF clone E10 (Santa Cruz Biotechnology), and anti-V5 (AbD SeroTec, Raleigh) were also used to probe the array and generate peptide lists as described previously. Anti-cMyc, anti-Leu-Enkaphalin, and anti-V5 recognize epitope tags, while the anti-PBEF was epitope mapped using tiling peptides (current

authors, in preparation). Phage display datasets were selected from those listed in the "several binding sites" category in from Derda *et al.* (2011) that identified the greatest number of unique peptide and were downloaded from MimoDB. These phage display datasets include peptides selected to a diverse set of targets including two human extracellular proteins, one bacterial protein, and immune sera to a virus and a bacterium.

Implementation

GuiTope was implemented in Visual Basic, using the Microsoft .NET framework. It may be installed on any computer running Microsoft Windows XP or newer operating system. It has a memory footprint of 400mb and will take seconds to several minutes to run a set of hundreds of peptides against a single protein with 100 sampling iterations on a single Pentium 4 core, 3.2GHz and 2G RAM running Windows XP. On the same hardware, searching a protein database of ~20,000 proteins with a set of several hundred peptides with a single sampling iteration, it will utilize <3G (virtual) and use approximately 20 hours direct CPU time.

Results and Discussion

The optimal combination of parameters for GuiTope was tested on a previously described dataset of monoclonal antibodies with known epitopes that had been used to probe a random-sequence peptide array. Epitope predictions were evaluated using ROC analysis. The most critical parameter appears to be the scoring matrix, with the BLOSUM62 matrix having an AUROC 0.15 less than the GuiTope

method which adjusts for altered amino acid frequencies. The dipeptide inversion method also had a substantial improvement in the AUROC score. Since dipeptide inversion alignment is a novel approach, we subsequently evaluated the other datasets with and without using this method. The library subtraction method only yields a small improvement to the score, but a large number off sampling iterations are required to accurately estimate the average library score, we only used library subtraction for evaluating individual proteins rather than for database searches in order to keep run times reasonable.

Table 6.1 Parameter Optimization. The AUROC (Area Under the Receiver Operator Characteristics Curve) is shown for each parameter value tested on the 1st Known Epitope Monoclonal Dataset shown below. The best parameter value was highlighted and that value was used when each other value was varied.

Previous Methods	RELIC	Glam2	
	0.87	0.90	
GuiTope			
Scoring Matrix	Blosum62	GuiTope Method	
	0.75	0.90	
Score Cutoff (X average identity score)	3X	3.5X	4X
	0.84	0.90	0.76
Moving Average	None		5 aa
	0.90		0.89
Library Subtraction	No		Yes
	0.89		0.90
Inversion Weight	0	0.8	1
	0.81	0.84	0.90

GuiTope was tested on two independent datasets that probed random-sequence peptide microarrays with antibodies. Dataset number 1 used monoclonal antibodies with known linear epitopes to probe a set of 10,000 random-sequence peptides. Note that this dataset used a peptide microarray with 10,000 completely

different sequences than those used in the training set. The monoclonal epitopes could be predicted with an AUROC score of 0.75 using the inversion method and 0.78 without inversions, from a microarray of completely random peptides. The poorer performance on this dataset may indicate that the parameters were overfit on the training dataset, or that this set simply contains monoclonals that are more difficult to predict. The second peptide microarray evaluation dataset was generated from polyclonal anti-peptide sera. Here GuiTope performs similarly to previously tested methods with an AUROC of 0.68 using the inversion method and 0.56 without inversions, compared to an AUROC of 0.48 using RELIC method and 0.68 using Glam2 (Figure 6.1). These microarray datasets are likely considerably more difficult than phage display datasets because of sparse sampling of sequence space.

Phage display datasets to evaluate in GuiTope were selected based on the summary of the MimoDB published in Derda *et al.* (2011). Two of these datasets consisted of peptides selected to polyclonal sera. The phage display peptides selected against the anti-Nipah virus were used to map three epitopes on the nucleoprotein, and GuiTope also identified these epitope regions (Figure 6.2A). GuiTope also predicted an epitope on Glycoprotein G that was also predicted by DiscoTope, which uses the crystal structure to identify accessible regions (Figure 2B). GuiTope analysis would not have been successful at predicting the *Mycoplasma hyopneumoniae* lipoprotein epitopes that were later discovered (Meens et al. 2010). It is difficult to evaluate whether the epitopes predicted by GuiTope are plausible because crystal structures are not available.

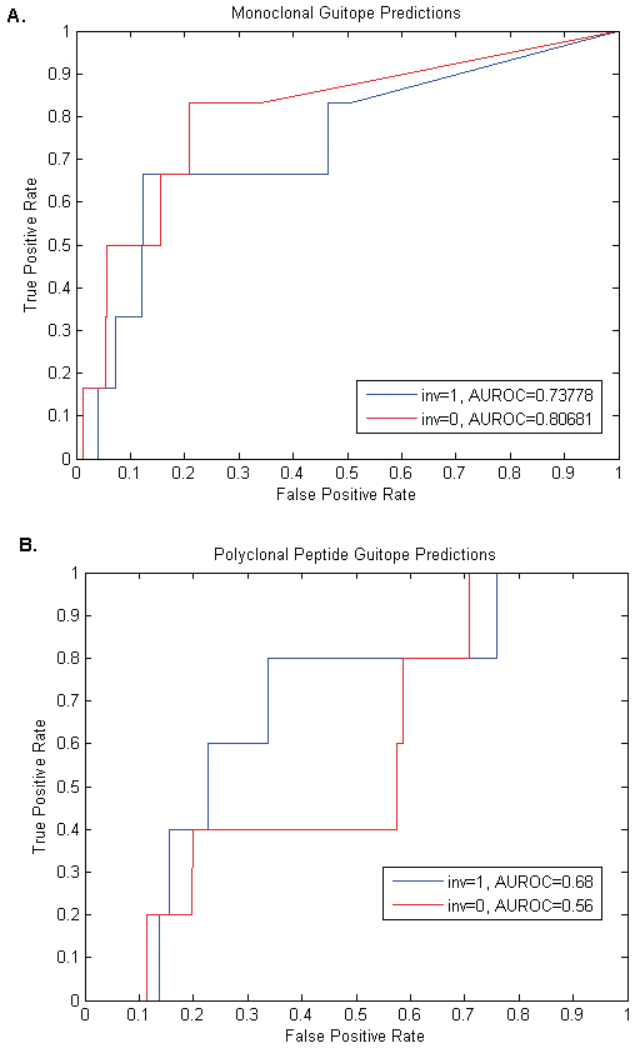


Figure 6.1 Peptide Microarray Evaluation Datasets. Peptides selected to bind known epitope monoclonals (A) or anti-peptide polyclonal sera (B) were used to predict the epitope (A) or immunizing peptide (B) in GuiTope within a database of decoy sequences. The significance scores of the true epitope or immunizing peptide sequences was compared to the decoy sequences using ROC plots, where the true positive rate is plotted against the false positive rate for all possible score thresholds. The results using the inversion weight as one are plotted in blue and the results without inversions are plotted in red. The AUROC value shown in the legend indicates the probability that true sequence would score higher than a decoy sequence for that dataset.

Table 6.2 Phage display database search

MimoDB/ Reference	**Target**	**Database (number of Proteins)**	**Known Interactor**	**Rank, p-value (Inv/No Inv)**
288 (White et al. 2005)	Endothelial protein C receptor	Human Extracellular and Cell Surface Proteins (5074)	Protein C	NA*
148 (Eshaghi, Tan, and Yusoff 2005)	Polyclonal Anti-Nipah Virus	Nipah Proteome (9)	Nucleoprotein	2, <0.1/ 2, <0.1
753,754,755 (Kraft et al. 1999)	Integerin α5β6	Human Extracellular and Cell Surface Proteins (5074)	TGF beta 1 TGF beta 3	2,<0.0002/ 1, <0.0002 3,<0.0002/ 2,<0.0002
204-205 (Yang et al. 2005)	Anti-*M. hyopneumoniae* polyclonal antibody	*Mycoplasma hyopneumoniae* Proteome (691)	Lipoproteins and p97	None matched correct region
1127 (Carettoni et al. 2003)	*Escherichia coli* FtsA	*Escherichia coli* (4311)	FtsA	1106,0.25/ 2067,0.58

Three protein panning datasets were also evaluated. In the first example, White *et al.* did not identify any similarity between the peptides found to bind to the endothelial protein C receptor (EPCR) and Protein C or any other known EPCR binding partners. GuiTope did not find any significant similarity between any known EPCR interactors either (Figure 6.3C). In the second case, the peptides selected to bind to integrin α5β6 were mapped by GuiTope to the known interactors TGF beta 1 and TGF beta 3 as two of the top three hits (Table 2) and correctly identified the important interacting amino acids (Figure 6.3B). Since these interactions were

discovered after the publication of the phage display study, one may suppose that they could have been predicted from the phage display data if the proper analysis tools had been available. In the third set, the peptides selected to bind to FtsA do not have a clear similarity to a known FtsA interactor. Carettoni *et al.* identified a weak motif that matched a site on FtsA, and used that site to develop a model for the structure of the FtsA dimer (2003). While several lines of evidence suggest that *E. Coli* FtsA does form a dimer, it is not clear whether the model proposed based on this phage-display data is correct (Adams and Errington 2009).

A.

B.

SEQ	SCORE	ALIGN LE	PRO POS	PEP POS	PRO WINDOW	PEP WINDOW
GSPNFTQ	7.651999	5	48	2	SafnT	SpnfT
FNVTQGP	7.088	4	50	1	FNtv	FNvt
FQINPTQ	6.698	7	69	1	iiqNyTr	fqiNpTq
SITNPTQ	6.684999	5	160	1	iscpn	sitnp
SNRTQGE	7.141	4	174	4	qteg	tgqe
NQTQHSV	7.137	6	174	2	QTegvs	QTqhsv
LGPNHTQ	6.869	4	183	1	gIPN	lgPN
IPPNPSQ	6.911	5	273	2	PPNPn	PPNPs
NPNQTRM	6.9575	5	274	1	pnpnT	npnqT
SGPNPTQ	6.8425	4	275	3	npnT	pnpT
YLPNVTQ	6.801	4	276	3	PNtv	PNvt
STPNVTQ	6.801	4	276	3	PNtv	PNvt
GSSLHDI	6.964001	4	311	1	sgSL	gsSL
NSTQGPV	8.857	7	446	1	dSIqqPV	nStqqPV
NSTQMRI	7.04	4	528	1	snqt	nstq
THANSTQ	7.04	4	528	4	snqt	nstq
NITQEAF	6.937	4	530	3	qtae	tqea
QTVKFAP	7.07	4	538	2	TVfk	TVkf

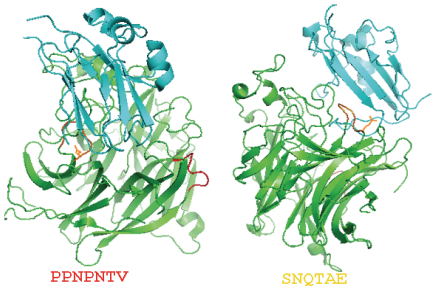
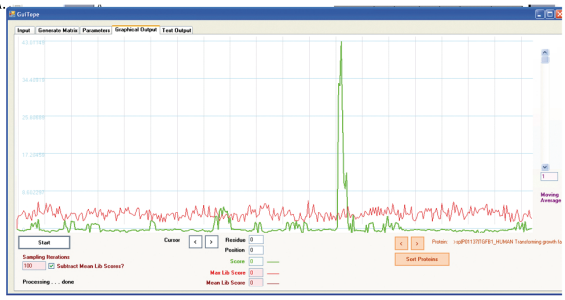


Figure 6.2 Analysis of Anti-Nipah Dataset. A. Screen shot of GuiTope mapping of anti-Nipah Virus selected peptides to the Nipah Nucleoprotein. Epitopes previously predicted and validated from this phage display peptide set are indicated with arrows. B. Novel GuiTope predictions using the inversion method of Nipah Glycoprotein G epitopes. The GuiTope alignment detail is shown as well as the locations of these epitopes in the crystal structure. The underlined glutamic acid is part of the receptor binding site.

A.



B.

SEQ	SCORE	ALIGN LEN	PRO POS	PEP POS	PRO WINDOW	PEP WINDOW
RGDLMLPFWPRL	7.974	6	12	4	LlplfW	LmlpfW
SSDLHALKKRYG	8.001	8	37	3	DmelvKrk	DlhalKkr
ASDISALSARLR	7.5215	6	54	7	LSkIrl	LSarlr
RGDAPSPNIFRL	6.805	5	60	4	AspPn	AqsPn
RGDYYANLANWL	7.687	4	102	3	DYYA	DYYA
HHELDSLRARLM	8.276	6	150	6	lsRAeL	sIRArL
TSDLETLRLRLH	10.014	11	151	2	SraElrLRLk	SdlEtrLRLh
RDDLHMLRLQLW	9.88	11	152	1	RaeLrtLRLkL	RddLhmLRLqL
GDLDLLKLRLTR	7.798	9	154	2	eLrLrLrLkL	dLdLLKLrL
VSDVYQLTERLR	6.7625	4	154	9	ElrI	Erlr
RDFQYLWNRVSQ	6.974	7	179	3	wrYLsNR	fqYLwNR
GDLHSLRQLLSR	8.796	7	204	6	rRQwLSR	IRQILSR
RGDLGYPRSLTL	7.542999	9	212	2	GeiegRls	GdlqypRsl
QSAHRGDIPWVL	*7.847*	*7*	*242*	*3*	*grRGDla*	*ahRGDip*
QRSDLASLKLTL	*8.172*	*9*	*243*	*1*	*rRgDLAtih*	*qRsDLAsIk*
RGDLKQLSELTW	7.363	7	244	1	RGDLati	RGDLkql
RGDLAALSAPPV	9.583	7	244	1	RGDLAti	RGDLAal
SHARGDLVGLTP	7.722001	9	244	4	RGDLatihg	RGDLvglltp
ILARGDLSTLSA	9.229	7	244	4	RGDLaTi	RGDLsTl
NARGDLGTLPPR	9.498001	7	244	3	RGDLaTi	RGDLgTl
RGDLPHQYQHPS	7.717	5	244	1	RGDLa	RGDLp
RGDLAAPHRTGA	9.561001	8	244	1	RGDLAtiH	RGDLAapH
RGDLPPVPMARE	7.856001	7	244	1	RGDLati	RGDLppv
RGDLPQERHYPR	8.09	12	244	1	RGDLatihgmnR	RGDLqaerhypR
RGDLLHASARAS	7.242001	4	244	1	RGDL	RGDL
VAPRGDRPTLPI	8.007	7	244	4	RGDIaTi	RGDspTI
RGDFAQLLITWQ	8.893	7	244	1	RGDIAti	RGDIAql
RGDIPSMSTLPR	9.365999	12	244	1	RGDIatihgmnR	RGDipsmstlpR
HPRTDLASLAKR	*7.816*	*7*	*244*	*3*	*RgDLAti*	*RdLAsI*
RDDLQTLLNRLG	*7.019*	*7*	*244*	*1*	*RgDLaTi*	*RdDLqTl*
RSDLTALTQLLL	*10.092*	*10*	*244*	*1*	*RgDLatihgm*	*RsDLtaltql*
KGDLATLRTNL	*10.164*	*7*	*244*	*1*	*rGDLATi*	*kGDLATl*
QSGDLYTLRNSR	*7.632*	*7*	*245*	*3*	*GDLaTih*	*GDLyTlr*
AGDLVTFKLRLH	*7.245*	*7*	*245*	*2*	*GDLaTih*	*GDLvTlk*
RSDLDTLLARHQ	7.596001	6	265	7	Lerarh	Llarhq
RGDLNNYLLMML	8.365	7	278	1	RaldtNY	RgdlnNY
RTDLHTLSSTRH	8.366	11	281	2	dtnycfSSTek	tdlhtlSSTrh
RHDLSELIRRTT	6.856	9	377	2	eqLSnmlvR	hdLSelIrR

C.

Figure 3. Protein interaction predictions. A. Peptides selected to bind Integrin AlphaV Beta6 clearly aligned in GuiTope to the integrin binding site on TGF beta 1. B. Detailed alignments of the peptides to TGF beta 1, with those that align to the binding site highlighted in yellow and those that do not contain the RGB motif shown in italic. Below the WebLogo view of peptides aligning to the region illustrates the relative importance of amino acids C. Peptides selected to bind to EPCR do not align to a particular region on protein C.

The peptides that bind to a given target do not always have sequences that are similar to biologically relevant proteins. This problem is confounded when peptide array approaches are used because peptides that are highly similar to a given protein are unlikely to be present in the library. GuiTope was able to take these loosely similar sequences and predict antibody epitopes with modest accuracy (AUC 0.75-0.9) in line with previously tested methods (Halperin, Stafford, and Johnston 2011). Random-sequence peptide microarrays have shown great promise in profiling the humoral immune response (Legutki et al. 2010; Restrepo et al. 2011), and it would be of great utility to be able to use the peptide sequences to trace back to the

antigen that elicited the immune response. However, the current prediction accuracy would not be sufficient for this task (Halperin, Stafford, and Johnston 2011). In contrast to the peptide array datasets, the phage display selected peptides can sometimes be used to predict interaction partners from a database very accurately. As less biased methods of molecular display methods are developed and higher density peptide arrays become available, we expect that the information content of the peptide sequences will improve, making the type of analysis facilitated by GuiTope even more useful.

Conclusions

GuiTope is a flexible and easy to use software application for comparing peptide to protein sequences. Although the RELIC MATCH program served to some extent as a model, the additional features and open source availability set GuiTope apart. The accuracy of epitope prediction is similar to previously described methods but GuiTope has other advantages including statistical analysis of results. It also allows searching peptides against protein databases, which has utility in predicting antibody targets and protein binding partners. Peptide library selection experiments do not always yield sequences that are similar to biologically relevant proteins, but when they do, GuiTope provides an efficient method to identify these relationships.

Availability and requirements

The executable is available on www.immunosiganture.com/software and will install and run on any PC with Windows XP or later. The source code is written in

Visual Basic and available on sourceforge.net. The Microsoft .NET framework is required.

References

Adams, D. W. and J. Errington. 2009. Bacterial cell division: Assembly, maintenance and disassembly of the z ring. *Nature Reviews. Microbiology* 7, no. 9: 642-53.

Bastas, G., S. R. Sompuram, B. Pierce, K. Vani, and S. A. Bogen. 2008. Bioinformatic requirements for protein database searching using predicted epitopes from disease-associated antibodies. *Molecular & Cellular Proteomics* 7, no. 2: 247-56.

Breitling, Frank, Alexander Nesterov, Volker Stadler, Thomas Felgenhauer, and F. Ralf Bischoff. 2009. High-density peptide arrays. *Molecular BioSystems* 5, no. 3: 224-234.

Cao, B. and C. Mao. 2009. Identification of microtubule-binding domains on microtubule-associated proteins by major coat phage display technique. *Biomacromolecules* 10, no. 3: 555-64.

Carettoni, D., P. Gomez-Puertas, L. Yim, J. Mingorance, O. Massidda, M. Vicente, A. Valencia, E. Domenici, and D. Anderluzzi. 2003. Phage-display and correlated mutations identify an essential region of subdomain 1c involved in homodimerization of escherichia coli ftsa. *Proteins* 50, no. 2: 192-206.

Carter, D. M., J. N. Gagnon, M. Damlaj, S. Mandava, L. Makowski, D. J. Rodi, P. D. Pawelek, and J. W. Coulton. 2006. Phage display reveals multiple contact sites between fhua, an outer membrane receptor of escherichia coli, and tonb. *Journal of Molecular Biology* 357, no. 1: 236-51.

Derda, R., S. K. Tang, S. C. Li, S. Ng, W. Matochko, and M. R. Jafari. 2011. Diversity of phage-displayed libraries of peptides during panning and amplification. *Molecules* 16, no. 2: 1776-803.

Eshaghi, M., W. S. Tan, and K. Yusoff. 2005. Identification of epitopes in the nucleocapsid protein of nipah virus using a linear phage-displayed random peptide library. *Journal of Medical Virology* 75, no. 1: 147-52.

Halperin, Rebecca F., Phillip Stafford, and Stephen Albert Johnston. 2011. Exploring antibody recognition of sequence space through random-sequence peptide microarrays. *Molecular & Cellular Proteomics* 10, no. 3.

Kraft, S., B. Diefenbach, R. Mehta, A. Jonczyk, G. A. Luckenbach, and S. L. Goodman. 1999. Definition of an unexpected ligand recognition motif for alphav beta6 integrin. *Journal of Biological Chemistry* 274, no. 4: 1979-85.

Legutki, J. B., D. M. Magee, P. Stafford, and S. A. Johnston. 2010. A general method for characterization of humoral immunity induced by a vaccine or infection. *Vaccine* 28, no. 28: 4529-37.

Mandava, S., L. Makowski, S. Devarapalli, J. Uzubell, and D. J. Rodi. 2004. Relic--a bioinformatics server for combinatorial peptide analysis and identification of protein-ligand interaction sites. *Proteomics* 4, no. 5: 1439-60.

Meens, J., V. Bolotin, R. Frank, J. Bohmer, and G. F. Gerlach. 2010. Characterization of a highly immunogenic mycoplasma hyopneumoniae lipoprotein mhp366 identified by peptide-spot array. *Veterinary Microbiology* 142, no. 3-4: 293-302.

Nie, J., B. Chang, D. O. Traktuev, J. Sun, K. March, L. Chan, E. H. Sage, R. Pasqualini, W. Arap, and M. G. Kolonin. 2008. Ifats collection: Combinatorial peptides identify alpha5beta1 integrin as a receptor for the matricellular protein sparc on adipose stromal cells. *Stem Cells* 26, no. 10: 2735-45.

Restrepo, Lucas, Phillip Stafford, D. Mitch Magee, and Stephen Albert Johnston. 2011. Application of immunosignatures to the assessment of alzheimer's disease. *Annals of Neurology* 70, no. 2: 286-295.

Rodi, D. J., R. W. Janes, H. J. Sanganee, R. A. Holton, B. A. Wallace, and L. Makowski. 1999. Screening of a library of phage-displayed peptides identifies human bcl-2 as a taxol-binding protein. *Journal of Molecular Biology* 285, no. 1: 197-203.

Stephen, C. W. and D. P. Lane. 1992. Mutant conformation of p53. Precise epitope mapping using a filamentous phage epitope library. *Journal of Molecular Biology* 225, no. 3: 577-83.

Takakusagi, Y., K. Kuramochi, M. Takagi, T. Kusayanagi, D. Manita, H. Ozawa, K. Iwakiri, K. Takakusagi, Y. Miyano, A. Nakazaki, S. Kobayashi, F. Sugawara, and K. Sakaguchi. 2008. Efficient one-cycle affinity selection of binding proteins or peptides specific for a small-molecule using a t7 phage display pool. *Bioorganic & Medicinal Chemistry* 16, no. 22: 9837-46.

Ullman, Christopher G., Laura Frigotto, and R. Neil Cooley. In vitro methods for peptide display and their applications. 2011. *Briefings in Functional Genomics* 10, no. 3: 125-134.

White, S. J., R. E. Simmonds, D. A. Lane, and A. H. Baker. 2005. Efficient isolation of peptide ligands for the endothelial cell protein c receptor (epcr) using candidate receptor phage display biopanning. *Peptides* 26, no. 7: 1264-9.

Yang, W. J., J. F. Lai, K. C. Peng, H. J. Chiang, C. N. Weng, and D. Shiuan. 2005. Epitope mapping of mycoplasma hyopneumoniae using phage displayed peptide libraries and the immune responses of the selected phagotopes. *Journal of Immunological Methods* 304, no. 1-2: 15-29.

Zhao, S. and E. Y. Lee. 1997. A protein phosphatase-1-binding motif identified by the panning of a random peptide display library. *Journal of Biological Chemistry* 272, no. 45: 28368-72.

CHAPTER 7

CONCLUSIONS

Monoclonal antibodies have proven to be a useful tool in characterizing the immunosignaturing platform. Individual antibody signatures have proven to be detectable at low nanomolar to picomolar concentrations and are largely unaffected by competition from naïve sera. A polyreactive antibody appears to bind more peptides, but with lower apparent affinity than antibodies raised to specific antigens. Some trends were observed with respect to sequence properties, but these do not account for very much of the variance in overall binding. Sequence similarity to the antibodies' epitope did not drive binding very much either, though there was some predictive power in mapping back to the epitope. Software was developed to facilitate this mapping between random-sequence peptides and natural proteins sequences. The utility of the sequence analysis approaches is expected to increase as higher density peptide arrays are available.

The most basic question underlying the immunosignaturing concept, is do antibodies bind the random-sequence peptide arrays. Phage display experiments had previously demonstrated that it was possible to identify peptides that bound to diverse antibodies and were not necessarily similar to the epitope (Meloen, Puijk, and Slootstra 2000). However, phage display libraries are several orders of magnitude larger than peptide libraries that are feasible with current technology. From these studies, it seemed as though mimotope peptides were extremely rare, and large libraries were needed to find uncommon peptide(s) that the antibody of interest

would recognize (Rodi and Makowski 1999). However, the peptides resulting from phage display experiments do not necessarily reflect the best or only binders. In the amplification phase, peptides that give the phage even a slight growth advantage may outcompete other phage even if they are more highly enriched in the selection phase (Derda et al. 2011). One previous study was able to show that a random-sequence peptide array could be used to identify antibody binders (Reineke et al. 2002). My work in Chapter 1 examines a large set of monoclonal antibodies on the immunosignaturing arrays.

I found that all of the antibodies that I have tested bound to some of the random-sequence peptides above background. Importantly, each antibody appears to have a unique pattern of peptide binding. What I find most interesting in this set of experiments is that the antibodies that I have tested exhibit such a wide range of binding distributions. Some bind as much as 70% of the peptides on the array and others bind only 0.1% of the peptides. What accounts for this difference remains an open question. One trend that I have observed is that antibodies that have higher background binding to the slide surface, also tend to bind more peptides above that background level (Figure A-1). One possibility is that some antibodies have a nonspecific attraction to the slide surface, resulting in a higher local concentration at the slide surface, which leads to a higher effective concentration seen by the peptides. Another explanation is that these antibodies bind the slide surface through a paratope mediated interaction. This weak interaction could lead to enhanced affinity for the peptides through a heteroligation model similar to what was observed with an anti-HIV monoclonal (Mouquet et al. 2010). This hypothesis could be tested

by comparing the binding of a high background/high binding antibody to the Fab fragment of the same antibody. A second trend that I observed is that antibodies that are able to recognize their antigen in a denatured state tend to bind to more peptides on the array than those that recognize a conformation dependent epitope (Appendix A-2). Surveys of antibody crystal structures have found that those that recognize peptides tend to have a more grooved paratope than those that recognize proteins (Chen, Van Regenmortel, and Pellequer 2009). It is possible that antibodies with a grooved paratope are more adept at binding peptides in general.

There is still much work to be done in elucidating what types of antibodies bind better to the peptide array and contribute more to the immunosignature. An important question that I have not addressed is the relationship between antibody affinity for its cognate antigen and its tendency to bind random-sequence peptides. The direction of this relationship is difficult to predict. On one hand, higher affinity antibodies could simply be better binders and also recognize other targets better. On the other hand classically polyreactive antibodies tend to have flexible paratopes enabling the recognition of diverse shapes, but at an entropic cost resulting in lower affinity. Addressing this question would be important with respect to whether immunosignaturing is better able to pick up antibodies arising in a primary response, or those that have undergone affinity maturation.

Most of the antibodies that I have worked with have been raised against protein or peptide targets. However, antibodies recognizing glycans can be very important in the immune response against a wide variety of pathogens and tumors (Astronomo and Burton 2010). Antibodies recognizing other protein modifications,

such as glycosylation and citrullination, and nucleic acids, can be important in a variety of autoimmune and inflammatory disease states (Dotan et al. 2006; Klareskog et al. 2008; Yung and Chan 2008). In the set of monoclonals I have tested, only one was known to be against a non-protein target (anti-TP). This antibody recognizes a glycan epitope on the fungus *coccidioidomycosis*. It behaved similarly to many of the anti-protein antibodies, recognizing over a thousand peptides above background. Though there is much precedent for peptides to be able to mimic diverse non-protein targets in the phage display literature (Meloen, Puijk, and Slootstra 2000), it would be unwise to assume that this should directly translate to our arrays. Testing a diverse set of monoclonals against a variety of non-protein targets would be important to better understand the capabilities and limitations of the current immunosignaturing platform. If any type of non-protein directed antibody was found to not be detected well on the peptide array, appropriately modified peptides could be included on the array to increase the chances of detecting such antibodies.

A critical question in assessing the power of the immunosignatuing technology is the sensitivity to small quantities of antibodies. A healthy human has typically ~4-12mg/ml of immunoglobulin in their serum, which translates to 50-160nM immunoglobulin on the array when the serum is diluted at 1:500 in the standard protocol. While most of my monoclonal experiments were done at 100nM, in the range of the total sera antibody concentration, sera is far from monoclonal. During a specific immune response, about 100ug/ml of immunoglobulin against the antigen may be produced (Mei et al. 2009), corresponding to a ~1nM concentration when diluted on the array. I have looked at two antibodies diluted out on the entire

10K array and another two on a small set of peptides selected to bind to those antibodies. For the P53Ab1 antibody, 7 peptides were detectable above background at 1nM, and one peptide was even detectable at 100pM. The 87G antibody did not bind to as many of the random-sequence peptides, but still had two peptides above background at 1nM. The anti-PBEF was found to bind several of the selected peptides below 1nM. The polyreactive antibody was only able to bind to peptides above background down to 5-10nM. From these antibodies it appears that the specific antibodies will still have some peptides detectable below 1nM. Although the polyreactive antibody also binds to 57% of peptides above background at 100nM, the signal drops off more quickly with concentration, and this antibody will be discussed in more detail later. If the antibodies I have tested are a fairly representative set, I would expect most specific antibodies to be detectable below 1nM. However, a larger sampling of monoclonal antibodies should be used in concentration series to get a better idea of how many may be detected the relevant concentrations. Since the amount of antibody produced to a specific antigen would be diluted to about 1nM in the standard immunosignaturing protocol, these results suggest that we should be able to detect most specific antibodies produced in an immune response.

In addition to the low concentration of individual antibodies in the serum, there is also a large complexity of different antibodies in the serum. An important consideration is whether antibodies are competing for a limited number of binding sites on the spots and if so, how the competition shapes the pattern detected. Two monoclonals were mixed in varying proportions and detected with the same

secondary antibody. One antibody (P53Ab1) bound many more peptides than the other (P53Ab8). P53Ab1 specific peptides have fairly constant signals as the antibody was diluted into the other, while some of the P53Ab8 peptides start to decrease in signal intensity when the P53Ab8 is only 40% to 20% percent of the mixture. Eight monoclonals were also mixed in equimolar proportions and the signal intensities were compared to the signals of those eight run individually and averaged. The correlation between the physically mixed and the averaged signals was low (r=0.59). Some of the noise is probably due to differences in binding curves. From the dilution experiments one can see that some peptides drop off in signal with concentration more rapidly than others. It would be interesting to see if a set of monoclonals were all run at a low concentration and then pooled at that concentration, if the sum of the signal intensities would correlate with the pooled signal intensities. I would predict that the sum of the signal intensities would correlate well with the pooled antibody signal intensities. If this high correlation was observed, it would imply that the patterns of individual antibodies are additive. Additivity of binding patterns would greatly simplify the deconvolution of multiple disease states in one patient.

While eight monoclonals are a substantial increase in complexity compared to one, it does not begin to approach the antibody complexity in sera. To get a better idea of how the complexity of serum antibodies would effect an individual clone, I diluted a mouse monoclonal (P53Ab1) at 100nM into 10 fold or 100 fold excess pooled human IgG. Surprisingly, I did not see any significant change in the binding pattern with or without the excess IgG. While it is tempting to conclude that the

monoclonal antibody out competes the normal serum antibodies, I do not believe that to be the case. It is important to note that the addition of the monoclonal does not change the binding pattern of the human sera either (Figure A-7). I believe this indicates that the binding sites for antibodies on the spots are not being saturated so the antibodies are not competing for binding sites. This idea is supported by the dilution series data that illustrates that most peptides do not appear to be approaching saturation at 100nM for the monoclonal. The pooled naïve immunoglobulin also continues to increase in signal intensity above 100nM (Figure A,-8).

While the previous experiment was important to establish the effect of competitor antibody on a monoclonal, perhaps a more realistic reconstruction experiment would involve lower antibody concentrations. For example, a dilution series of a monoclonal going from 100nM down to 100pM in the presence of 100nM naïve immunoglobulin would more closely mirrors the dynamics of a real infection. Since competition effects were not observed in the high concentration competition experiment, I would be surprised if they were observed in an experiment at more realistic concentrations. It would be good make sure that the presence of naïve IgG does not somehow interfere with detecting an antibody at low concentrations as well.

Based on the theoretical density of attachment sites on aminosilane, we would expect peptides to be close enough for antibodies to bind bivalently. The binding enhancement from bivalent binding can be multiplicative, which seemed to be a likely explanation for being able to detect so much binding to random sequence peptides. In order to examine the effect of the peptide spacing, I created custom

microarrays were peptides are intentionally printed at different concentrations. As would be expected, the signal intensity drops off with lower concentrations. Another way to control the peptide density is to use NSB slides that have peptide attachment sites precisely spaced at 3nM or 9nM. Less binding was observed on the 3nM slides than the aminosilane, and very little binding was observed on the 9nM slides. While there are likely differences in the surface properties of the aminosilane compared to the NSB slides, this data also supports the idea that the peptide spacing is critical to observing binding. Surprisingly, when I compared the binding of an Fab fragment to that of the intact antibody of the same clone, I found very similar binding patterns. The effect of the peptide spacing must not be due to antibodies binding bivalently, but more likely due to the high local concentration of the peptide leading to rebinding effects.

Differences in peptide density among peptides and between print runs are likely a source of noise. Several methods were tested to measure the relative peptide density on the slide surface. Amine reactive dyes were found to preferentially label positively charged peptides and not detect negatively charged peptides very well. Amine reactive biotin appeared to label peptides without any particular bias for peptide properties. However, the biotinylation did have relatively high background, likely due to residual free amines that had not reacted with the linker. It also was not able to detect all peptides above background, some of which had observable binding with sera. Despite these apparent limitations, I developed an algorithm to normalize the binding experiments using the biotinylation data. The normalization was able to

give small but consistent improvements in batch to batch reproducibility (Appendix E).

Batch to batch variability in peptide array production is a persistent problem facing the immunosignaturing technology. I have seen great improvements in the reproducibility during my time in CIM. Automation of the slide processing using the Tecan system is probably the biggest source of improvement. Other changes to the process including the printing technology, spotting buffers, and improved peptide handling have each contributed to the better reproducibility. Despite these improvements, the batch to batch reproducibility issues still make performing large studies difficult. Continuing sources of noise probably include degradation or aggregation of the peptides over time, variability in the aminosilane coating, differences in activation efficiency, and variable printing of the full length peptide compared other peptide synthesis byproducts that are present. While getting the peptides purified would obviously address the last issue, there does not seem to be any readily available solution for the other sources of noise. The *in situ* synthesis would likely eliminate many of these sources of noise by avoiding the need to store peptides in solution. Preliminary experiments show the reproducibility of the biomicrochips to be in the same range as the slides. However, the biomicrochips were processed manually, so if automation of the chip processing has as big of an improvement in reproducibility has was seen on the slides, the chips would clearly be more reproducible.

The goal of immunosignaturing is to take the pattern identified in a particular disease state and use it as a diagnostic. However, current regulations do not permit

the inclusion of markers not relevant to a disease state in a diagnostic test. The reproducibility issues of the current platform also make the current format unappealing as a clinical diagnostic. Others in the lab are currently developing methods for printing and processing custom arrays with disease specific panels of peptides. While working with a smaller set of peptides may allow better control over the peptide handling, which may lend some improvement in reproducibility, it does not address most of the sources of noise on the 10K platform. In any solid phase assay, there is likely to be some variability in immobilization efficiency and differential non-specific binding to the surface between samples. Solution phase assays have been shown to have improved sensitivity and specificity compared to solid phase assays (Liu and Eisenbarth 2007). However, the peptides probably have relatively poor solution phase affinity for the antibodies, which could make solution phase detection difficult. It is possible that using a detector with high sensitivity could enable measurement of these interactions. Another approach could be to identify amino acid substitutions that improve affinity of the peptides for the antibodies. While the array format may be ideal for high throughput screening, translating the results into a diagnostic may require exploring other assay formats, particularly for diseases where the differences in the immunosignatures are subtle.

Polyreactive antibody

Polyreactive antibodies are produced in the absence of any specific antigen and play an important role in host maintenance and defense (Zhou, Tzioufas, and Notkins 2007). I sought to compare the reactivity of specific and polyreactive

antibodies to random-sequence and protein sequence derived peptides. I found that the polyreactive antibody had a broader distribution of binding to the 10K peptide array than an isotype matched specific antibody (anti-PBEF) or any other specific antibody I have tested. I selected a set of peptides that preferentially bound the polyreactive antibody, the anti-PBEF, or both antibodies. I also had peptides tiling the region of PBEF that the antibody was raised against synthesized as well as peptides tiling AKT1 as negative control. These peptides from the random-sequence array and the protein sequences were printed together in a custom peptide array. The two antibodies were used to probe the array at concentrations ranging from 100nM to less than 1nM. The binding curves were fit to estimate half maximal binding concentrations. While I would not expect the half maximal binding to correspond to the solution phase dissociation constant, it should give a better estimate of relative affinity than the signal intensity at a signal concentration point (Tapia et al. 2007). Three overlapping PBEF peptides were found to bind anti-PBEF in the 5-20nM range, while the selected random-sequence peptides bound the anti-PBEF 20-60nM. In contrast, none of the peptides appeared to approach saturation for the polyreactive antibody, so the half-maximal binding was estimated to be >100nM for about half of the protein tiling peptides. This dataset supports the model that polyreactive antibodies bind to more targets, but with lower affinity than antibodies raised to specific targets.

The observation that the polyreactive antibody recognizes multiple peptides tiling a protein sequence may imply that it recognizes multiple epitopes on the same protein. However, the recognition of a peptide does not necessarily mean that the

antibody would recognize that sequence in a folded structure. A survey of crystal structures of antibody-peptide complexes found that most peptides bound in conformations that would result in steric clashes in the folded proteins (Chen, Van Regenmortel, and Pellequer 2009). If several of these regions are accessible on the folded protein, it would imply that it might be possible for the polyreactive antibody to bind to multiple epitopes simultaneously. Multivalent binding would greatly increase the avidity of the antibody for the target, and have a large effect on the ability of it to bind *in vivo*.

Polyreactive antibodies play important roles in health and disease, and in some cases their presence or absence may be indicative of a disease state (Merbl et al. 2009; Shimomura et al. 2008; Zhang et al. 2009). While the polyreactive antibody bound a large percentage of peptides on the array at 100nM, the signals dropped of quickly at lower concentrations for the small number of peptides tested on the custom array. If most of the other peptides behave similarly, binding of polyreactive antibodies would be difficult to detect on the immunosignaturing platform. Extending the incubation time appears to enhance the detection of the polyreactive antibody more so than the specific antibody (Figure B-17). This difference in kinetics could be explained by the flexibility of polyreactive antibody paratopes. The polyreactive antibody may take longer to find the right conformation to bind a peptide, thus resulting in a slower on-rate. If polyreactive antibodies are thought to be important in a particular disease state, increasing the incubation time in the immunosignaturing protocol may provide a simple way to enhance their detection.

While the kinetic discrimination appears to be a plausible mechanism to enhance or reduce the contribution of polyreactive antibodies to the immunosignature, a more direct measure of their role would be desirable. A method to enrich or deplete polyreactive antibodies from serum was used in Zhou et al. (2007). Applying the enriched and depleted fractions to the peptide array would enable the analysis of exactly how much polyreactive antibodies are contributing to immunosignatures. This fractionation would also enable the evaluation of the hypothesis that increasing incubation time would enhance the detection of polyreactive antibodies as well as explore other methods of discriminating between polyreactive and specific antibodies.

Sequence Properties

I sought to determine the extent to which peptide properties could predict antibody binding. A variety of sequence properties have previously been shown to be associated with antibody epitopes, including hydrophilicity (Parker, Guo, and Hodges 1986), flexibility (Karplus and Schulz 1985), accessibility (Emini et al. 1985), beta turn propensity (Chou and Fasman 1978), and antigenicity (Kolaskar and Tongaonkar 1990). In addition to these properties, I also examined isoelectric point, residue volume, and sequence complexity. We have used two different 10K peptide libraries for immunosignaturing and I have included datasets from both libraries in this analysis. Another peptide microarray technology is under development by HealthTel in collaboration with CIM, which enables the production of much higher density arrays. I was able to test a small number of these arrays which had 100K

peptides, with 12mer peptides containing eight different amino acids. The 100K chip had substantially less sequence complexity than the 10K arrays do to the limited amino acid set, and also had a moderate increase in residue volume. Otherwise, the distributions of the peptide properties were similar across all three platforms. Overall the strongest correlations were observed between peptide isoelectric point and antibody binding, though the size and direction of the correlation varied greatly between antibodies. The largest correlations observed were less than 0.5, indicating that isoelectric point explains less than 25% of the variance in antibody binding. The 10Kv1 array and 100K chip, antibody binding generally has a negative correlation with accessibility, flexibility, and hydrophilicity, while on the 10Kv2 binding was typically positively correlated with these properties. On the 100K chip, there was a moderate correlation with sequence complexity for all three antibodies. The denser sampling of sequence space on the 100K array allowed the analysis of the peptides by repeated 5mers. One of the antibodies was found to preferentially recognize a set of similar 5mers.

While trends were observed with respective to antibody binding levels and peptide properties, perhaps the more important question for immunosignaturing is how peptides could be identified that would be highly specific. I did generate specificity scores for each peptide/antibody pair and examined the correlations with peptide properties. There was more variability in the size and direction of the correlations between the specificity and the properties than the binding. This variability is not surprising, as peptides can only be specific for one antibody. For example, if one antibody strongly prefers peptides with higher isoelectric points, then

the other antibodies would not have high specificity scores for the high isoelectric point peptides. Perhaps a better metric for examining what peptide properties are important for informative peptides would be dynamic range or signal to noise ratio.

An important question for immunosignaturing is how to better detect antibodies that were not detected well on the current platform. For example, the anti-MHC antibody bound very few peptides above background and none were found to be highly specific for that antibody. This antibody only recognizes the MHC beta2-microglobulin quartenary structure, and will not bind to either domain by itself. I have also observed that antibodies that are known to recognize their target denatured in a western blot tend to bind to more peptides on the array than those that don't, so it may be more difficult for antibodies to highly structured targets to find good mimotopes. It is possible that some antibodies may be better detected by more structured targets. Cyclized peptides libraries have been shown to provide higher affinity ligands for some targets (Katz 1997). These libraries generally utilize Cys to cyclize peptides through a disulfide bond. This strategy would not be feasible on our current platform because we use Cys to attach the peptides to the slide. Perhaps an orthogonal chemical reaction could be used to design cyclic peptides or different attachment chemistry could be used so that Cys could be used for cyclization. First, more antibodies with well characterized structural requirements for binding should be tested on the peptide arrays to determine if recognition of structure is really a limitation.

In designing the next generation of peptide libraries, there will be an important tradeoff to consider with respect to redundancy vs. complexity. Having

5mer or longer sub-sequences repeated on the array has obvious advantages for identifying an antibody epitope. Of course it would be very difficult to identify an epitope if the amino acids contained in that epitope are not present on the array, and the more amino acids are included, the less redundancy there will be. As I have discussed previously, there may be advantages to including non-canonical or modified amino acids on the array, but this would also take space away from repetition of potential linear epitopes. Of course, the cost and ease of synthesis would also be an important consideration in the design.

Epitope Mapping

The ability to take an immunisignature and use the peptide sequences to map back to the antigen that raised the immune response would be incredibly useful for gaining biological insight into an immunosignature. I used a set of monoclonals with known epitopes to evaluate the feasibility of this mapping. Surprisingly, peptides that were the best binders for each antibody were generally not enriched for peptides that had sequence similarity to the epitope sequence. In spite of this negative result, I tested two methods for predicting epitopes from these peptide sequences. The first was a motif based approach using the glam2 program (Frith et al. 2008) and the second was an alignment approach using the RELIC program (Mandava et al. 2004). I found that while there was some predictive power using these approaches, they would not be sufficient to identify an unknown antigen from a database.

The RELIC webserver was taken offline sometime in the spring of 2011 which motivated me to further develop an alternative program for the same task.

While the GuiTope program had originally been written with the intention of improving on the predictive power of the RELIC program (see Appendix D), but my attempts were unsuccessful. I did make several improvements to the functionality. The GuiTope program is more transparent and flexible than the RELIC program as the algorithm is described in detail in a publication (submitted), it is open source, and the user has access to all of the parameters. It also enables the searching of a set of proteins and estimates the statistical significance of results. Additionally, as it is a standalone program, the user does not need to be concerned with long queues on a webserver, as was often a problem with RELIC. While GuiTope does not make a significant improvement in the prediction power, it provides a flexible platform for implementing future improvements.

A novel alignment aspect discovered during the development of GuiTope was that dipeptide inversions tended to appear in alignments (ex. RS aligns with SR). When the alignment algorithm was modified to specifically search for sequence inversions, I found an improvement in epitope prediction accuracy for two of the three known epitope datasets tested. How the dipeptide inversion recognition works structurally is difficult to explain. Some experimental evidence supporting the dipeptide inversion recognition was found using peptide immune sera. Sera raised against the FT02 peptide was found to crossreact against the FT03 peptide and vice versa. These peptides were found to have a region of similar sequence that contained a dipeptide inversion (Figure I-1). In a valley fever dataset, I have found that small amino acids tend to occur more often in the inversions than in large ones (Figure I-2). Small amino acids may allow more flexibility in arranging the peptide

chain in the paratope. To begin to understand how this recognition works and in what context inversions may be recognized, libraries of inversion mutants could be made and the relative binding tested. If some inversion mutants are found to bind well, crystal structures of inversion mutants compared to the wild type crystal structures could enable seeing how these inversions are recognized.

Predicting the epitope of an unknown antigen from the peptide array data is a very difficult problem for several reasons. First, searching a database requires extremely good accuracy in order for the true antigen to be among the top hits. For example, the UniRef100 database currently has over 13 million protein sequences. Even with an AUROC score of 0.99, about 130 thousand other sequences would be expected to score better than the true epitope. Usually, biological information would be available such as organism, molecular weight of the protein, and/or sub-cellular localization that could help to narrow down the search space. The second difficulty is the sparse sampling of sequence space on the current 10K arrays, which have less than 5% of all possible 5mers present. This sampling would be substantially improved with the recent capability to make 100K chips. For example, 100K 17mers with 19 amino acids could cover more than half of all possible 5mers. Finally, the current assay does not typically enrich for sequence similar peptides among the binders. This lack of enrichment is much more difficult to address.

It is possible that changes to the experimental parameters could lead to better enrichment for epitope similar peptides. I think that the best way to approach this problem would be to select a set of peptides from the array that are similar to several known epitopes and exhibit a wide range of binding levels, and also include a set of

peptides that bind well to those antibodies but don't have any sequence similarity. The binding kinetics between these peptides and antibodies should be measured. If any differences are found between the kinetics of epitope similar peptides and mimotope peptides, the assay temperature, incubation time, and/or wash time could be varied to favor the epitope peptides.

The best approach for improving the prediction algorithm is not obvious. The simplest approach is probably to use a machine learning algorithm to optimize the alignment parameters. Previous attempts at this approach were not successful, most likely because the training set was too small. A larger training may enable this approach to be more successful. A more thorough exploration of the best way to approach the best motif finding algorithm and parameters would also be warranted. Incorporating structural predictions could potentially enhance the prediction capability. However, since bioinformatics structural predictions are limited in accuracy, it is hard would be difficult to imagine that these could greatly enhance the prediction accuracy. I believe that the greatest improvements in epitope prediction will likely result from higher density chips and from improvements in the assay, rather than from more sophisticated algorithms.

References

Astronomo, Rena D. and Dennis R. Burton. 2010. Carbohydrate vaccines: Developing sweet solutions to sticky situations? *Nat Rev Drug Discov* 9, no. 4: 308-324.

Chen, S. W., M. H. Van Regenmortel, and J. L. Pellequer. 2009. Structure-activity relationships in peptide-antibody complexes: Implications for epitope prediction and development of synthetic peptide vaccines. *Curr Med Chem* 16, no. 8: 953-64.

Chou, P. Y. and G. D. Fasman. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Advances in Enzymology & Related Areas of Molecular Biology* 47: 45-148.

Derda, R., S. K. Tang, S. C. Li, S. Ng, W. Matochko, and M. R. Jafari. 2011. Diversity of phage-displayed libraries of peptides during panning and amplification. *Molecules* 16, no. 2: 1776-803.

Dotan, N., R. T. Altstock, M. Schwarz, and A. Dukler. 2006. Anti-glycan antibodies as biomarkers for diagnosis and prognosis. *Lupus* 15, no. 7: 442-450.

Emini, E. A., J. V. Hughes, D. S. Perlow, and J. Boger. 1985. Induction of hepatitis a virus-neutralizing antibody by a virus-specific synthetic peptide. *Journal of Virology* 55, no. 3: 836-9.

Frith, M. C., N. F. Saunders, B. Kobe, and T. L. Bailey. 2008. Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Computational Biology* 4, no. 4: e1000071.

Karplus, P. A. and G. E. Schulz. 1985. Prediction of chain flexibility in proteins. *Naturwissenschaften* 72, no. 4: 212-213.

Katz, Bradley A. 1997. Structural and mechanistic determinants of affinity and specificity of ligands discovered or engineered by phage display. *Annual Review of Biophysics and Biomolecular Structure* 26, no. 1: 27-45.

Klareskog, L., M. Widhe, M. Hermansson, and J. Ronnelid. 2008. Antibodies to citrullinated proteins in arthritis: Pathology and promise. *Current Opinion in Rheumatology* 20, no. 3: 300-5.

Kolaskar, A. S. and P. C. Tongaonkar. 1990. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Letters* 276, no. 1-2: 172-4.

Liu, E. and G. S. Eisenbarth. 2007. Accepting clocks that tell time poorly: Fluid-phase versus standard elisa autoantibody assays. *Clinical Immunology* 125, no. 2: 120-6.

Mandava, S., L. Makowski, S. Devarapalli, J. Uzubell, and D. J. Rodi. 2004. Relic--a bioinformatics server for combinatorial peptide analysis and identification of protein-ligand interaction sites. *Proteomics* 4, no. 5: 1439-60.

Mei, Henrik E., Taketoshi Yoshida, Wondossen Sime, Falk Hiepe, Kathi Thiele, Rudolf A. Manz, Andreas Radbruch, and Thomas DÃ¶rner. 2009. Blood-borne human plasma cells in steady state are derived from mucosal immune responses. *Blood* 113, no. 11: 2461-2469.

Meloen, R. H., W. C. Puijk, and J. W. Slootstra. 2000. Mimotopes: Realization of an unlikely concept. *J Mol Recognit* 13, no. 6: 352-9.

Merbl, Y., R. Itzchak, T. Vider-Shalit, Y. Louzoun, F. J. Quintana, E. Vadai, L. Eisenbach, and I. R. Cohen. 2009. A systems immunology approach to the host-tumor interaction: Large-scale patterns of natural autoantibodies distinguish healthy and tumor-bearing mice. *PLoS ONE [Electronic Resource]* 4, no. 6: e6053.

Mouquet, Hugo, Johannes F. Scheid, Markus J. Zoller, Michelle Krogsgaard, Rene G. Ott, Shetha Shukair, Maxim N. Artyomov, John Pietzsch, Mark Connors, Florencia Pereyra, Bruce D. Walker, David D. Ho, Patrick C. Wilson, Michael S. Seaman, Herman N. Eisen, Arup K. Chakraborty, Thomas J. Hope, Jeffrey V. Ravetch, Hedda Wardemann, and Michel C. Nussenzweig. 2010. Polyreactivity increases the apparent affinity of anti-hiv antibodies by heteroligation. *Nature* 467, no. 7315: 591-595.

Parker, J. M., D. Guo, and R. S. Hodges. 1986. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: Correlation of predicted surface residues with antigenicity and x-ray-derived accessible sites. *Biochemistry* 25, no. 19: 5425-32.

Reineke, U., C. Ivascu, M. Schlief, C. Landgraf, S. Gericke, G. Zahn, H. Herzel, R. Volkmer-Engert, and J. Schneider-Mergener. 2002. Identification of distinct antibody epitopes and mimotopes from a peptide array of 5520 randomly generated sequences. *J Immunol Methods* 267, no. 1: 37-51.

Rodi, Diane J. and Lee Makowski. 1999. Phage-display technology â€“ finding a needle in a vast molecular haystack. *Current Opinion in Biotechnology* 10, no. 1: 87-93.

Shimomura, Y., E. Mizoguchi, K. Sugimoto, R. Kibe, Y. Benno, A. Mizoguchi, and A. K. Bhan. 2008. Regulatory role of b-1 b cells in chronic colitis. *International Immunology* 20, no. 6: 729-37.

Tapia, V., J. Bongartz, M. Schutkowski, N. Bruni, A. Weiser, B. Ay, R. Volkmer, and M. Or-Guil. 2007. Affinity profiling using the peptide microarray technology: A case study. *Analytical Biochemistry* 363, no. 1: 108-18.

Yung, Susan and Tak Mao Chan. 2008. Anti-DNA antibodies in the pathogenesis of lupus nephritis â€" the emerging mechanisms. *Autoimmunity Reviews* 7, no. 4: 317-321.

Zhang, J., A. M. Jacobi, J. Wang, R. Berlin, B. T. Volpe, and B. Diamond. 2009. Polyreactive autoantibodies in systemic lupus erythematosus have pathogenic potential. *J Autoimmun.* 33, no. 3-4: 270-4. Epub 2009 Apr 26.

Zhou, Z. H., A. G. Tzioufas, and A. L. Notkins. 2007. Properties and function of polyreactive antibodies and polyreactive antigen-binding b cells. *Journal of Autoimmunity* 29, no. 4: 219-28.

Zhou, Z. H., J. Zhang, Y. F. Hu, L. M. Wahl, J. O. Cisar, and A. L. Notkins. 2007. The broad antibacterial activity of the natural antibody repertoire is due to polyreactive antibodies. *Cell Host Microbe.* 1, no. 1: 51-61.

REFERENCES

Adams, D. W. and J. Errington. 2009. Bacterial cell division: Assembly, maintenance and disassembly of the z ring. *Nature Reviews. Microbiology* 7, no. 9: 642-53.

Adda, C. G., R. F. Anders, L. Tilley, and M. Foley. 2002. Random sequence libraries displayed on phage: Identification of biologically important molecules. *Combinatorial Chemistry and High Throughput Screening* 5, no. 1: 1-14.

Afonin, P. V., A. V. Fokin, I. N. Tsygannik, I. Y. Mikhailova, L. V. Onoprienko, Mikhaleva, II, V. T. Ivanov, T. Y. Mareeva, V. A. Nesmeyanov, N. Li, W. A. Pangborn, W. L. Duax, and V. Z. Pletnev. 2001. Crystal structure of an anti-interleukin-2 monoclonal antibody fab complexed with an antigenic nonapeptide. *Protein Science* 10, no. 8: 1514-21.

Allen, C. D., T. Okada, and J. G. Cyster. 2007. Germinal-center organization and cellular dynamics. *Immunity* 27, no. 2: 190-202.

Amanna, I. J. and M. K. Slifka. Contributions of humoral and cellular immunity to vaccine-induced protection in humans. 2011. *Virology* 411, no. 2: 206-15.

Amur, Shashi, Felix W. Frueh, Lawrence J. Lesko, and Shiew-Mei Huang. 2008. Integration and use of biomarkers in drug development, regulation and clinical practice: A us regulatory perspective. *Biomarkers in Medicine* 2, no. 3: 305-311.

Anderson, Karen S. and Joshua LaBaer. 2005. The sentinel within:Â€‰ exploiting the immune system for cancer biomarkersâ€ *Journal of Proteome Research* 4, no. 4: 1123-1133.

Arbuckle, M. R., M. T. McClain, M. V. Rubertone, R. H. Scofield, G. J. Dennis, J. A. James, and J. B. Harley. 2003a. Development of autoantibodies before the clinical onset of systemic lupus erythematosus. *N Engl J Med* 349, no. 16: 1526-33.

Astronomo, Rena D. and Dennis R. Burton. 2010. Carbohydrate vaccines: Developing sweet solutions to sticky situations? *Nat Rev Drug Discov* 9, no. 4: 308-324.

Bacarese-Hamilton, T., J. Gray, and A. Crisanti. 2003. Protein microarray technology for unraveling the antibody specificity repertoire against microbial proteomes. *Curr Opin Mol Ther* 5, no. 3: 278-84.

Bailey, T. L., N. Williams, C. Misleh, and W. W. Li. 2006. Meme: Discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research* 34, no. Web Server issue: W369-73.

Bastas, G., S. R. Sompuram, B. Pierce, K. Vani, and S. A. Bogen. 2008a. Bioinformatic requirements for protein database searching using predicted epitopes from disease-associated antibodies. *Mol Cell Proteomics* 7, no. 2: 247-56.

Blank, M., O. Barzilai, and Y. Shoenfeld. 2007. Molecular mimicry and auto-immunity. *Clinical Reviews in Allergy & Immunology* 32, no. 1: 111-8.

Blythe, M. J. and D. R. Flower. 2005. Benchmarking b cell epitope prediction: Underperformance of existing methods. *Protein Science* 14, no. 1: 246-8.

Boltz, K. W., M. J. Gonzalez-Moa, P. Stafford, S. A. Johnston, and S. A. Svarovsky. 2009. Peptide microarrays for carbohydrate recognition. *Analyst* 134, no. 4: 650-2.

Bongartz, J., N. Bruni, and M. Or-Guil. 2009. Epitope mapping using randomly generated peptide libraries. *Methods Mol Biol* 524: 237-46.

Bonifacio, E, V Lampasona, L Bernasconi, and AG Ziegler. 2000. Maturation of the humoral autoimmune response to epitopes of gad in preclinical childhood type 1 diabetes. *Diabetes* 49, no. 2: 202-208.

Bowman, Christal C. and John D. Clements. 2001. Differential biological and adjuvant activities of cholera toxin and escherichia coli heat-labile enterotoxin hybrids. *Infect. Immun.* 69, no. 3: 1528-1535.

Braden, B. C. and R. J. Poljak. 1995. Structural features of the reactions between antibodies and protein antigens. *FASEB J* 9, no. 1: 9-16.

Breitling, F., T. Felgenhauer, A. Nesterov, V. Lindenstruth, V. Stadler, and F. R. Bischoff. 2009. Particle-based synthesis of peptide arrays. *Chembiochem* 10, no. 5: 803-8.

Breitling, F. and M. Little. 1986. Carboxy-terminal regions on the surface of tubulin and microtubules. Epitope locations of yol1/34, dm1a and dm1b. *Journal of Molecular Biology* 189, no. 2: 367-70.

Breitling, Frank, Alexander Nesterov, Volker Stadler, Thomas Felgenhauer, and F. Ralf Bischoff. 2009. High-density peptide arrays. *Molecular BioSystems* 5, no. 3: 224-234.

Brettschneider, S., N. G. Morgenthaler, S. J. Teipel, C. Fischer-Schulz, K. Burger, R. Dodel, Y. Du, H. J. Moller, A. Bergmann, and H. Hampel. 2005. Decreased serum amyloid beta(1-42) autoantibody levels in alzheimer's disease, determined by a newly developed immuno-precipitation assay with radiolabeled amyloid beta(1-42) peptide. *Biol Psychiatry* 57, no. 7: 813-6.

Burbelo, P. D., K. E. Bren, K. H. Ching, E. S. Gogineni, S. Kottilil, J. I. Cohen, J. A. Kovacs, and M. J. Iadarola. 2011. Lips arrays for simultaneous detection of antibodies against partial and whole proteomes of hcv, hiv and ebv. *Mol Biosyst* 7, no. 5: 1453-62.

Burbelo, P. D., K. H. Ching, E. R. Bush, B. L. Han, and M. J. Iadarola. 2010. Antibody-profiling technologies for studying humoral responses to infectious agents. *Expert Rev Vaccines* 9, no. 6: 567-78.

Burbelo, P. D., K. H. Ching, T. L. Mattson, J. S. Light, L. R. Bishop, and J. A. Kovacs. 2007. Rapid antibody quantification and generation of whole proteome antibody response profiles using lips (luciferase immunoprecipitation systems). *Biochemical & Biophysical Research Communications* 352, no. 4: 889-95.

Burritt, James B., Frank R. DeLeo, Connie L. McDonald, Justin R. Prigge, Mary C. Dinauer, Michio Nakamura, William M. Nauseef, and Algirdas J. Jesaitis. 2001. Phage display epitope mapping of human neutrophil flavocytochromeb 558. *Journal of Biological Chemistry* 276, no. 3: 2053-2061.

Butler, J. E. 2000. Solid supports in enzyme-linked immunosorbent assay and other solid-phase immunoassays. *Methods (Duluth)* 22, no. 1: 4-23.

Caiazzo, R. J., Jr., D. J. O'Rourke, T. J. Barder, B. P. Nelson, and B. C. Liu. 2011. Native antigen fractionation protein microarrays for biomarker discovery. *Methods Mol Biol* 723: 129-48.

Cao, B. and C. Mao. 2009. Identification of microtubule-binding domains on microtubule-associated proteins by major coat phage display technique. *Biomacromolecules* 10, no. 3: 555-64.

Carettoni, D., P. Gomez-Puertas, L. Yim, J. Mingorance, O. Massidda, M. Vicente, A. Valencia, E. Domenici, and D. Anderluzzi. 2003. Phage-display and correlated mutations identify an essential region of subdomain 1c involved in homodimerization of escherichia coli ftsa. *Proteins* 50, no. 2: 192-206.

Carter, D. M., J. N. Gagnon, M. Damlaj, S. Mandava, L. Makowski, D. J. Rodi, P. D. Pawelek, and J. W. Coulton. 2006. Phage display reveals multiple contact sites between fhua, an outer membrane receptor of escherichia coli, and tonb. *Journal of Molecular Biology* 357, no. 1: 236-51.

Carter, J. M. and L. Loomis-Price. 2004. B cell epitope mapping using synthetic peptides. *Current Protocols in Immunology* Chapter 9: Unit 9 4.

Casey, J. L., A. M. Coley, K. Parisi, and M. Foley. 2009. Peptide mimics selected from immune sera using phage display technology can replace native antigens in the diagnosis of epstein-barr virus infection. *Protein Engineering, Design & Selection* 22, no. 2: 85-91.

Cenci, Simone and Roberto Sitia. 2007. Managing and exploiting stress in the antibody factory. *FEBS Letters* 581, no. 19: 2652-3657.

Chapman, C., A. Murray, J. Chakrabarti, A. Thorpe, C. Woolston, U. Sahin, A. Barnes, and J. Robertson. 2007. Autoantibodies in breast cancer: Their use as an aid to early diagnosis. *Ann Oncol* 18, no. 5: 868-73.

Chen, S. W., M. H. Van Regenmortel, and J. L. Pellequer. 2009. Structure-activity relationships in peptide-antibody complexes: Implications for epitope prediction and development of synthetic peptide vaccines. *Curr Med Chem* 16, no. 8: 953-64.

Chen, Z. J., F. Shimizu, J. Wheeler, and A. L. Notkins. 1996. Polyreactive antigen-binding b cells in the peripheral circulation are igd+ and b7. *European Journal of Immunology* 26, no. 12: 2916-23.

Chen, Z. J., C. J. Wheeler, W. Shi, A. J. Wu, C. H. Yarboro, M. Gallagher, and A. L. Notkins. 1998. Polyreactive antigen-binding b cells are the predominant cell type in the newborn b cell repertoire. *European Journal of Immunology* 28, no. 3: 989-94.

Chen, Z. J., J. Wheeler, and A. L. Notkins. 1995. Antigen-binding b cells and polyreactive antibodies. *European Journal of Immunology* 25, no. 2: 579-86.

Cho-Chung, Y. S. 2006. Autoantibody biomarkers in the detection of cancer. *Biochim Biophys Acta* 1762, no. 6: 587-91.

Chou, P. Y. and G. D. Fasman. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Advances in Enzymology & Related Areas of Molecular Biology* 47: 45-148.

Christen, U., E. Hintermann, M. Holdener, and M. G. von Herrath. 2010. Viral triggers for autoimmunity: Is the 'glass of molecular mimicry' half full or half empty? *Journal of Autoimmunity* 34, no. 1: 38-44.

Colasanti, T., C. Barbati, G. Rosano, W. Malorni, and E. Ortona. 2010. Autoantibodies in patients with alzheimer's disease: Pathogenetic role and potential use as biomarkers of disease progression. *Autoimmunity Reviews* 9, no. 12: 807-11.

Cooperman, Jonathan, Robert Neely, David T. Teachey, Stephen Grupp, and John Kim Choi. 2004. Cell division rates of primary human precursor b cells in culture reflect in vivo rates. *Stem Cells* 22: 1111-1120.

Cortese, R., F. Felici, G. Galfre, A. Luzzago, P. Monaci, and A. Nicosia. 1994. Epitope discovery using peptide libraries displayed on phage. *Trends in Biotechnology* 12, no. 7: 262-7.

Darnell, Robert B. and Jerome B. Posner. 2003. Paraneoplastic syndromes involving the nervous system. *New England Journal of Medicine* 349, no. 16: 1543-1554.

Davies, J. M. 1997. Molecular mimicry: Can epitope mimicry induce autoimmune disease? *Immunology & Cell Biology* 75, no. 2: 113-26.

Denisova, G. F., D. A. Denisov, and J. L. Bramson. 2010. Applying bioinformatics for antibody epitope prediction using affinity-selected mimotopes - relevance for vaccine design. *Immunome Res* 6 Suppl 2: S6.

Derda, R., S. K. Tang, S. C. Li, S. Ng, W. Matochko, and M. R. Jafari. 2011. Diversity of phage-displayed libraries of peptides during panning and amplification. *Molecules* 16, no. 2: 1776-803.

Dimitrov, Jordan D., Cyril Planchais, Jonghoon Kang, Anastas Pashov, Tchavdar L. Vassilev, Srinivas V. Kaveri, and Sebastien Lacroix-Desmazes. 2010. Heterogeneous antigen recognition behavior of induced polyspecific antibodies. *Biochemical and Biophysical Research Communications* 398, no. 2: 266-271.

Dotan, N., R. T. Altstock, M. Schwarz, and A. Dukler. 2006. Anti-glycan antibodies as biomarkers for diagnosis and prognosis. *Lupus* 15, no. 7: 442-450.

Dudak, F. C., I. H. Boyaci, and B. P. Orner. 2011. The discovery of small-molecule mimicking peptides through phage display. *Molecules* 16, no. 1: 774-89.

Eggert, M., U. K. Zettl, and G. Neeck.2010. Autoantibodies in autoimmune diseases. *Current Pharmaceutical Design* 16, no. 14: 1634-43.

El-Manzalawy, Y. and V. Honavar. 2010. Recent advances in b-cell epitope prediction methods. *Immunome Res* 6 Suppl 2: S2.

Elrington, G. M., N. M. Murray, S. G. Spiro, and J. Newson-Davis. 1991. Neurological paraneoplastic syndromes in patients with small cell ling cancer: A prospective survey of 150 patients. *Journal of Neurosurgery and Psychiatry* 54, no. 9: 764-767.

Emini, E. A., J. V. Hughes, D. S. Perlow, and J. Boger. 1985. Induction of hepatitis a virus-neutralizing antibody by a virus-specific synthetic peptide. *Journal of Virology* 55, no. 3: 836-9.

Erkkila, A. T., O. Narvanen, S. Lehto, M. I. Uusitupa, and S. Yla-Herttuala. 2000. Autoantibodies against oxidized low-density lipoprotein and cardiolipin in patients with coronary heart disease. *Arterioscler Thromb Vasc Biol* 20, no. 1: 204-9.

Eshaghi, M., W. S. Tan, and K. Yusoff. 2005. Identification of epitopes in the nucleocapsid protein of nipah virus using a linear phage-displayed random peptide library. *Journal of Medical Virology* 75, no. 1: 147-52.

Fack, F., B. Hugle-Dorr, D. Song, I. Queitsch, G. Petersen, and E. K. Bautz. 1997. Epitope mapping by phage display: Random versus gene-fragment libraries. *J Immunol Methods* 206, no. 1-2: 43-52.

Fang, Y., A. G. Frutos, and J. Lahiri. 2002. Membrane protein microarrays. *Journal of the American Chemical Society* 124, no. 11: 2394-5.

Fauchere, J. L., M. Charton, L. B. Kier, A. Verloop, and V. Pliska. 1988. Amino acid side chain parameters for correlation studies in biology and pharmacology. *International Journal of Peptide & Protein Research* 32, no. 4: 269-78.

FDA. 2010. Table of valid genomic biomarkers in the context of approved drug labels. http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm accessed Date Accessed)|.

Fodor, S. P., J. L. Read, M. C. Pirrung, L. Stryer, A. T. Lu, and D. Solas. 1991. Light-directed, spatially addressable parallel chemical synthesis. *Science* 251, no. 4995: 767-73.

Forster, Irmgard and Klaus Rajewsky. 1990. The bulk of the peripheral b-cell pool in mice is stable and not rapidly renewed from the bone marrow. *Proceedings of the National Academy of Sciences of the United States of America* 87: 4781-4784.

Frith, M. C., N. F. Saunders, B. Kobe, and T. L. Bailey. 2008. Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Computational Biology* 4, no. 4: e1000071.

Fu, M., P. S. Fan, W. Li, C. X. Li, Y. Xing, J. G. An, G. Wang, X. L. Fan, T. W. Gao, Y. F. Liu, and S. Ikeda. 2007. Identification of poly-reactive natural igm antibody that recognizes late apoptotic cells and promotes phagocytosis of the cells. *Apoptosis* 12, no. 2: 355-62.

Fujinami, R. S., M. B. Oldstone, Z. Wroblewska, M. E. Frankel, and H. Koprowski. 1983. Molecular mimicry in virus infection: Crossreaction of measles virus phosphoprotein or of herpes simplex virus protein with human intermediate filaments. *Proceedings of the National Academy of Sciences of the United States of America* 80, no. 8: 2346-50.

Gao, X., J. P. Pellois, Y. Na, Y. Kim, E. Gulari, and X. Zhou. 2004. High density peptide microarrays. In situ synthesis and applications. *Molecular Diversity* 8, no. 3: 177-87.

Garcia de la Nava, J., S. van Hijum, and O. Trelles. 2004. Saturation and quantization reduction in microarray experiments using two scans at different sensitivities. *Stat Appl Genet Mol Biol* 3, no. 8: Article11.

Geijersstam, VeronikaÂ af, Mari Kibur, Zhaohui Wang, Pentti Koskela, EeroÂ Pukka la, John Schiller, Matti Lehtinen, and Joakim Dillner. 1998. Stability over time of serum antibody levels to human papillomavirus type 16. *The Journal of Infectious Diseases* 177, no. 6: 1710-1714.

Gershoni, J. M., A. Roitburd-Berman, D. D. Siman-Tov, N. Tarnovitski Freund, and Y. Weiss. 2007. Epitope mapping: The first step in developing epitope-based vaccines. *BioDrugs* 21, no. 3: 145-56.

Giraudi, G., I. Rosso, C. Baggiani, and C. Giovannoli. 1999. Affinity between immobilised monoclonal and polyclonal antibodies and steroid-enzyme tracers increases sharply at high surface density. *Analytica Chimica Acta* 381, no. 2-3: 133-146.

Greving, M. P., P. E. Belcher, C. D. Cox, D. Daniel, C. W. Diehnelt, and N. W. Woodbury. 2010. High-throughput screening in two dimensions: Binding intensity and off-rate on a peptide microarray. *Analytical Biochemistry* 402, no. 1: 93-5.

Gultekin, S. Humayun, Myrna R. Rosenfeld, Raymond Voltz, Joseph Eichen, Jerome B. Posner, and Josep Dalmau. 2000. Paraneoplastic limbic encephalitis: Neurological symptoms, immunological findings and tumour association in 50 patients. *Brain* 123, no. 7: 1481-1494.

Halperin, R. F., P. Stafford, and S. A. Johnston. 2011. Exploring antibody recognition of sequence space through random-sequence peptide microarrays. *Molecular & Cellular Proteomics* 10, no. 3: M110 000786.

Hampton, Tracy. 2003. Autoantibodies predict lupus. *Journal of the American Medical Association* 290: 3186.

Hanash, Sam. 2003. Disease proteomics. *Nature* 422, no. 6928: 226-232.

Hansson, G. K. and A. Hermansson. 2011. The immune system in atherosclerosis. *Nature Immunology* 12, no. 3: 204-12.

Hao, Zhenyue and Klaus Rajewsky. 2001. Homeostasis of peripheral b cells in the absence of b cell influx from the bone marrow. *The Journal of Experimental Medicine* 194, no. 8: 1151-1164.

Hoess, R. H., A. J. Mack, H. Walton, and T. M. Reilly. 1994. Identification of a structural epitope by using a peptide library displayed on filamentous bacteriophage. *Journal of Immunology* 153, no. 2: 724-9.

Irving, M. B., O. Pan, and J. K. Scott. 2001. Random-peptide libraries and antigen-fragment libraries for epitope mapping and the development of vaccines and diagnostics. *Current Opinion in Chemical Biology* 5, no. 3: 314-24.

James, L. C., P. Roversi, and D. S. Tawfik. 2003. Antibody multispecificity mediated by conformational diversity. *Science* 299, no. 5611: 1362-7.

Jonassen, I. 1997. Efficient discovery of conserved patterns using a pattern graph. *Computer Applications in the Biosciences* 13, no. 5: 509-22.

Kajihara, Y., N. Yamamoto, R. Okamoto, K. Hirano, and T. Murase. Chemical synthesis of homogeneous glycopeptides and glycoproteins. 2010. *Chemical Record: an Official Publication of the Chemical Society of Japan* 10, no. 2: 80-100.

Kalaycioglu, A. T., P. H. Russell, and C. R. Howard. Peptide mimics of hapten dnp: The effect of affinity of anti-dnp monoclonal antibodies for the selection of phage-displayed mimotopes. *Journal of Immunological Methods* 366, no. 1-2: 36-42.

Karplus, P. A. and G. E. Schulz. 1985. Prediction of chain flexibility in proteins. *Naturwissenschaften* 72, no. 4: 212-213.

Katz, Bradley A. 1997. Structural and mechanistic determinants of affinity and specificity of ligands discovered or engineered by phage display. *Annual Review of Biophysics and Biomolecular Structure* 26, no. 1: 27-45.

Kaufman, E. N. and R. K. Jain. 1992. Effect of bivalent interaction upon apparent antibody affinity: Experimental confirmation of theory using fluorescence photobleaching and implications for antibody binding assays. *Cancer Res* 52, no. 15: 4157-67.

Kawashima, S., P. Pokarowski, M. Pokarowska, A. Kolinski, T. Katayama, and M. Kanehisa. 2008. Aaindex: Amino acid index database, progress report 2008. *Nucleic Acids Research* 36, no. Database issue: D202-5.

Kieber-Emmons, T. 1998. Peptide mimotopes of carbohydrate antigens. *Immunologic Research* 17, no. 1-2: 95-108.

Kijanka, G., S. Ipcho, S. Baars, H. Chen, K. Hadley, A. Beveridge, E. Gould, and D. Murphy. 2009. Rapid characterization of binding specificity and cross-reactivity of antibodies using recombinant human protein arrays. *J Immunol Methods* 340, no. 2: 132-7.

Klareskog, L., M. Widhe, M. Hermansson, and J. Ronnelid. 2008. Antibodies to citrullinated proteins in arthritis: Pathology and promise. *Current Opinion in Rheumatology* 20, no. 3: 300-5.

Koide, S. and S. S. Sidhu. 2009. The importance of being tyrosine: Lessons in molecular recognition from minimalist synthetic binding proteins. *ACS Chemical Biology [Electronic Resource]* 4, no. 5: 325-34.

Kolaskar, A. S. and P. C. Tongaonkar. 1990. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Letters* 276, no. 1-2: 172-4.

Kracun, S. K., E. Clo, H. Clausen, S. B. Levery, K. J. Jensen, and O. Blixt. 2010. Random glycopeptide bead libraries for seromic biomarker discovery. *J Proteome Res* 9, no. 12: 6705-14.

Kraft, S., B. Diefenbach, R. Mehta, A. Jonczyk, G. A. Luckenbach, and S. L. Goodman. 1999. Definition of an unexpected ligand recognition motif for alphav beta6 integrin. *Journal of Biological Chemistry* 274, no. 4: 1979-85.

Kurian, Sunil M., Raymond Heilman, Tony S. Mondala, Aleksey Nakorchevsky, Johannes A. Hewel, Daniel Campbell, Elizabeth H. Robison, Lin Wang, Wen Lin, Lillian Gaber, Kim Solez, Hamid Shidban, Robert Mendez, Randolph L. Schaffer, Jonathan S. Fisher, Stuart M. Flechner, Steve R. Head, Steve Horvath, John R. Yates, III, Christopher L. Marsh, and Daniel R. Salomon. 2009. Biomarkers for early and late stage chronic allograft nephropathy by proteogenomic profiling of peripheral blood. *PLoS ONE* 4, no. 7: e6212.

Kurth, Dirk G. and Thomas Bein. 1993. Surface reactions on thin layers of silane coupling agents. *Langmuir* 9: 2965-2973.

Lapedes, A. and R. Farber. 2001. The geometry of shape space: Application to influenza. *Journal of Theoretical Biology* 212, no. 1: 57-69.

Larman, H. B., Z. Zhao, U. Laserson, M. Z. Li, A. Ciccia, M. A. Gakidis, G. M. Church, S. Kesari, E. M. Leproust, N. L. Solimini, and S. J. Elledge. 2011. Autoantigen discovery with a synthetic human peptidome. *Nat Biotechnol* 29, no. 6: 535-41.

Legutki, J. B., D. M. Magee, P. Stafford, and S. A. Johnston. 2010. A general method for characterization of humoral immunity induced by a vaccine or infection. *Vaccine* 28, no. 28: 4529-37.

Leslie, R. D. G., M. A. Atkinson, and Abner L. Notkins. 1999. Autoantigens ia-2 and gad in type i (insulin-dependent) diabetes. *Diabetologia* 42: 3-14.

Liang, S., D. Zheng, D. M. Standley, B. Yao, M. Zacharias, and C. Zhang. 2010. Epsvr and epmeta: Prediction of antigenic epitopes using support vector regression and multiple server results. *BMC Bioinformatics* 11: 381.

Litvack, M. L., P. Djiadeu, S. D. Renganathan, S. Sy, M. Post, and N. Palaniyar. 2010. Natural igm and innate immune collectin sp-d bind to late apoptotic cells and enhance their clearance by alveolar macrophages in vivo. *Molecular Immunology* 48, no. 1-3: 37-47.

Liu, E. and G. S. Eisenbarth. 2007. Accepting clocks that tell time poorly: Fluid-phase versus standard elisa autoantibody assays. *Clinical Immunology* 125, no. 2: 120-6.

Madi, A., I. Hecht, S. Bransburg-Zabary, Y. Merbl, A. Pick, M. Zucker-Toledano, F. J. Quintana, A. I. Tauber, I. R. Cohen, and E. Ben-Jacob. 2009. Organization of the autoantibody repertoire in healthy newborns and adults revealed by system level informatics of antigen microarray data. *Proceedings of the National Academy of Sciences of the United States of America* 106, no. 34: 14484-9.

Mandava, S., L. Makowski, S. Devarapalli, J. Uzubell, and D. J. Rodi. 2004. Relic--a bioinformatics server for combinatorial peptide analysis and identification of protein-ligand interaction sites. *Proteomics* 4, no. 5: 1439-60.

Manivel, V., N. C. Sahoo, D. M. Salunke, and K. V. Rao. 2000. Maturation of an antibody response is governed by modulations in flexibility of the antigen-combining site. *Immunity* 13, no. 5: 611-20.

Marchalonis, J. J., M. K. Adelman, I. F. Robey, S. F. Schluter, and A. B. Edmundson. 2001. Exquisite specificity and peptide epitope recognition promiscuity, properties shared by antibodies from sharks to humans. *Journal of Molecular Recognition* 14, no. 2: 110-21.

Mariuzza, R. A. 2006. Multiple paths to multispecificity. *Immunity* 24, no. 4: 359-61.

Mattoon, D., G. Michaud, J. Merkel, and B. Schweitzer. 2005. Biomarker discovery using protein microarray technology platforms: Antibody-antigen complex profiling. *Expert Rev Proteomics* 2, no. 6: 879-89.

Meens, J., V. Bolotin, R. Frank, J. Bohmer, and G. F. Gerlach. 2010. Characterization of a highly immunogenic mycoplasma hyopneumoniae lipoprotein mhp366 identified by peptide-spot array. *Veterinary Microbiology* 142, no. 3-4: 293-302.

Mei, Henrik E., Taketoshi Yoshida, Wondossen Sime, Falk Hiepe, Kathi Thiele, Rudolf A. Manz, Andreas Radbruch, and Thomas DÃ¶rner. 2009. Blood-borne human plasma cells in steady state are derived from mucosal immune responses. *Blood* 113, no. 11: 2461-2469.

Meloen, R. H., W. C. Puijk, and J. W. Slootstra. 2000. Mimotopes: Realization of an unlikely concept. *J Mol Recognit* 13, no. 6: 352-9.

Merbl, Y., R. Itzchak, T. Vider-Shalit, Y. Louzoun, F. J. Quintana, E. Vadai, L. Eisenbach, and I. R. Cohen. 2009a. A systems immunology approach to the host-tumor interaction: Large-scale patterns of natural autoantibodies distinguish healthy and tumor-bearing mice. *PLoS ONE [Electronic Resource]* 4, no. 6: e6053.

Merbl, Y., M. Zucker-Toledano, F. J. Quintana, and I. R. Cohen. 2007. Newborn humans manifest autoantibodies to defined self molecules detected by antigen microarray informatics. *Journal of Clinical Investigation* 117, no. 3: 712-8.

Michaud, G. A., M. Salcius, F. Zhou, R. Bangham, J. Bonin, H. Guo, M. Snyder, P. F. Predki, and B. I. Schweitzer. 2003. Analyzing antibody specificity with whole proteome microarrays. *Nat Biotechnol* 21, no. 12: 1509-12.

Mohan, S., K. Kourentzi, K. A. Schick, C. Uehara, C. A. Lipschultz, M. Acchione, M. E. Desantis, S. J. Smith-Gill, and R. C. Willson. 2009. Association energetics of cross-reactive and specific antibodies. *Biochemistry* 48, no. 6: 1390-8.

Morales Betanzos, C., M. J. Gonzalez-Moa, K. W. Boltz, B. D. Vander Werf, S. A. Johnston, and S. A. Svarovsky. 2009. Bacterial glycoprofiling by using random sequence peptide microarrays. *Chembiochem* 10, no. 5: 877-88.

Morell, A., W. D. Terry, and T. A. Waldmann. 1970. Metabolic properties of igg subclasses in man. *Journal of Clinical Investigation* 49, no. 4: 673-80.

Mouquet, H., J. F. Scheid, M. J. Zoller, M. Krogsgaard, R. G. Ott, S. Shukair, M. N. Artyomov, J. Pietzsch, M. Connors, F. Pereyra, B. D. Walker, D. D. Ho, D. B. Wilson, M. S. Seaman, H. N. Eisen, A. K. Chakraborty, T. J. Hope, J. V. Ravetch, H. Wardemann, and M. C. Nussenzweig. 2010. Polyreactivity increases the apparent affinity of anti-hiv antibodies by heteroligation. *Nature.* 467, no. 7315: 591-5.

Mun, J., Y. H. Kim, J. Yu, J. Bae, D. Y. Noh, M. H. Yu, and C. Lee. 2010. A proteomic approach based on multiple parallel separation for the unambiguous identification of an antibody cognate antigen. *Electrophoresis* 31, no. 20: 3428-36.

Nell, V. P., K. P. Machold, T. A. Stamm, G. Eberl, H. Heinzl, M. Uffmann, J. S. Smolen, and G. Steiner. 2005. Autoantibody profiling as early diagnostic and prognostic tool for rheumatoid arthritis. *Ann Rheum Dis* 64, no. 12: 1731-6.

Nie, J., B. Chang, D. O. Traktuev, J. Sun, K. March, L. Chan, E. H. Sage, R. Pasqualini, W. Arap, and M. G. Kolonin. 2008. Ifats collection: Combinatorial peptides identify alpha5beta1 integrin as a receptor for the matricellular protein sparc on adipose stromal cells. *Stem Cells* 26, no. 10: 2735-45.

Notkins, A. L. 2004a. Polyreactivity of antibody molecules. *Trends in Immunology* 25, no. 4: 174-9.

Notkins, Abner Louis. 2004b. Polyreactivity of antibody molecules. *Trends in immunology* 25, no. 4: 174-179.

Oldstone, M. B. 1998. Molecular mimicry and immune-mediated diseases. *FASEB Journal* 12, no. 13: 1255-65.

Papini, A. M. 2009. The use of post-translationally modified peptides for detection of biomarkers of immune-mediated diseases. *J Pept Sci* 15, no. 10: 621-8.

Park, Joon Won. 2007. Nsb amine slides. no. http://www.nsbpostech.com/2007/order/order3_detail.html?g_no=34 accessed Date Accessed)|.

Parker, J. M., D. Guo, and R. S. Hodges. 1986. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: Correlation of predicted surface residues with antigenicity and x-ray-derived accessible sites. *Biochemistry* 25, no. 19: 5425-32.

Pasqualini, Renata and Erkki Ruosiahti. 1996. Organ targeting *in vivo* using phage display peptide libraries. *Nature* 380: 364-366.

Pellequer, J. L., E. Westhof, and M. H. Van Regenmortel. 1993. Correlation between the location of antigenic sites and the prediction of turns in proteins. *Immunology Letters* 36, no. 1: 83-99.

Pinilla, C., R. Martin, B. Gran, J. R. Appel, C. Boggiano, D. B. Wilson, and R. A. Houghten. 1999. Exploring immunological specificity using synthetic peptide combinatorial libraries. *Curr Opin Immunol.* 11, no. 2: 193-202.

Quintana, F. J., P. H. Hagedorn, G. Elizur, Y. Merbl, E. Domany, and I. R. Cohen. 2004. Functional immunomics: Microarray analysis of igg autoantibody repertoires predicts the future response of mice to induced diabetes. *Proceedings of the National Academy of Sciences of the United States of America* 101 Suppl 2: 14615-21.

Radbruch, A., G. Muehlinghaus, E. O. Luger, A. Inamine, K. G. Smith, T. Dorner, and F. Hiepe. 2006. Competence and competition: The challenge of becoming a long-lived plasma cell. *Nature Reviews. Immunology* 6, no. 10: 741-50.

Ramachandran, N., J. V. Raphael, E. Hainsworth, G. Demirkan, M. G. Fuentes, A. Rolfs, Y. Hu, and J. LaBaer. 2008. Next-generation high-density self-assembling functional protein arrays. *Nature Methods* 5, no. 6: 535-8.

Reimer, U., U. Reineke, and J. Schneider-Mergener. 2002. Peptide arrays: From macro to micro. *Curr Opin Biotechnol* 13, no. 4: 315-20.

Reineke, U. 2009. Antibody epitope mapping using de novo generated synthetic peptide libraries. *Methods in Molecular Biology* 524: 203-11.

Reineke, U., C. Ivascu, M. Schlief, C. Landgraf, S. Gericke, G. Zahn, H. Herzel, R. Volkmer-Engert, and J. Schneider-Mergener. 2002. Identification of distinct antibody epitopes and mimotopes from a peptide array of 5520 randomly generated sequences. *J Immunol Methods* 267, no. 1: 37-51.

Restrepo, Lucas, Phillip Stafford, D. Mitch Magee, and Stephen Albert Johnston. 2011. Application of immunosignatures to the assessment of alzheimer's disease. *Annals of Neurology* 70, no. 2: 286-295.

Reuschenbach, M., M. von Knebel Doeberitz, and N. Wentzensen. 2009. A systematic review of humoral immune responses against tumor antigens. *Cancer Immunology, Immunotherapy* 58, no. 10: 1535-44.

Rigoutsos, I. and A. Floratos. 1998. Combinatorial pattern discovery in biological sequences: The teiresias algorithm. *Bioinformatics* 14, no. 1: 55-67.

Robinson, W. H., C. DiGennaro, W. Hueber, B. B. Haab, M. Kamachi, E. J. Dean, S. Fournel, D. Fong, M. C. Genovese, H. E. de Vegvar, K. Skriner, D. L. Hirschberg, R. I. Morris, S. Muller, G. J. Pruijn, W. J. van Venrooij, J. S. Smolen, P. O. Brown, L. Steinman, and P. J. Utz. 2002. Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nat Med* 8, no. 3: 295-301.

Roche, Stéphane, Yves Dauvilliers, Laurent Tiers, Carine Couderc, Marie-Thérèse Piva, Monique Provansal, Audrey Gabelle, and Sylvain Lehmann. 2008. Autoantibody profiling on high-density protein microarrays for biomarker discovery in the cerebrospinal fluid. *Journal of Immunological Methods* 338, no. 1-2: 75-78.

Rodi, D. J., R. W. Janes, H. J. Sanganee, R. A. Holton, B. A. Wallace, and L. Makowski. 1999. Screening of a library of phage-displayed peptides identifies human bcl-2 as a taxol-binding protein. *Journal of Molecular Biology* 285, no. 1: 197-203.

Rodi, Diane J. and Lee Makowski. 1999. Phage-display technology â€“ finding a needle in a vast molecular haystack. *Current Opinion in Biotechnology* 10, no. 1: 87-93.

Schroeder, H. W., Jr. and L. Cavacini. Structure and function of immunoglobulins. *Journal of Allergy & Clinical Immunology* 125, no. 2 Suppl 2: S41-52.

Scofield, R. Hal. 2004. Autoantibodies as predictors of disease. *Lancet* 363: 1544-1546.

Sfriso, P., A. Ghirardello, C. Botsios, M. Tonon, M. Zen, N. Bassi, F. Bassetto, and A. Doria. 2010. Infections and autoimmunity: The multifaceted relationship. *Journal of Leukocyte Biology* 87, no. 3: 385-95.

Shimomura, Y., E. Mizoguchi, K. Sugimoto, R. Kibe, Y. Benno, A. Mizoguchi, and A. K. Bhan. 2008. Regulatory role of b-1 b cells in chronic colitis. *International Immunology* 20, no. 6: 729-37.

Silverman, Gregg J. 2011. Regulatory natural autoantibodies to apoptotic cells: Pallbearers and protectors. *Arthritis & Rheumatism* 63, no. 3: 597-602.

Smith, D. J., A. S. Lapedes, J. C. de Jong, T. M. Bestebroer, G. F. Rimmelzwaan, A. D. Osterhaus, and R. A. Fouchier. 2004. Mapping the antigenic and genetic evolution of influenza virus. *Science* 305, no. 5682: 371-6.

Smith, K. G., T. D. Hewitson, G. J. Nossal, and D. M. Tarlinton. 1996. The phenotype and fate of the antibody-forming cells of the splenic foci. *European Journal of Immunology* 26, no. 2: 444-8.

Sompuram, S. R., G. Bastas, K. Vani, and S. A. Bogen. 2008. Accurate identification of paraprotein antigen targets by epitope reconstruction. *Blood* 111, no. 1: 302-8.

Srinivasappa, J., J. Saegusa, B. S. Prabhakar, M. K. Gentry, M. J. Buchmeier, T. J. Wiktor, H. Koprowski, M. B. Oldstone, and A. L. Notkins. 1986. Molecular mimicry: Frequency of reactivity of monoclonal antiviral antibodies with normal tissues. *Journal of Virology* 57, no. 1: 397-401.

Stafford, Phillip and Marcel Brun. 2007. Three methods for optimization of cross-laboratory and cross-platform microarray expression data. *Nucleic Acids Research* 35, no. 10: e72.

Stemke-Hale, Katherine, Bernhard Kaltenboeck, Fred J. DeGraves, Kathryn F. Sykes, Jin Huang, Chun-hui Bu, and Stephen Albert Johnston. 2005. Screening the whole genome of a pathogen in vivo for individual protective antigens. *Vaccine* 23, no. 23: 3016-3025.

Stephen, C. W., P. Helminen, and D. P. Lane. 1995. Characterisation of epitopes on human p53 using phage-displayed peptide libraries: Insights into antibody-peptide interactions. *Journal of Molecular Biology* 248, no. 1: 58-78.

Stephen, C. W. and D. P. Lane. 1992. Mutant conformation of p53. Precise epitope mapping using a filamentous phage epitope library. *Journal of Molecular Biology* 225, no. 3: 577-83.

Stockert, Elisabeth, Elke Jäger, Yao-Tseng Chen, Matthew J. Scanlan, Ivan Gout, Julia Karbach, Michael Arand, Alexander Knuth, and Lloyd J. Old. 1998. A survey of the humoral immune response of cancer patients to a panel of human tumor antigens. *Journal of Experimental Medicine* 187, no. 8: 1349-1354.

Sulzer, Bernhard, J. Leo van Hemmen, Avidan U. Neumann, and Ulrich Behn. 1993. Memory in idiotypic networks due to competition between proliferation and differentiation. *Bulletin of Mathematical Biology* 55, no. 6: 1133-1182.

Sun, P., W. Chen, Y. Huang, H. Wang, Z. Ma, and Y. Lv. 2011. Epitope prediction based on random peptide library screening: Benchmark dataset and prediction tools evaluation. *Molecules* 16, no. 6: 4971-93.

Takakusagi, Y., K. Kuramochi, M. Takagi, T. Kusayanagi, D. Manita, H. Ozawa, K. Iwakiri, K. Takakusagi, Y. Miyano, A. Nakazaki, S. Kobayashi, F. Sugawara, and K. Sakaguchi. 2008. Efficient one-cycle affinity selection of binding proteins or peptides specific for a small-molecule using a t7 phage display pool. *Bioorganic & Medicinal Chemistry* 16, no. 22: 9837-46.

Tang, De-chu, Michael DeVit, and Stephen A. Johnston. 1992. Genetic immunization is a simple method for eliciting an immune response. *Nature* 356, no. 6365: 152-154.

Tapia, V., J. Bongartz, M. Schutkowski, N. Bruni, A. Weiser, B. Ay, R. Volkmer, and M. Or-Guil. 2007. Affinity profiling using the peptide microarray technology: A case study. *Analytical Biochemistry* 363, no. 1: 108-18.

Thorpe, I. F. and C. L. Brooks, 3rd. 2007. Molecular evolution of affinity and flexibility in the immune system. *Proceedings of the National Academy of Sciences of the United States of America* 104, no. 21: 8821-6.

Thurlings, RM, K Vos, DM Gerlag, and PP Tak. 2006. The humoral response in rheumatoid arthritis and the effect of b-cell depleting therapy. *Ned Tijdschr Geneeskd* 150, no. 30: 1657-1661.

Tsuchiya, N. and R. C. Williams, Jr. 1992. Molecular mimicry--hypothesis or reality? *Western Journal of Medicine* 157, no. 2: 133-8.

Ullman, Christopher G., Laura Frigotto, and R. Neil Cooley. 2011. In vitro methods for peptide display and their applications. *Briefings in Functional Genomics* 10, no. 3: 125-134.

Uttamchandani, M. and S. Q. Yao. 2008. Peptide microarrays: Next generation biochips for detection, diagnostics and high-throughput screening. *Curr Pharm Des* 14, no. 24: 2428-38.

Vigil, A., D. H. Davies, and P. L. Felgner. 2010. Defining the humoral immune response to infectious agents using high-density protein microarrays. *Future Microbiology* 5, no. 2: 241-51.

Vihinen, M., E. Torkkila, and P. Riikonen. 1994. Accuracy of protein flexibility predictions. *Proteins* 19, no. 2: 141-9.

Vlieghe, P., V. Lisowski, J. Martinez, and M. Khrestchatisky. 2010. Synthetic therapeutic peptides: Science and market. *Drug Discovery Today* 15, no. 1-2: 40-56.

Volkmer, R. 2009. Synthesis and application of peptide arrays: Quo vadis spot technology. *Chembiochem* 10, no. 9: 1431-42.

Vollmers, H. P. and S. Brandlein. 2007. Natural antibodies and cancer. *Journal of Autoimmunity* 29, no. 4: 295-302.

Voltz, Raymond. 2002. Paraneoplastic neurological syndromes: An update on diagnosis, pathogenesis, and therapy. *The Lancet Neurology* 1, no. 5: 294-305.

Walsh, G. 2010. Post-translational modifications of protein biopharmaceuticals. *Drug Discovery Today* 15, no. 17-18: 773-80.

Wang, L. F. and M. Yu. 2004. Epitope identification and discovery using phage display libraries: Applications in vaccine development and diagnostics. *Current Drug Targets* 5, no. 1: 1-15.

White, S. J., R. E. Simmonds, D. A. Lane, and A. H. Baker. 2005. Efficient isolation of peptide ligands for the endothelial cell protein c receptor (epcr) using candidate receptor phage display biopanning. *Peptides* 26, no. 7: 1264-9.

Whittingham, S., S. L. Byron, J. Tuomilehto, P. Z. Zimmet, M. A. Myers, G. Vidgren, M. J. Rowley, S. J. Feeney, P. Koskela, E. Tuomilehto-Wolf, and I. R. Mackay. 1997. Autoantibodies associated with presymptomatic insulin-dependent diabetes mellitus in women. *Diabet Med* 14, no. 8: 678-85.

Williams, B. A., C. W. Diehnelt, P. Belcher, M. Greving, N. W. Woodbury, S. A. Johnston, and J. C. Chaput. 2009. Creating protein affinity reagents by combining peptide ligands on synthetic DNA scaffolds. *Journal of the American Chemical Society* 131, no. 47: 17233-41.

Xu, Y., P. A. Ramsland, J. M. Davies, M. Scealy, K. S. Nandakumar, R. Holmdahl, and M. J. Rowley. 2004. Two monoclonal antibodies to precisely the same epitope of type ii collagen select non-crossreactive phage clones by phage display: Implications for autoimmunity and molecular mimicry. *Molecular Immunology* 41, no. 4: 411-9.

Yang, W. J., J. F. Lai, K. C. Peng, H. J. Chiang, C. N. Weng, and D. Shiuan. 2005. Epitope mapping of mycoplasma hyopneumoniae using phage displayed peptide libraries and the immune responses of the selected phagotopes. *Journal of Immunological Methods* 304, no. 1-2: 15-29.

Yip, Y. L. and R. L. Ward. 1999. Epitope discovery using monoclonal antibodies and phage peptide libraries. *Combinatorial Chemistry & High Throughput Screening* 2, no. 3: 125-38.

Yung, Susan and Tak Mao Chan. 2008. Anti-DNA antibodies in the pathogenesis of lupus nephritis â€" the emerging mechanisms. *Autoimmunity Reviews* 7, no. 4: 317-321.

Zhang, J., A. M. Jacobi, J. Wang, R. Berlin, B. T. Volpe, and B. Diamond. 2009. Polyreactive autoantibodies in systemic lupus erythematosus have pathogenic potential. *J Autoimmun.* 33, no. 3-4: 270-4. Epub 2009 Apr 26.

Zhang, Q., P. Wang, Y. Kim, P. Haste-Andersen, J. Beaver, P. E. Bourne, H. H. Bui, S. Buus, S. Frankild, J. Greenbaum, O. Lund, C. Lundegaard, M. Nielsen, J. Ponomarenko, A. Sette, Z. Zhu, and B. Peters. 2008. Immune epitope database analysis resource (iedb-ar). *Nucleic Acids Research* 36, no. Web Server issue: W513-8.

Zhang, Wen, Yi Xiong, Meng Zhao, Hua Zou, Xinghuo Ye, and Juan Liu. Prediction of conformational b-cell epitopes from 3d structures by random forest with a distance-based feature. *BMC Bioinformatics* 12, no. 1: 341.

Zhao, S. and E. Y. Lee. 1997. A protein phosphatase-1-binding motif identified by the panning of a random peptide display library. *Journal of Biological Chemistry* 272, no. 45: 28368-72.

Zhou, Z. H. , A. G. Tzioufas, and A. L. Notkins. 2007a. Properties and function of polyreactive antibodies and polyreactive antigen-binding b cells. *J Autoimmun.* 29, no. 4: 219-28. Epub 2007 Sep 20.

Zhou, Z. H., A. G. Tzioufas, and A. L. Notkins. 2007b. Properties and function of polyreactive antibodies and polyreactive antigen-binding b cells. *Journal of Autoimmunity* 29, no. 4: 219-28.

Zhou, Z. H., J. Zhang, Y. F. Hu, L. M. Wahl, J. O. Cisar, and A. L. Notkins. 2007a. The broad antibacterial activity of the natural antibody repertoire is due to polyreactive antibodies. *Cell Host Microbe.* 1, no. 1: 51-61.

Zhou, Zhao-Hua, Yahong Zhang, Ya-Fang Hu, Larry M. Wahl, John O. Cisar, and Abner Louis Notkins. 2007b. The broad antibacterial activity of the natural antibody repertoire is due to polyreactive antibodies. *Cell Host & Microbe* 1, no. 1: 51-61.

Zubair, A., P. D. Burbelo, L. G. Vincent, M. J. Iadarola, P. D. Smith, and N. Y. Morgan. 2011. Microfluidic lips for serum antibody detection: Demonstration of a rapid test for hsv-2 infection. *Biomed Microdevices.*

APPENDIX A

SUPPLEMENTAL FIGURES FOR CHAPTER 2

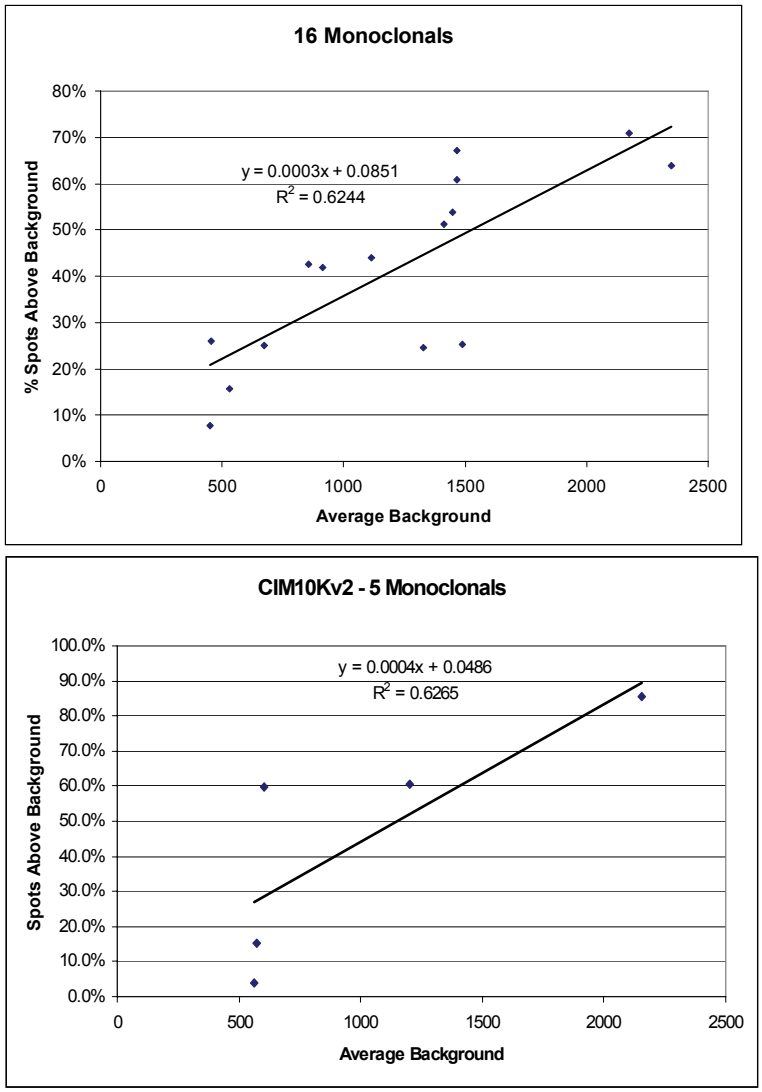


Figure A-1. Spots above background vs. Average Background. The average background binding to the slide surface is plotted against the number of spots that bind above background. The top graph shows 16 monoclonals run on the CIM10Kv1, and the bottom graph shows five monoclonals run on the 10Kv2. A line was fit to the data, and the equation of the line and the RSQ are shown on each graph.

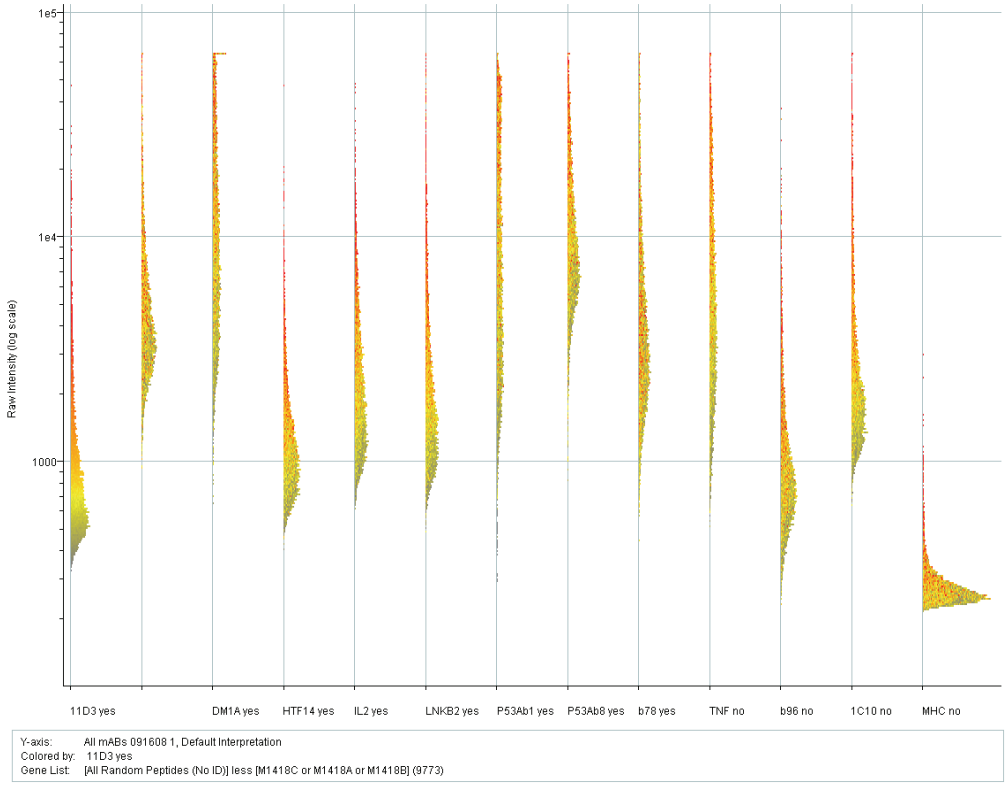
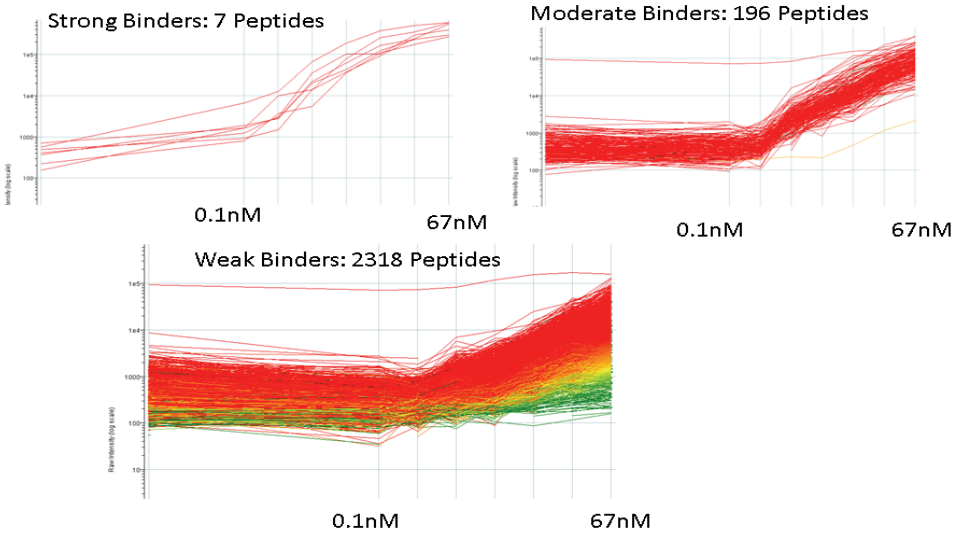


Figure A-2 Monoclonal Binding Distributions by Western Blot. Histograms of monoclonal binding distributions are organized by whether the antibody recognizes its target when it is denatured in a western blot. Only random sequence peptides are shown. Note that this is raw data, not secondary subtracted. *(include table)*



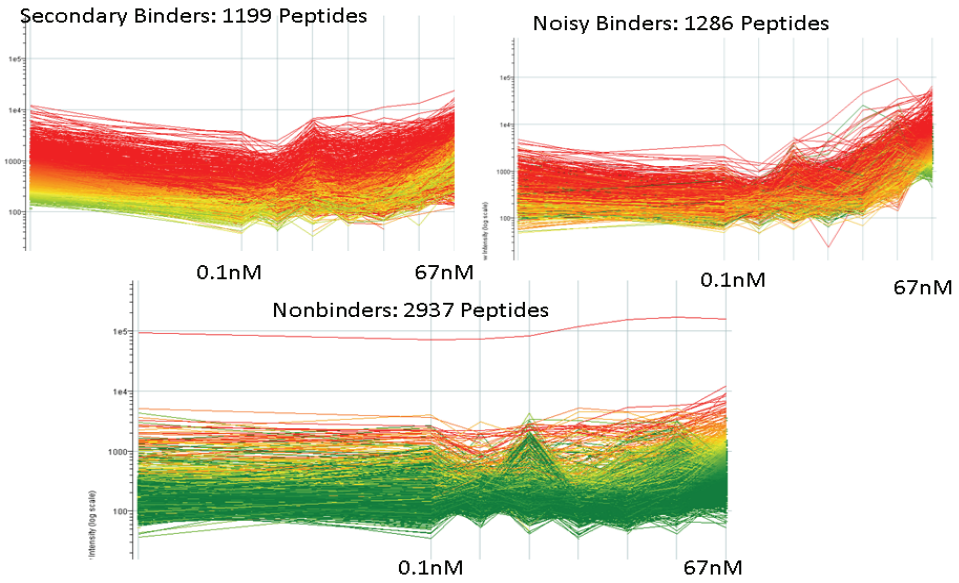


Figure A-3. Grouping of Peptides in Concentration Series. The concentration series of P53Ab1 was grouped by estimated half maximal binding (see Chapter 2, Figure 3A). The strong binders are <67nM, moderate are 66nM to 667nM and the weak binders are >667nM. Here the half maximal binding was estimating using constant rmax (signal at saturation), while the histogram in chapter 2 is based on estimates where rmax was also fit.

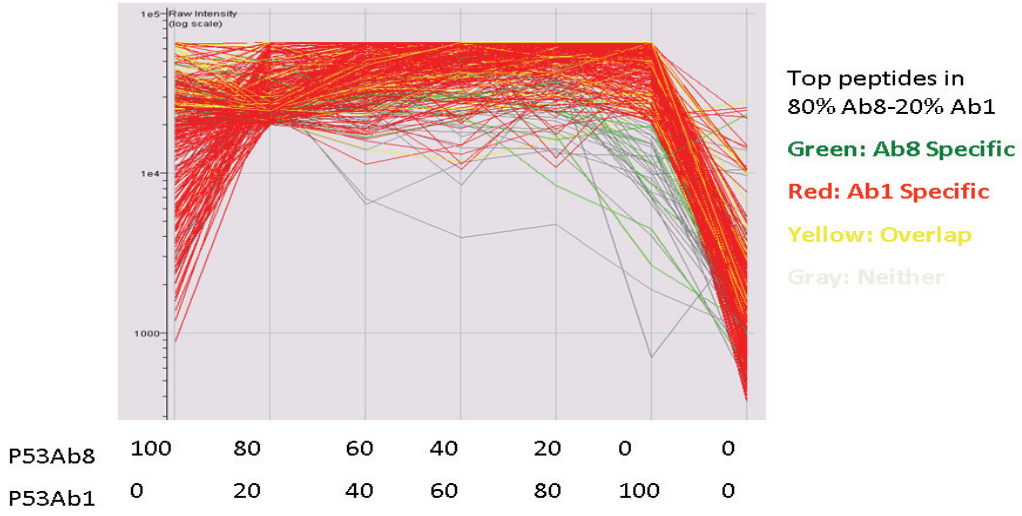


Figure A-4. Two Monoclonal Mixing. The top peptides in the 80% P5Ab8 20% P5Ab1 experiment are shown in the line graph above. They are colored by which antibody they are specific for.

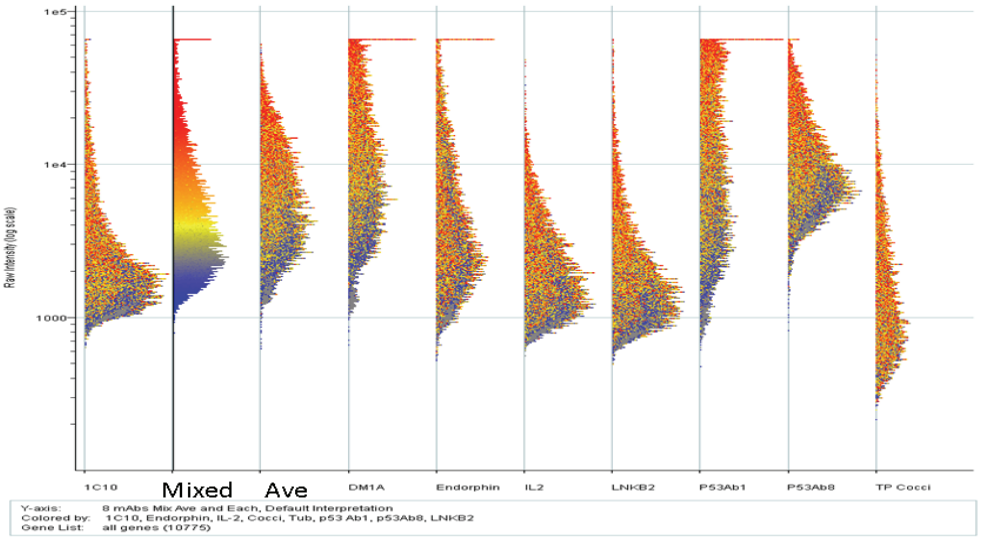


Figure A-5. Eight Monoclonal Mixing. Histograms of the binding distributions of the eight monoclonals used in the mixing experiment are shown, along with the mixed signal distribution and the average of the eight.

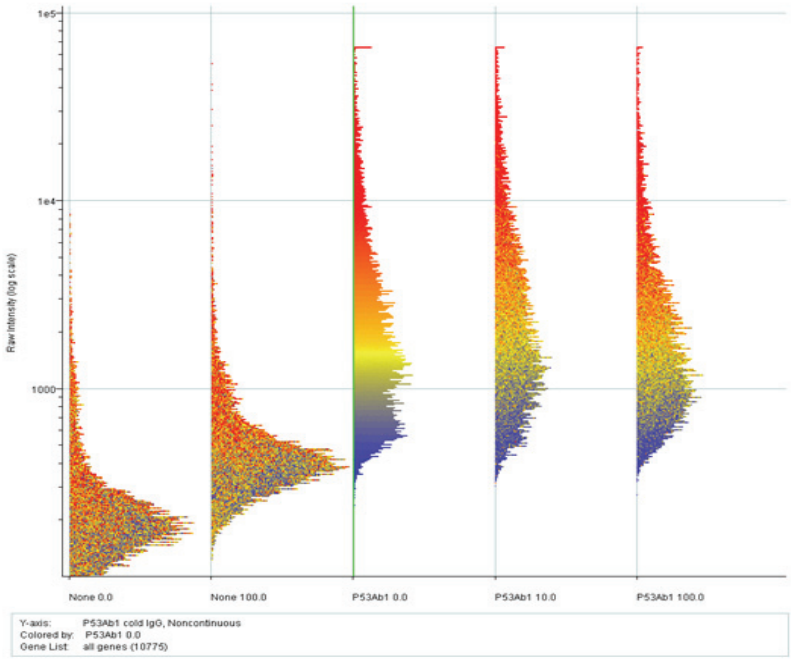


Figure A-6. Competition with Excess Naïve IgG. Histogram view of competition experiment shown in figure 6 of chapter 2.

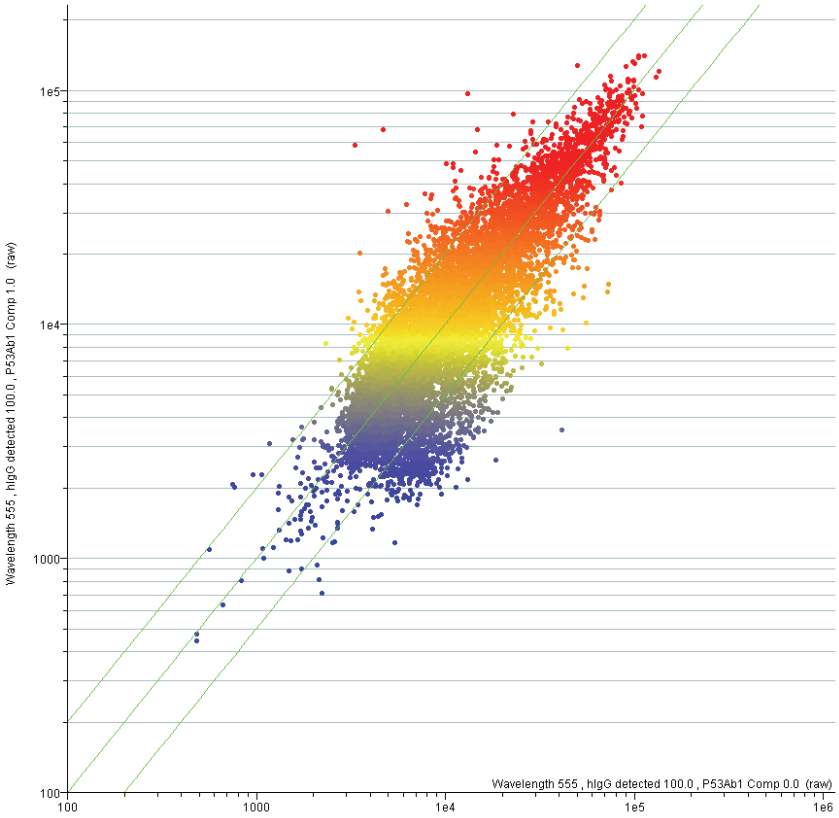


Figure A-7. Pooled human IgG with and without Monoclonal. The pooled naïve human IgG at 10uM (100X) was detected with a species specific secondary with (y-axis) and without the presence of the P53Ab1 monoclonal at 100nM.

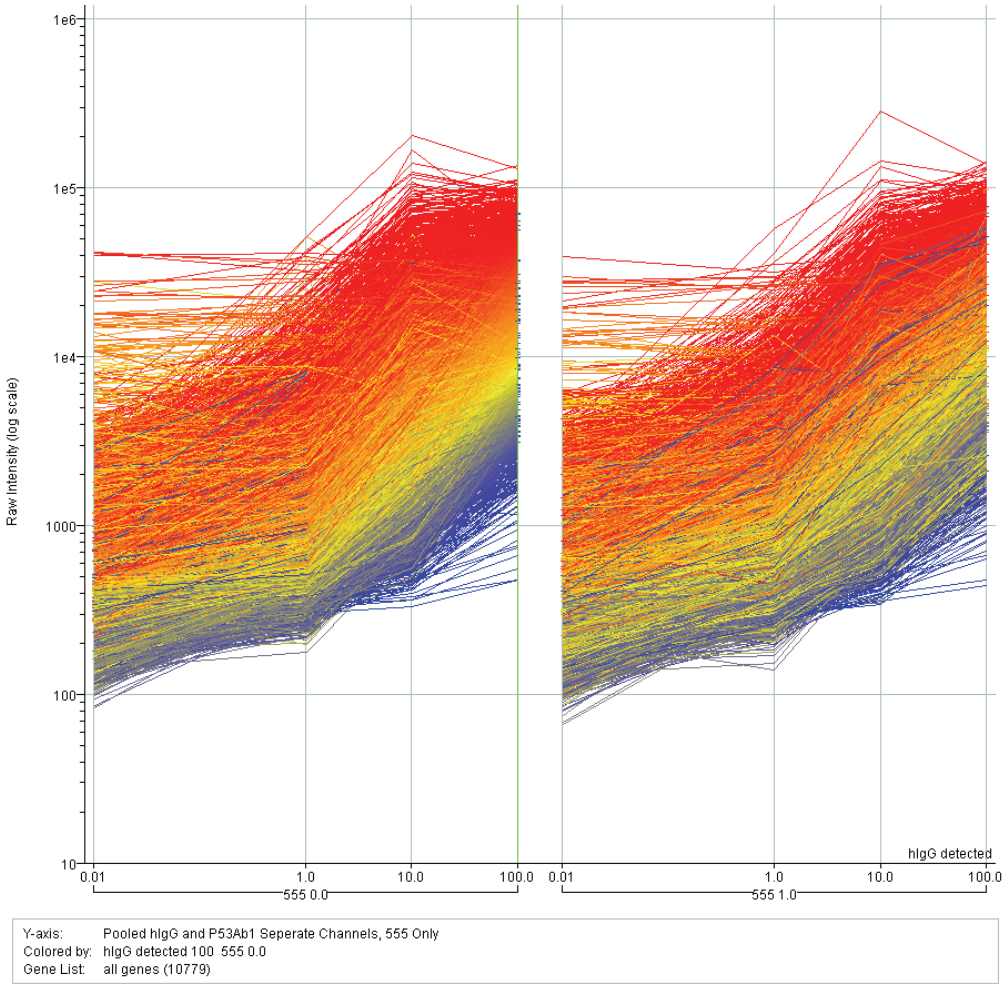


Figure A-8. Concentration Series of Pooled Human IgG with and without Monoclonal. The x-axis shown the human IgG concentration series, where 0.01 corresponds to no human IgG, 1.0 corresponds to 100nM, 10.0 corresponds to 1uM, and 100 corresponds to 10uM. The graph on the left shows the human IgG alone, and the graph on the right shows human IgG in the presence of the P53Ab1 monoclonal at 100nM. All of the peptides appear to increase in intensity from 100nM (1.0) to 1uM (10.0) in both graphs.

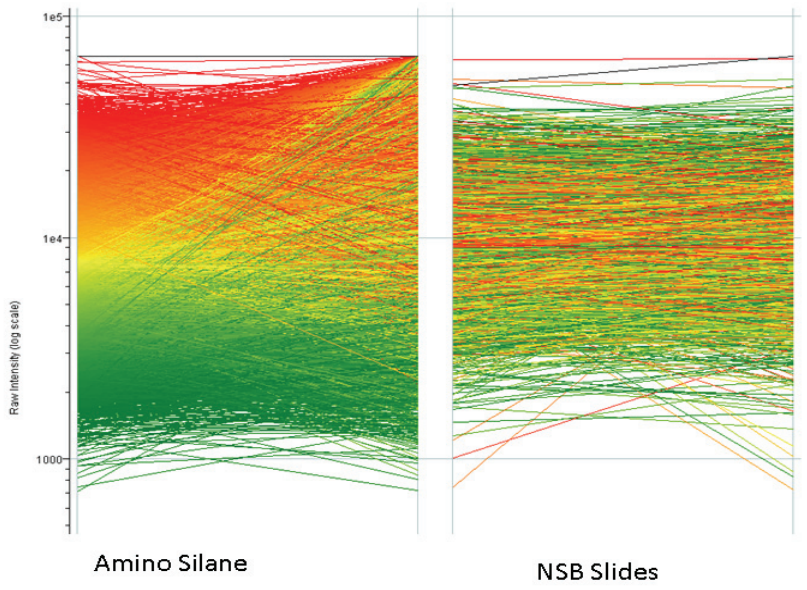


Figure A-9. Amino Silane vs NSB reproducibility. Replicates were run on the of the anti-FT05 sera on both the amino silane coated slides and the NSB slides. The line graph above illustrates how the replicates compare. The amino silane slides appear to have more differences in peptide intensities between replicates

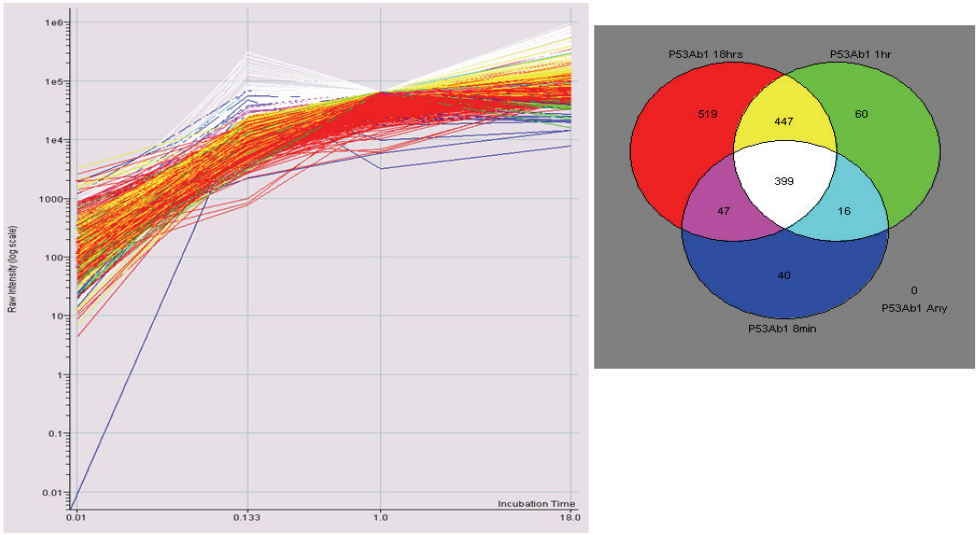


Figure A-10. Incubation Time. The p53Ab1 antibody was used to probe the array with different incubation times. The peptides having signal intensities greater than 60,000 RFU were selected for each incubation time. The signal intensities continue to increase up to 18hrs.

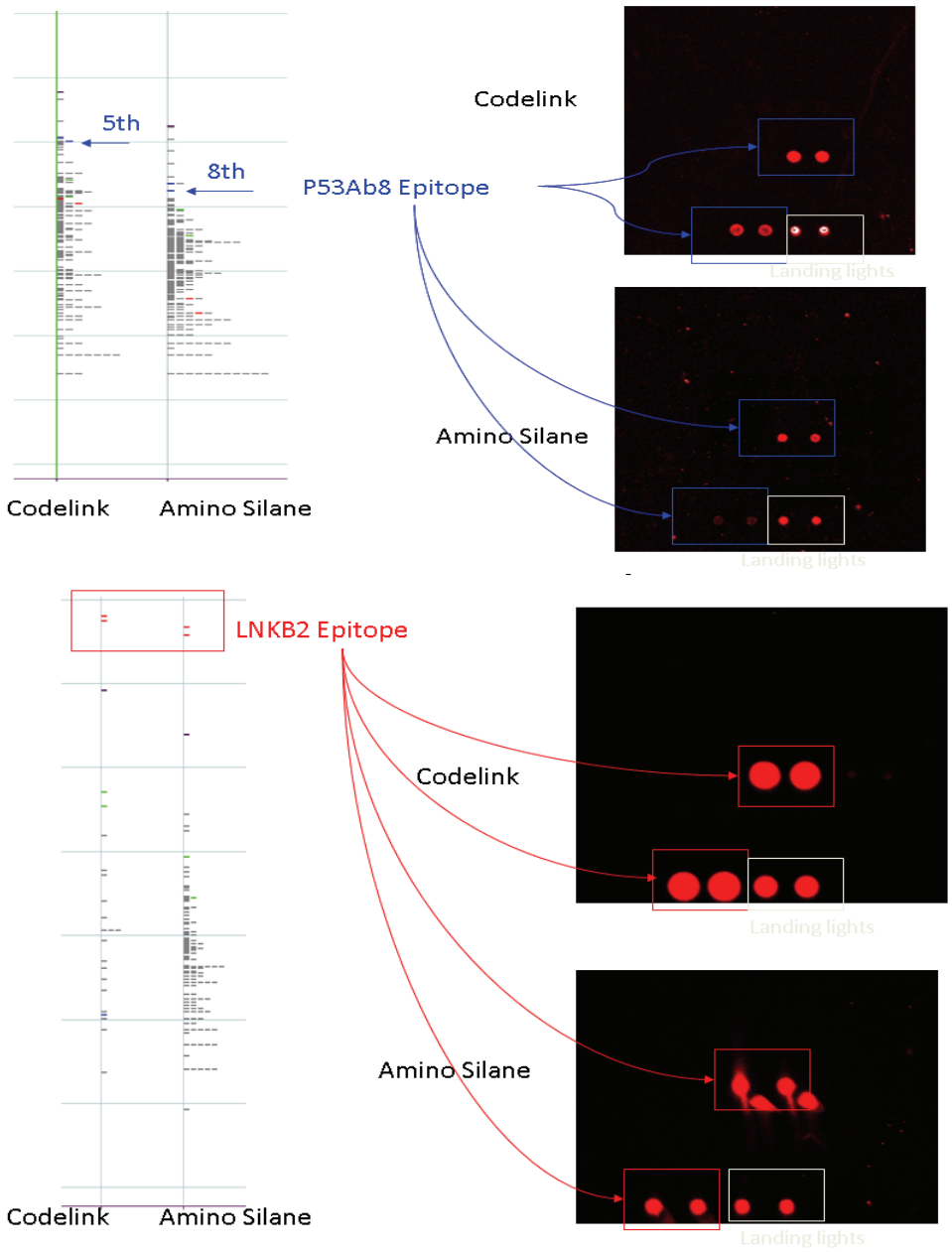


Figure A-11. Epitopes on Codelink vs. Aminosilane. A custom microarray containing to monoclonal epitopes was printed on codelink and aminosilane slides. The codelink appears to have better separation of the cognate epitopes from the other peptides

Name	Peptide Sequence	PolarIndex	Overall Aggregation Score (AGGRESCAN algo)	Number of Aggregating Hot Spots (AGGRESCAN algo)	Hotspot Aggregation Score (AGGRESCAN algo)	Predicted Secondary Structure	Molecular Wt...	Theoretical pI (Protparam)	Ext. Coefficient (Protparam)	Aliphatic Index (Protparam)	Grand average Hydropath...
FP6	DSPQKSDDDISQNFHGRGSC	-0.35	-0.64	0	0.00	LLLLLLLHHHLLLLLLLLL	2193	4.6	0	20	-1.60
FT02	TALPDPTKWPEHHQYFAGSC	0.07	0.02	1	2.72	LLLLLLLLLHHHHHHLLLLL	2348	6.0	6990	49	-0.41
FT03	KANWFDFKTPNQMTQVWGSC	0.02	0.03	1	1.19	LLLELLLLHHHHHHHHLLL	2439	8.2	11000	20	-0.55
FT05	KTFKSEPAVNIESNSSTGSC	-0.05	-0.23	0	0.00	LLLLLLLLLLLLLLLLLLL	2150	6.1	1490	25	-0.95
KA24	FQIHNRSWQQSMWPNEYGSC	-0.07	-0.07	0	0.00	LLLLLLLHHHHHHHLLLLL	2549	6.7	12490	20	-1.01
KA4	MPHKHKNTHAQVLPNWRGSC	-0.28	-0.31	0	0.00	LLLLLLLLHHHLLLLLLLL	2342	10.1	5500	39	-1.21
Neq1	EFSNPTAQVPPDFWMSDGSC	-0.01	-0.03	0	0.00	LLLLLLLLLLLLLLLLLLL	2265	3.5	5500	20	-0.32
Rco3	TIPAHNIFWTLLYFSTGTGSC	0.51	0.52	1	10.13	LLLLLLEEEEEEEELLLLL	2241	6.4	6990	103	0.87
	Set Average	**-0.02**	**-0.09**	**0.38**	**1.76**	**NA**	**2316**	**6.4**	**6245**	**37**	**-0.61**
	10K Average	**0.00**	**-0.05**	**0.70**	**2.42**	**NA**	**2302**	**6.9**	**6203**	**52**	**-0.51**

Figure A-12. Peptides used in custom array. These peptides were selected to be printed in a custom array because there was immune sera raised against them, and they represent a range in physiochemical properties.

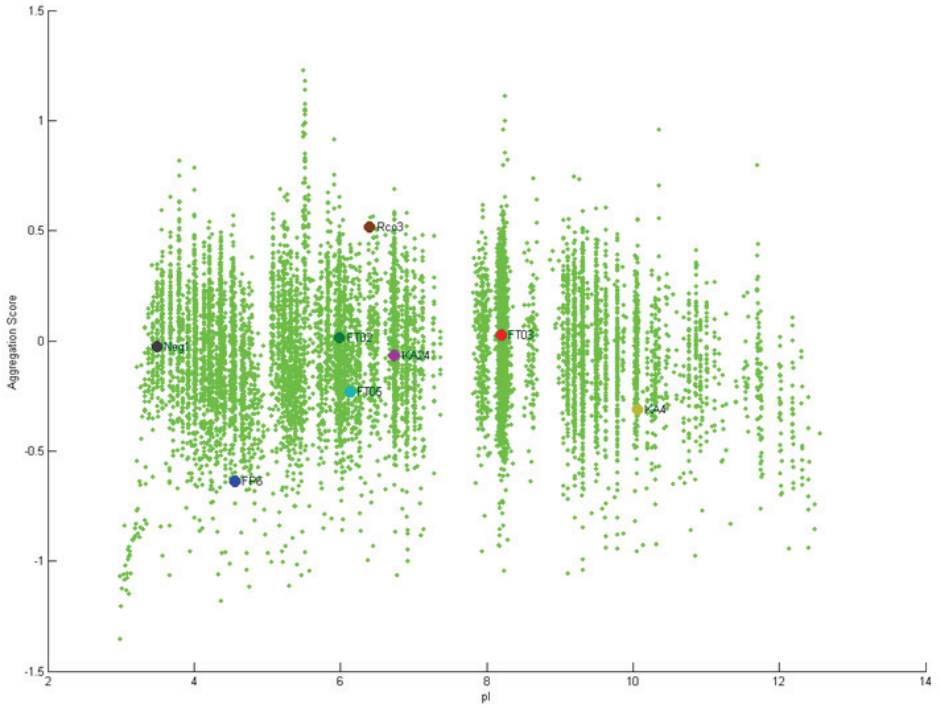


Figure A-13. Distribution of Peptides by pI and Aggregation Score. All peptides on the 10K array are plotted by there isoelectric point vs. their aggregation score. The peptides selected for the custom array are annotated.

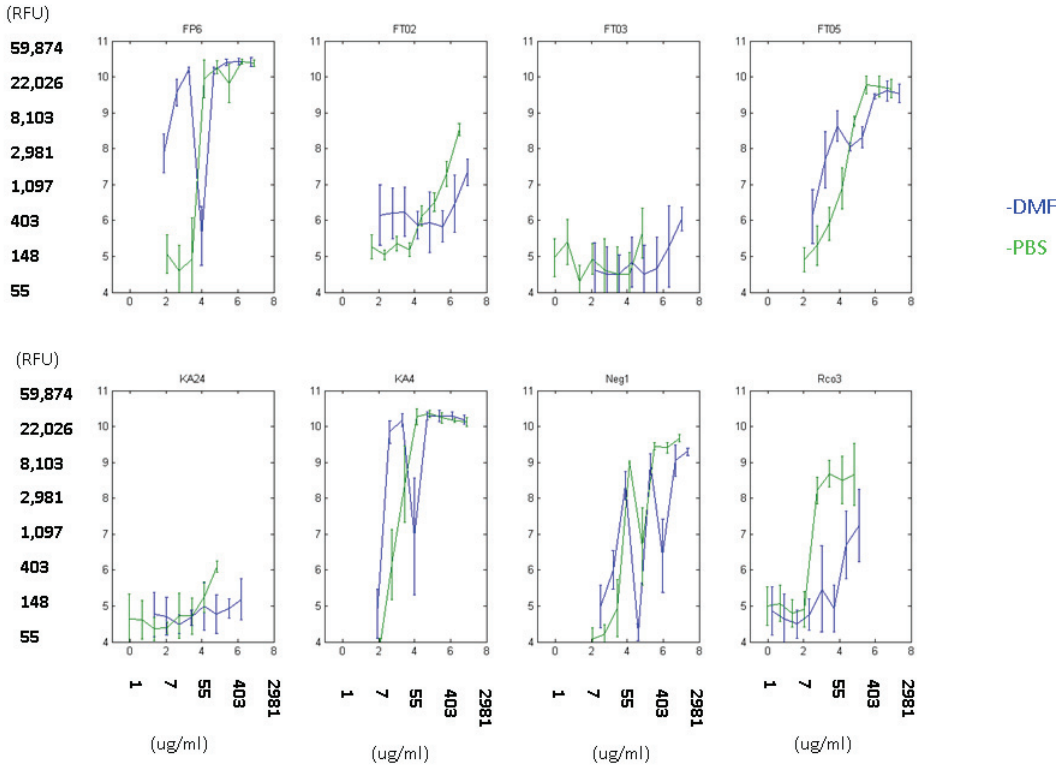


Figure A- 14. Anti-peptide sera on dilution series of peptides in custom array. Anti-peptide sera was used to probe the custom array. The signals for the cognate peptide binding are plotted against the peptide concentration. The color indicates whether the peptides were diluted in DMF (blue) or PBS (green).

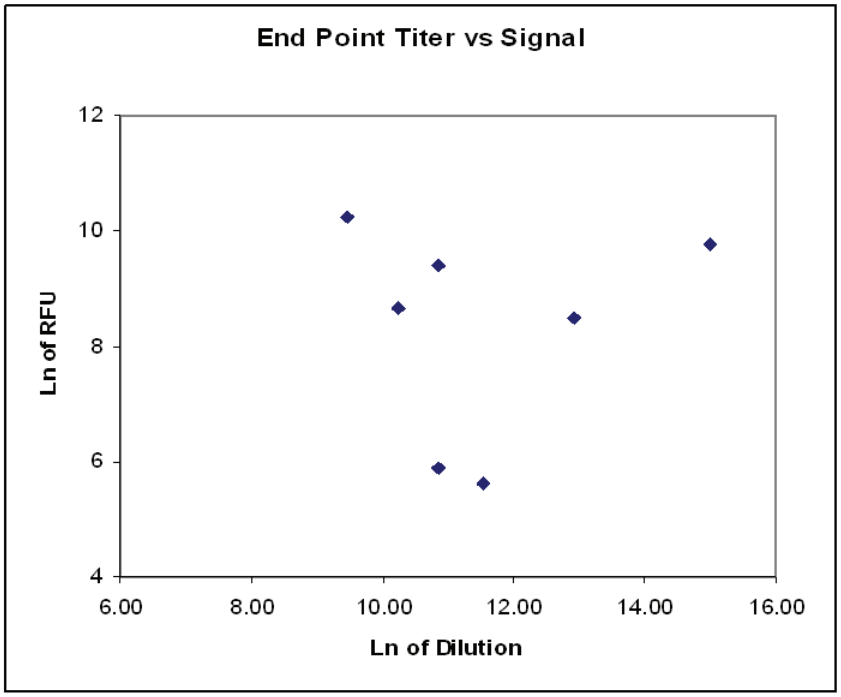
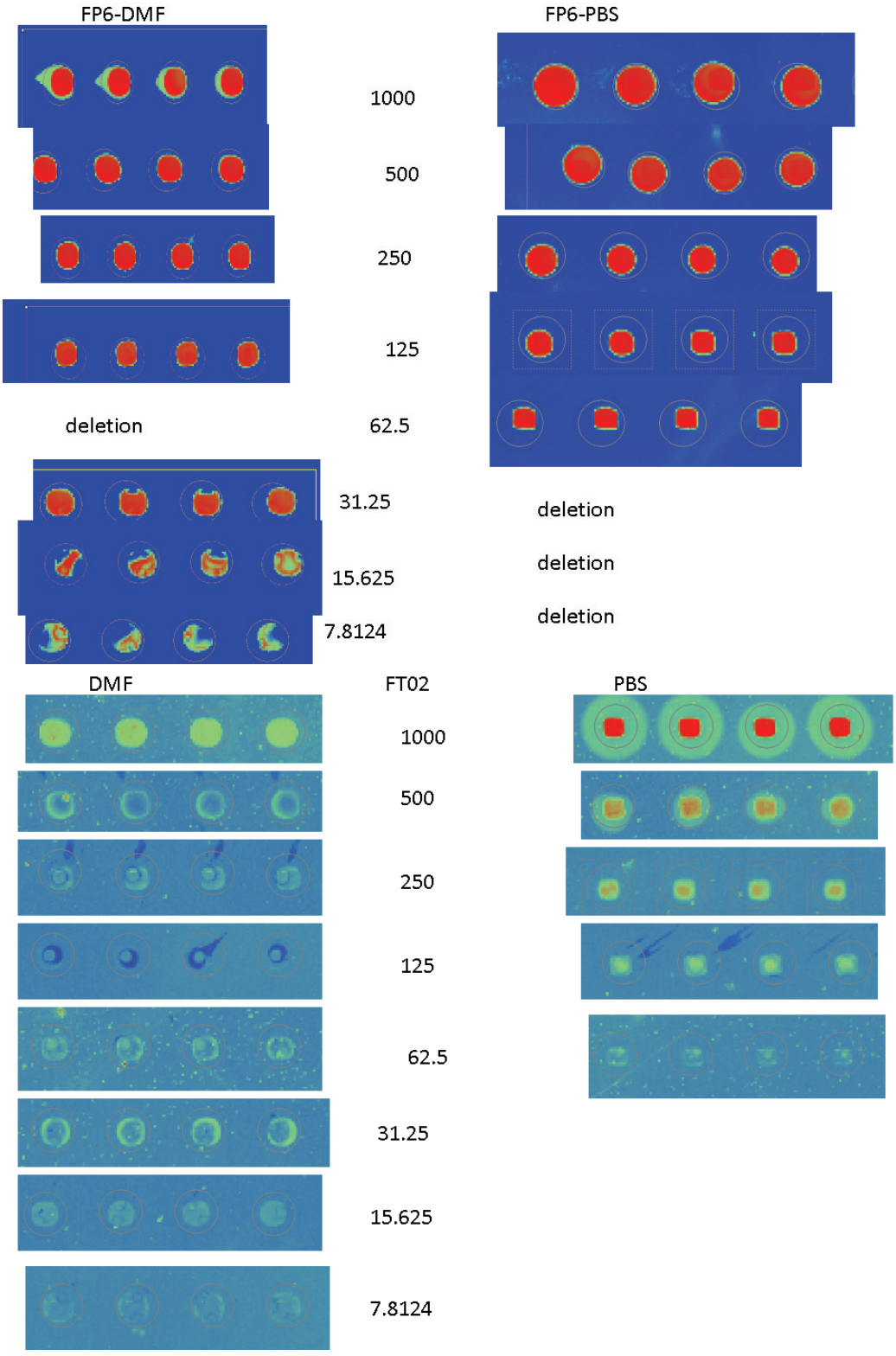


Figure A-15. Signal Intensity vs. End Point Titer. The signal intensity at the highest concentration point for anti-sera binding to the cognate peptide is plotted against the end point titer measured by ELISA. There does not appear to be any correlation between the signal intensity and relative affinity.


FP6-DMF
FP6-PBS
1000
500
250
125
deletion
62.5
31.25
15.625
7.8124
deletion
deletion
deletion
DMF
FT02
PBS
1000
500
250
125
62.5
31.25
15.625
7.8124

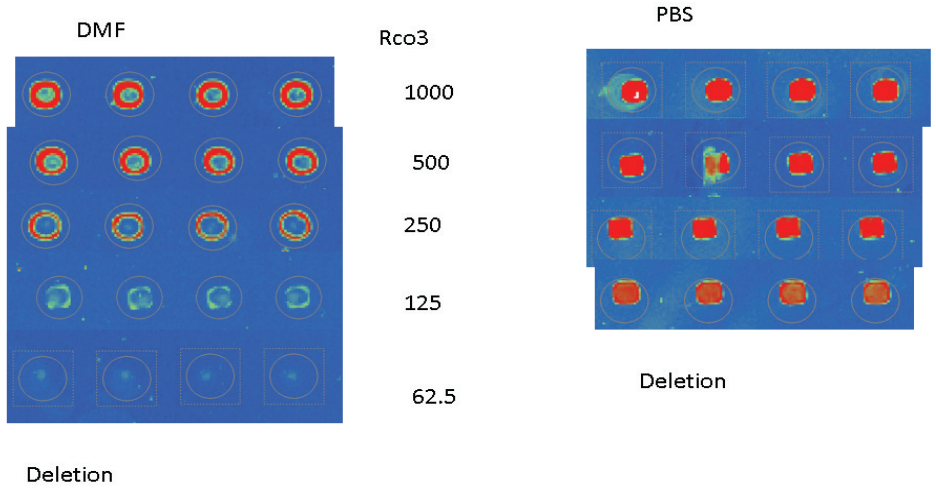
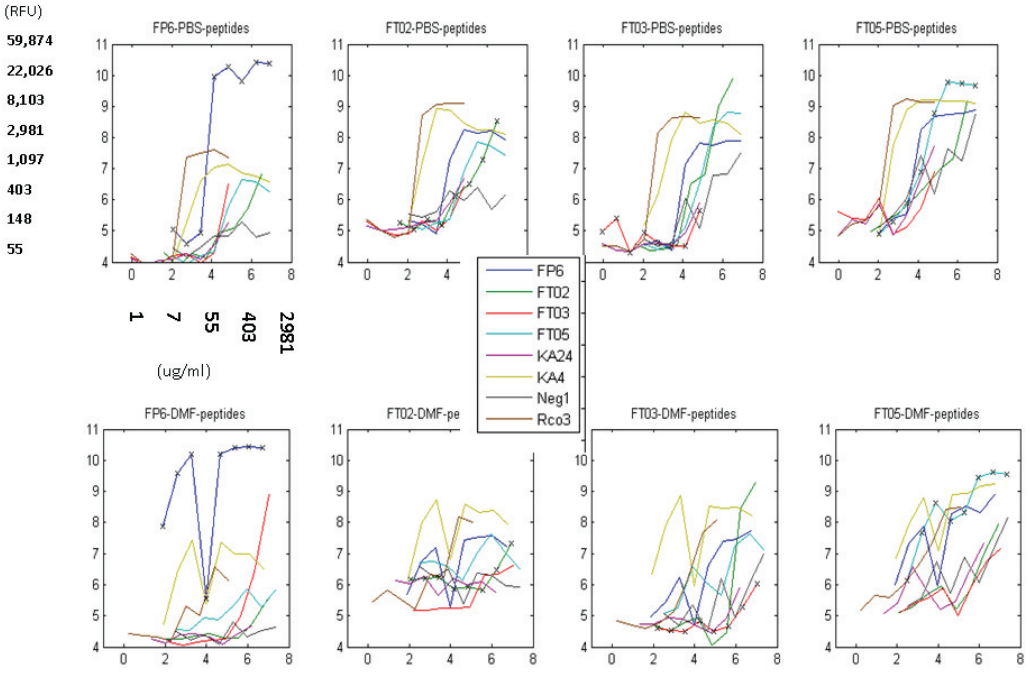


Figure A-16. Screen Shots of Peptide Dilution Series.



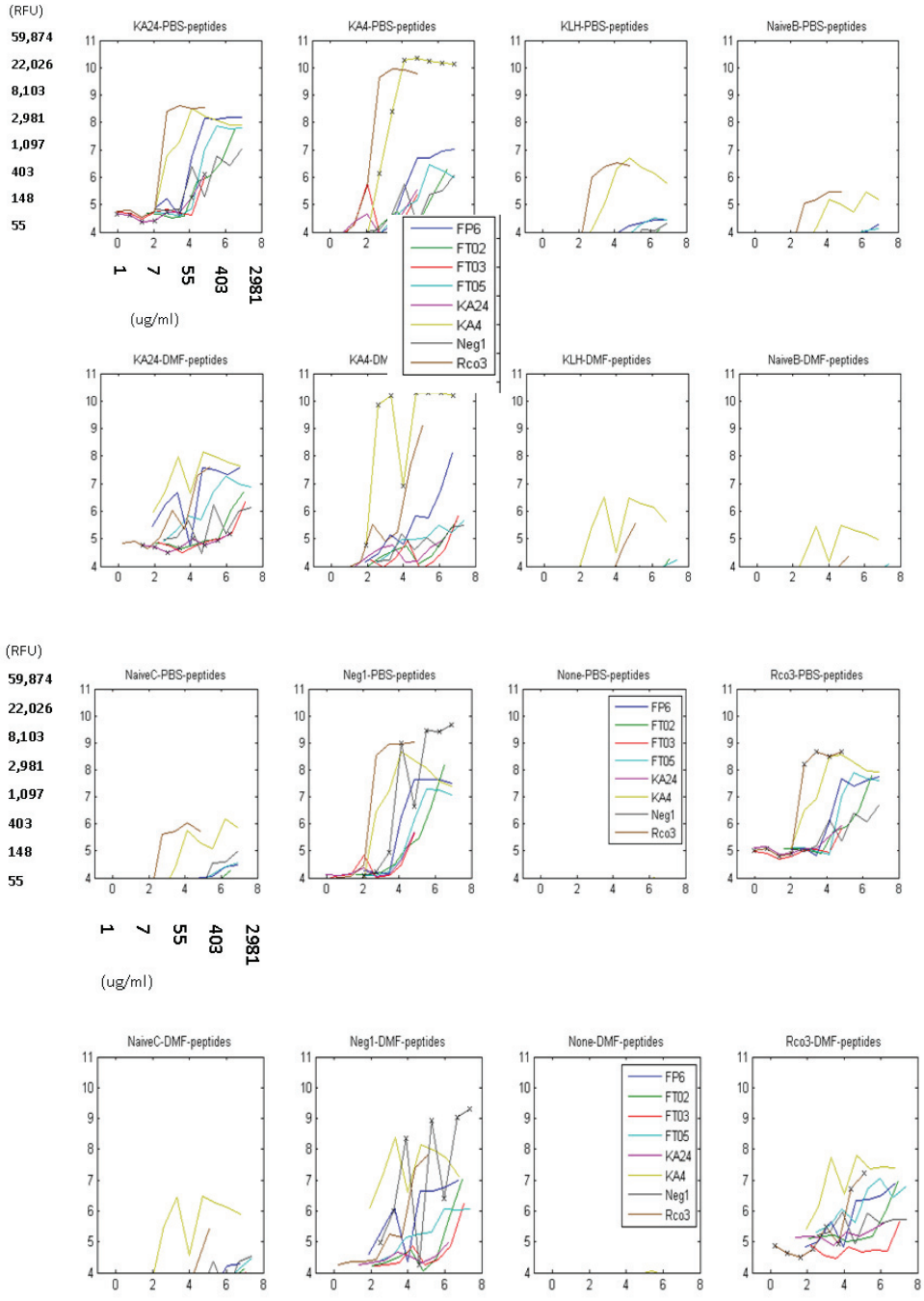


Figure A-17. Crossreactivity of peptide immune sera by peptide concentration. Signal intensities for all anti-peptide sera are plotted against each peptide. Each plot represents one peptide and the x-axis shows the dilutions of that peptide. There does not appear to be much difference between the peptide dilution that the specific immune sera recognizes compared to cross-reactive immune sera.

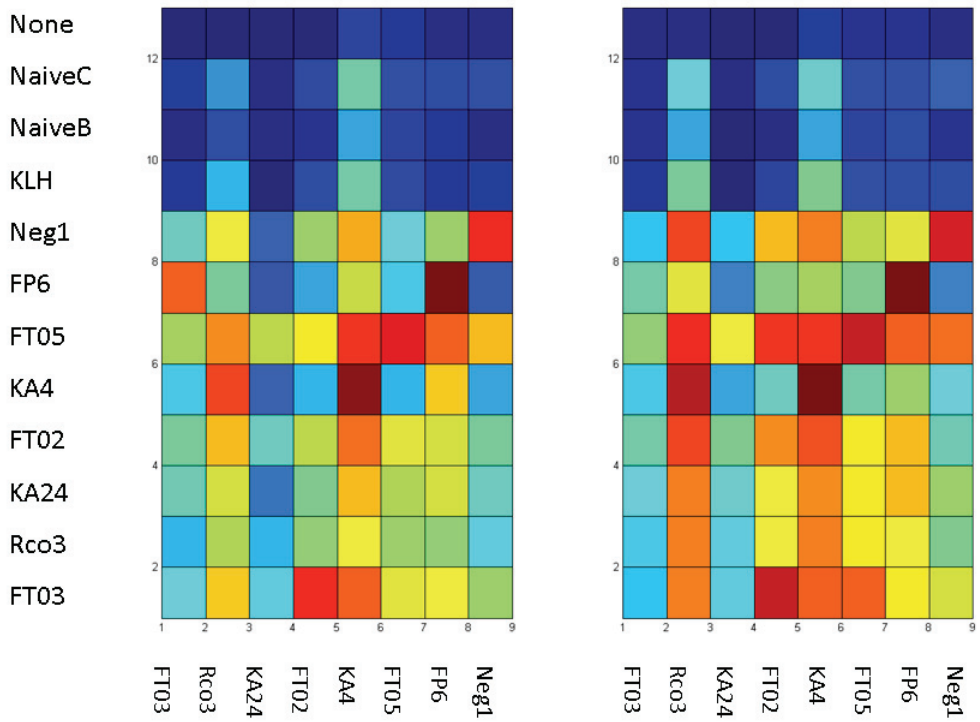


Figure A-18. Crossreactivity of peptide sera summary. Heat map shows the signal intensities at the tope peptide concentration for each anti-peptide sera peptide pair. Sera type is shown in the rows and peptides are shown in the columns. Data for peptides diluted in DMF is shown on the left and peptides in PBS is shown on the right.

APPENDIX B

SUPPLEMENTAL FIGURES FOR CHAPTER 3

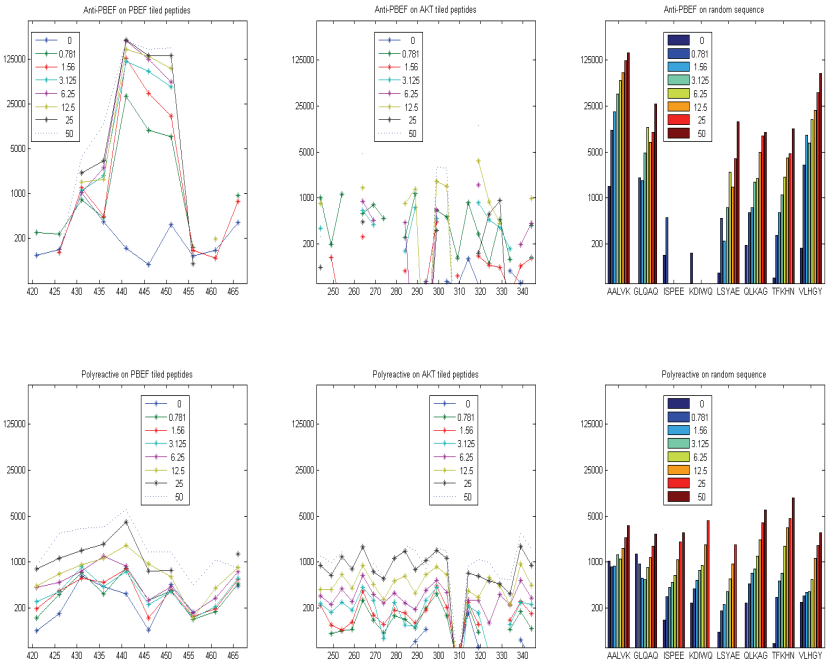


Figure B-1. PBEF and 2E4 Signal Intensities by Antibody Concentration at 1 hour. The signal intensities for the antibodies are plotted against for peptide on the x-axis. The top row shows the anti-PBEF experiments and the bottom row shows the 2E4 experiments. Each line or bar represents a different antibody concentration. The signals across peptide concentrations were averaged for this graph.

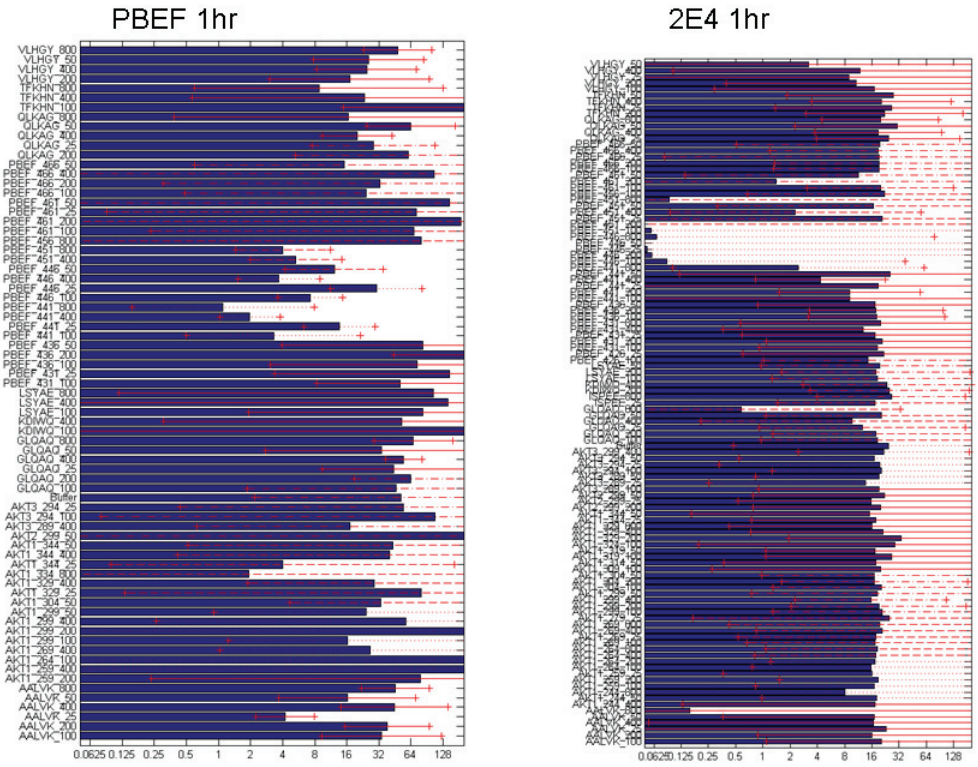
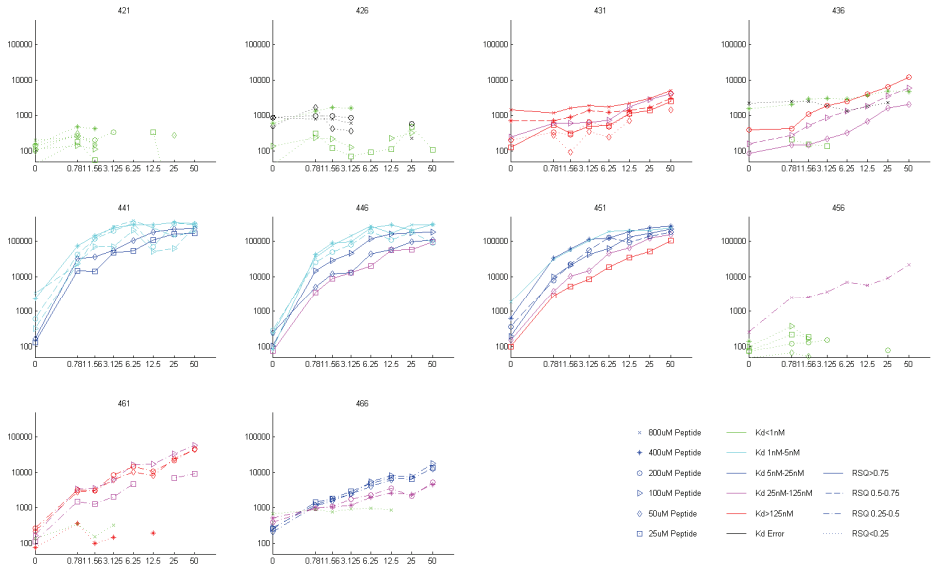


Figure B-2. Half maximal binding for all peptide conc 1hr incubation. The half maximal binding concentration is shown for each peptide concentration that could be fit for the anti-PBEF on the left and 2E4 on the right. Error bars represent 95% confidence intervals on the half maximal binding estimates.



Figurer B-3. Binding Curves for anti-PBEF on PBEF tiling peptides 1hr incubation. The signal intensities for the anti-PBEF are plotted against the anti-PBEF concentration on the x-axis. Each plot represents a PBEF tiling peptide and each line represents a peptide concentration. The color indicates the half maximal binding concentration and the solid or dashed indicate the R-squared for the curve fit.

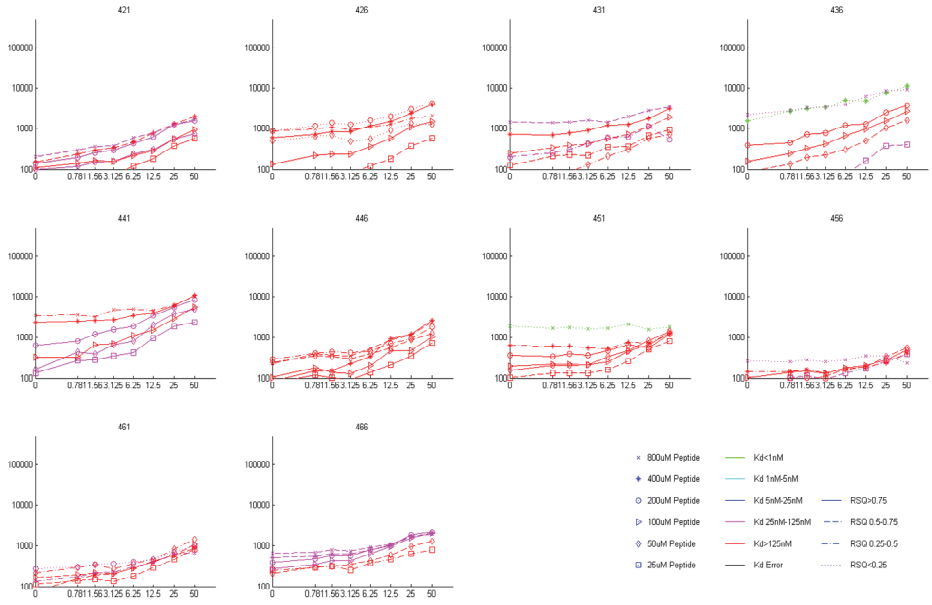


Figure B-4. Binding Curves for 2E4 on PBEF tiling peptides 1hr incubation. The signal intensities for the 2E4 are plotted against the 2E4 concentration on the x-axis. Each plot represents a PBEF tiling peptide and each line represents a peptide concentration. The color indicates the half maximal binding concentration and the solid or dashed indicate the R-squared for the curve fit.

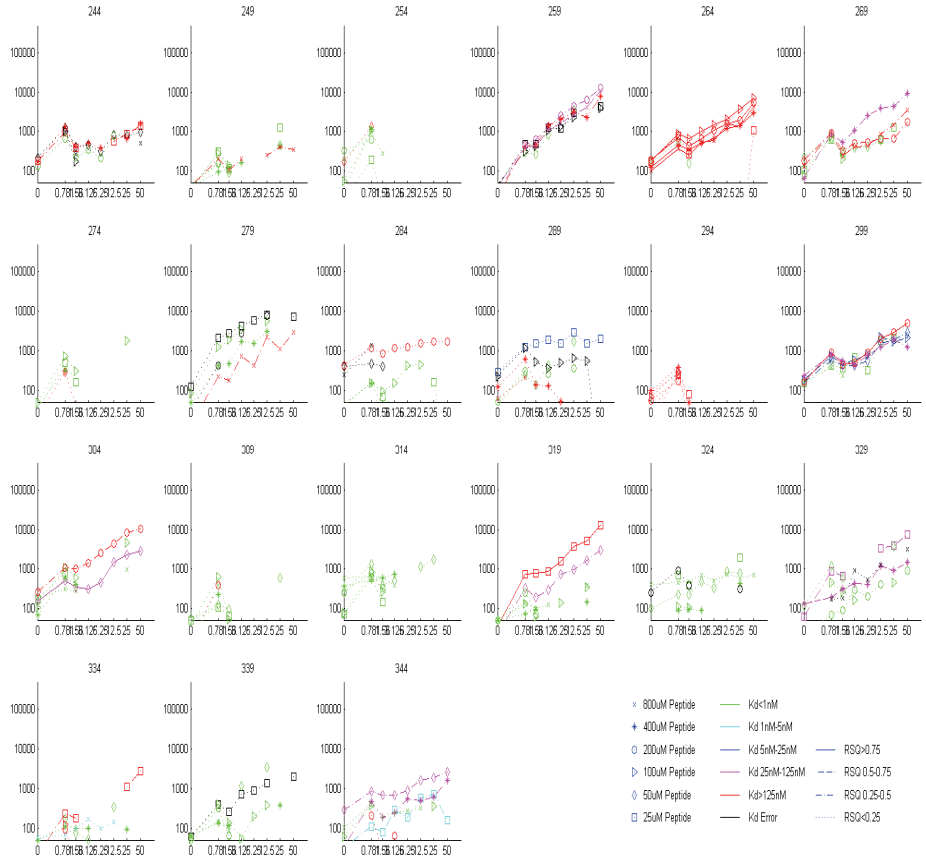


Figure B-5. Binding Curves for anti-PBEF on AKT peptides 1hr. The signal intensities for the anti-PBEF are plotted against the anti-PBEF concentration on the x-axis. Each plot represents an AKT tiling peptide and each line represents a peptide concentration. The color indicates the half maximal binding concentration and the solid or dashed indicate the R-squared for the curve fit.

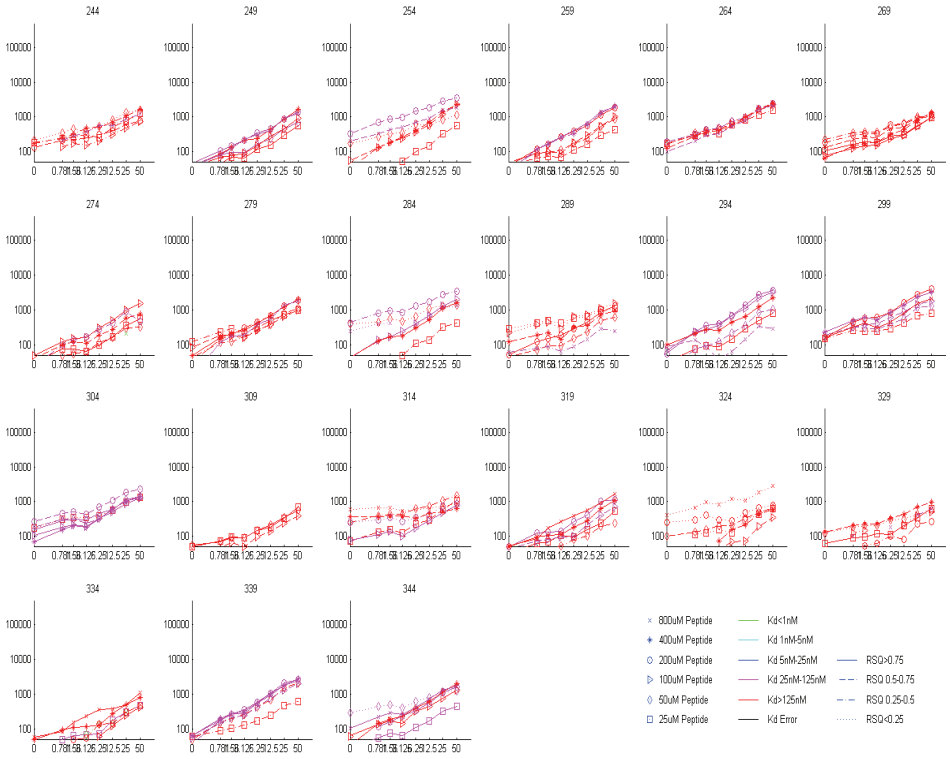


Figure B-6. Binding Curves for 2E4 on PBEF peptides 1hr incubation. The signal intensities for the 2E4 are plotted against the 2E4 concentration on the x-axis. Each plot represents an AKT tiling peptide and each line represents a peptide concentration. The color indicates the half maximal binding concentration and the solid or dashed indicate the R-squared for the curve fit.

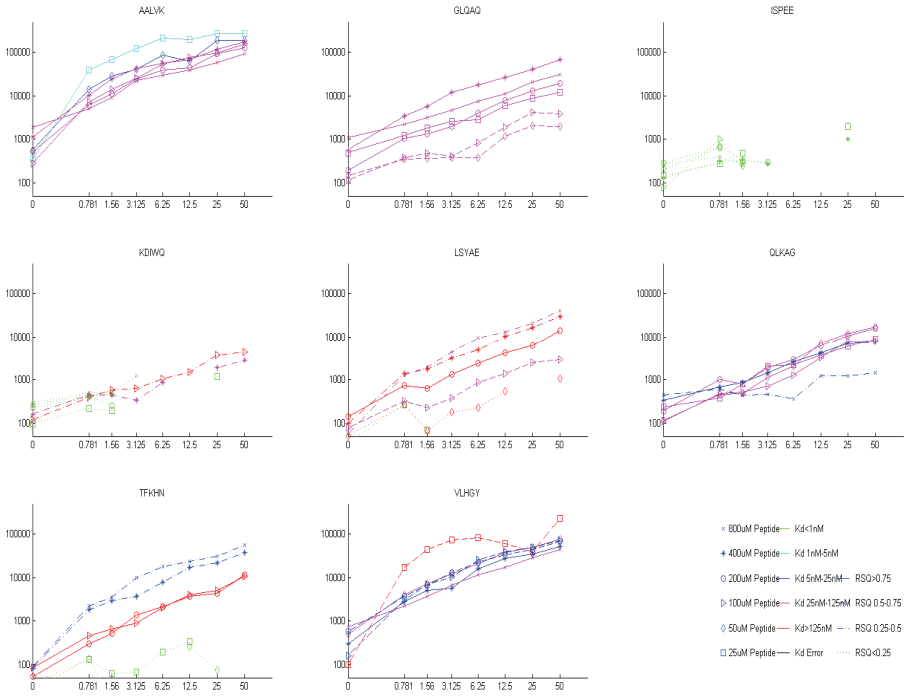


Figure B-7. Binding Curves for PBEF on Rand Peptides 1hr incubation. The signal intensities for the anti-PBEF are plotted against the anti-PBEF concentration on the x-axis. Each plot represents a peptide selected from the random sequence array and each line represents a peptide concentration. The color indicates the half maximal binding concentration and the solid or dashed indicate the R-squared for the curve fit.

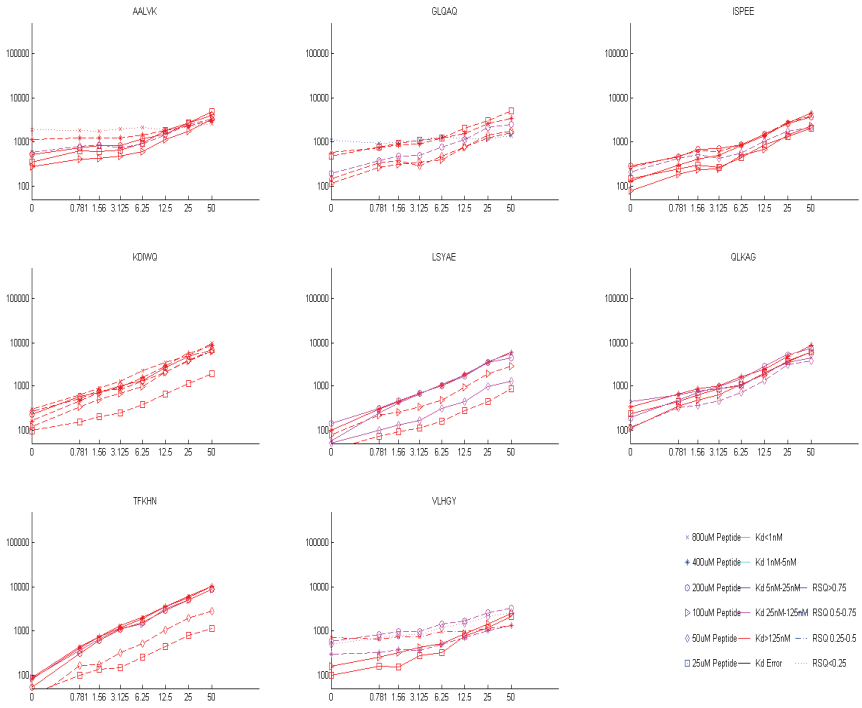


Figure B-8. Binding Curves for 2E4 on Rand Peptides 1hr Incubations. The signal intensities for the 2E4 are plotted against the 2E4 concentration on the x-axis. Each plot represents a peptide selected from the random sequence array and each line represents a peptide concentration. The color indicates the half maximal binding concentration and the solid or dashed indicate the R-squared for the curve fit.

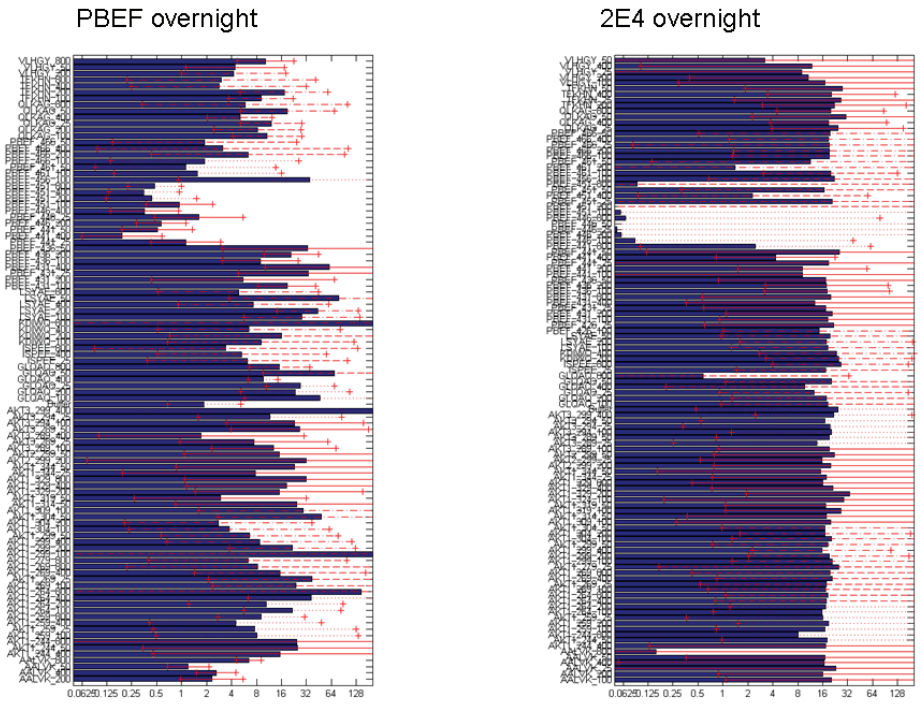


Figure B-9. Half maximal binding for all peptide concentrations overnight incubation. The half maximal binding concentration is shown for each peptide concentration that could be fit for the anti-PBEF on the left and 2E4 on the right. Error bars represent 95% confidence intervals on the half maximal binding estimates.

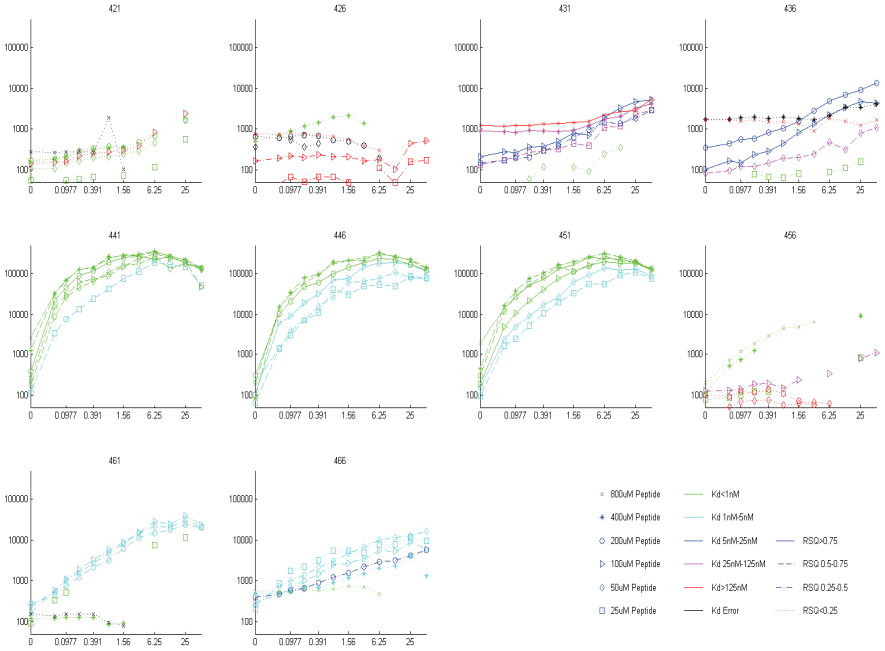


Figure B-10. Binding Curves for anti-PBEF on PBEF peptides overnight incubation. The signal intensities for the anti-PBEF are plotted against the anti-PBEF concentration on the x-axis. Each plot represents a PBEF tiling peptide and each line represents a peptide concentration. The color indicates the half maximal binding concentration and the solid or dashed indicate the R-squared for the curve fit.

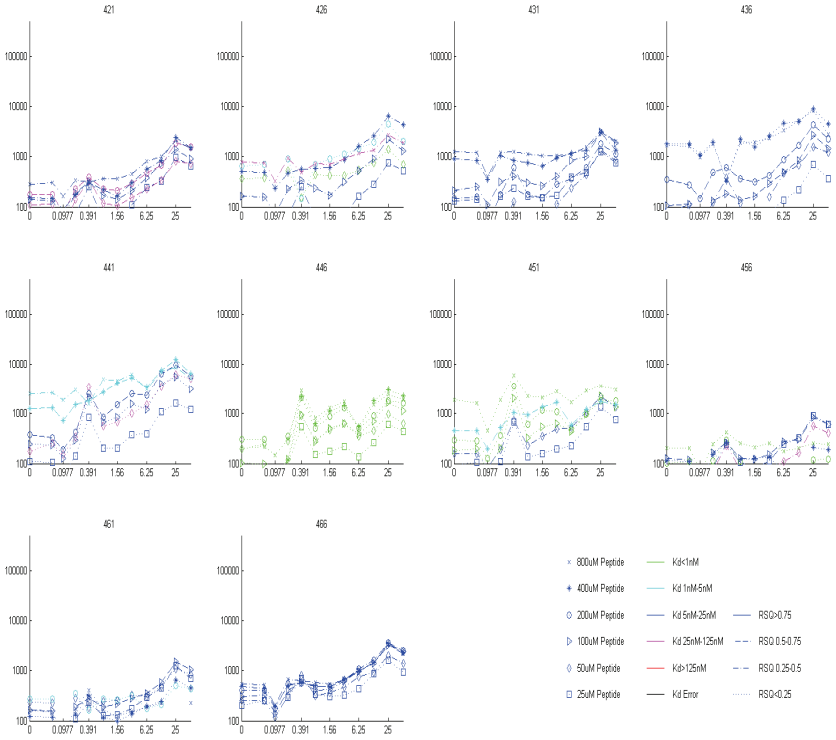


Figure B-11. Binding Curves for 2E4 on PBEF Peptides overnight incubation. The signal intensities for the 2E4 are plotted against the 2E4 concentration on the x-axis. Each plot represents a PBEF tiling peptide and each line represents a peptide concentration. The color indicates the half maximal binding concentration and the solid or dashed indicate the R-squared for the curve fit.

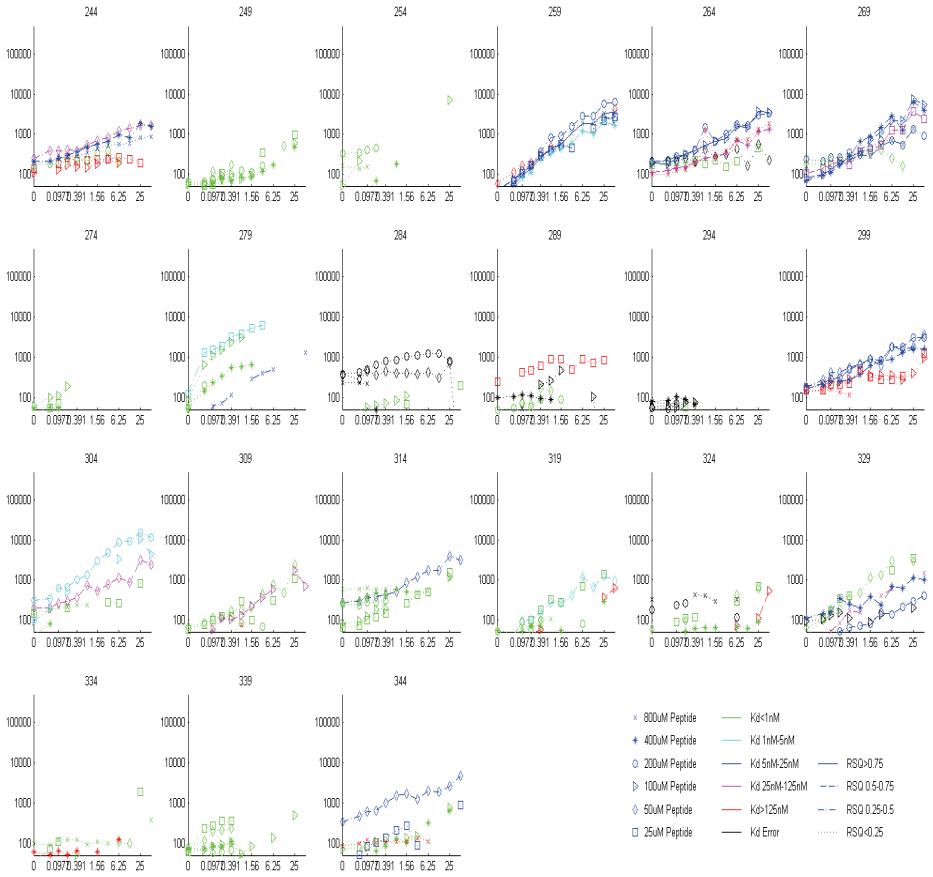


Figure B-12. Binding Curves for anti-PBEF on AKT peptides overnight incubation. The signal intensities for the anti-PBEF are plotted against the anti-PBEF concentration on the x-axis. Each plot represents an AKT tiling peptide and each line represents a peptide concentration. The color indicates the half maximal binding concentration and the solid or dashed indicate the R-squared for the curve fit.

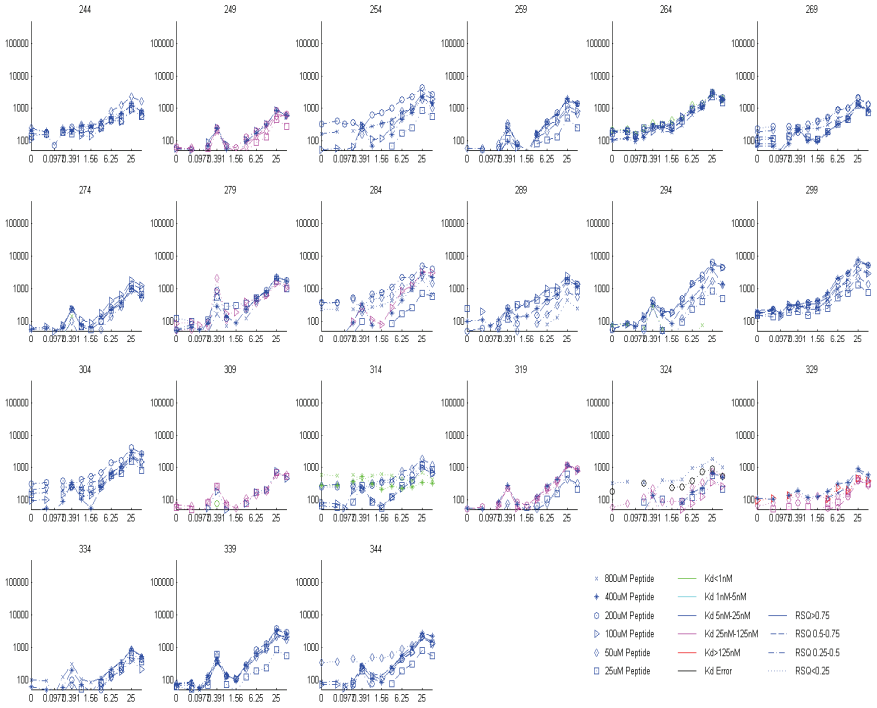


Figure B-13. Binding Curves for 2E4 on AKT peptides overnight incubation. The signal intensities for the 2E4 are plotted against the 2E4 concentration on the x-axis. Each plot represents an AKT tiling peptide and each line represents a peptide concentration. The color indicates the half maximal binding concentration and the solid or dashed indicate the R-squared for the curve fit.

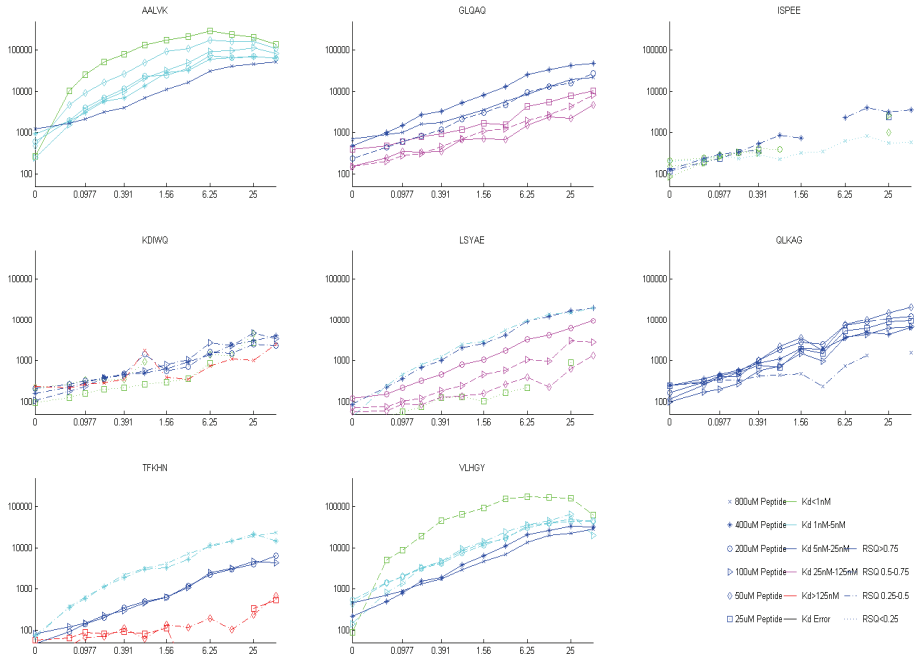


Figure B-14. Binding Curves for anti-PBEF on Rand Peptides. The signal intensities for the anti-PBEF are plotted against the anti-PBEF concentration on the x-axis. Each plot represents a peptide selected from the random sequence array and each line represents a peptide concentration. The color indicates the half maximal binding concentration and the solid or dashed indicate the R-squared for the curve fit.

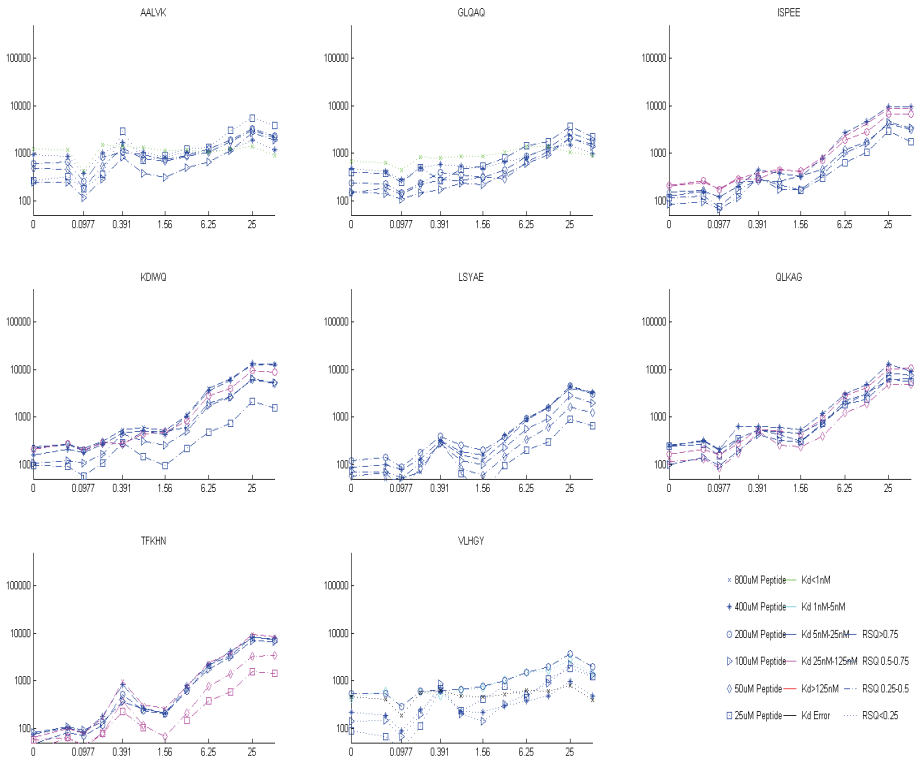


Figure B-15. Binding Curves for 2E4 on Random Peptides. The signal intensities for the anti-PBEF are plotted against the anti-PBEF concentration on the x-axis. Each plot represents a peptide selected from the random sequence array and each line represents a peptide concentration. The color indicates the half maximal binding concentration and the solid or dashed indicate the R-squared for the curve fit.

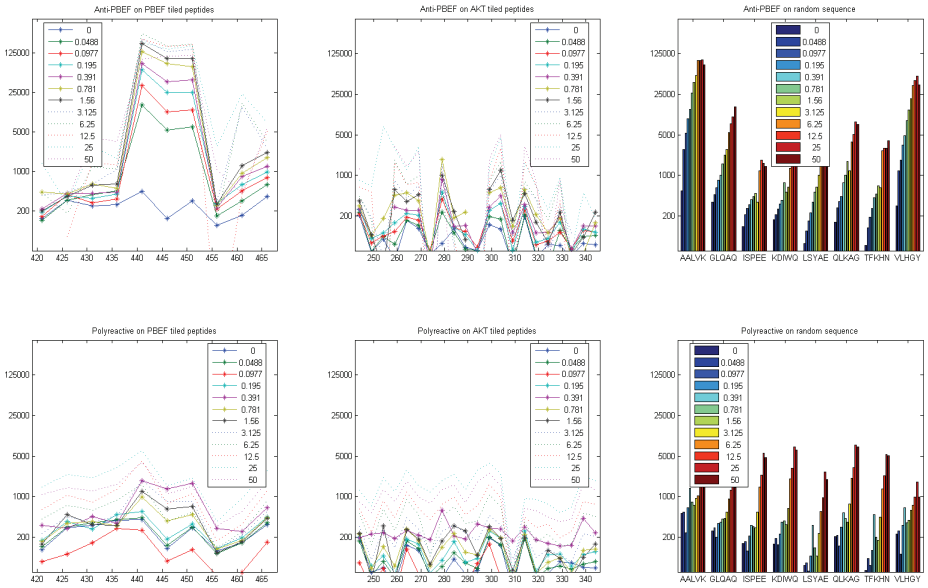


Figure B-16. . PBEF and 2E4 Signal Intensities by Antibody Concentration overnight. The signal intensities for the antibodies are plotted against for peptide on the x-axis. The top row shows the anti-PBEF experiments and the bottom row shows the 2E4 experiments. Each line or bar represents a different antibody concentration. The signals across peptide concentrations were averaged for this graph

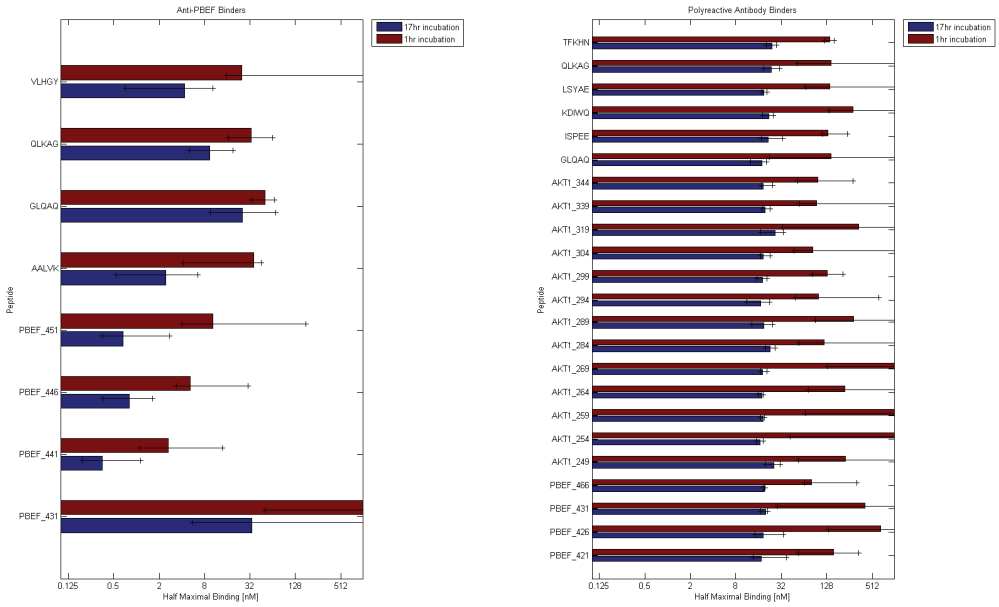


Figure B-17. Half Maximal binding of anti-PBEF and 2E4 on Selected Peptides. The half maximal binding concentrations were estimated from both 1hr incubation experiments (shown in red) and overnight incubations (shown in blue). The polyreactive antibody (on the right) shows decreases in the half maximal binding concentration at 17hrs indicating greater apparent affinity.

APPENDIX C

SUPPLEMENTAL DATA FOR CHAPTER 4

	Platform	binders count	specific count
11D3	CIM10Kv1	548	14
1C10	CIM10Kv1	1448	12
AbcamHA	CIM10Kv1	12	1
DM1A	CIM10Kv1	5201	388
Endorphin	CIM10Kv1	1813	115
HTF14	CIM10Kv1	1158	6
IL2	CIM10Kv1	2463	35
LNKB2	CIM10Kv1	2051	29
MHC	CIM10Kv1	5	0
P53Ab1	CIM10Kv1	5208	1401
P53Ab8	CIM10Kv1	4423	223
TNF	CIM10Kv1	5199	715
TP Cocci	CIM10Kv1	1489	71
b78	CIM10Kv1	3010	256
b96	CIM10Kv1	182	5
herceptin	CIM10Kv1	4618	1345
87G	CIM10Kv2	123	7
LeuEnk	CIM10Kv2	1440	33
P53Ab1	CIM10Kv2	288	8
P53Ab8	CIM10Kv2	28	1
PBEF	CIM10Kv2	2583	88
Poly	CIM10Kv2	5711	3931
V5	CIM10Kv2	27	1
cMyc	CIM10Kv2	5250	1948

Table C-1 Antibody Specific Peptides Counts. The "binders count" indicates the number of peptides found to bind above background for each antibody. The "specific count" indicates the number of peptides with specificity scores (as defined in chapter 4) over 0.7.

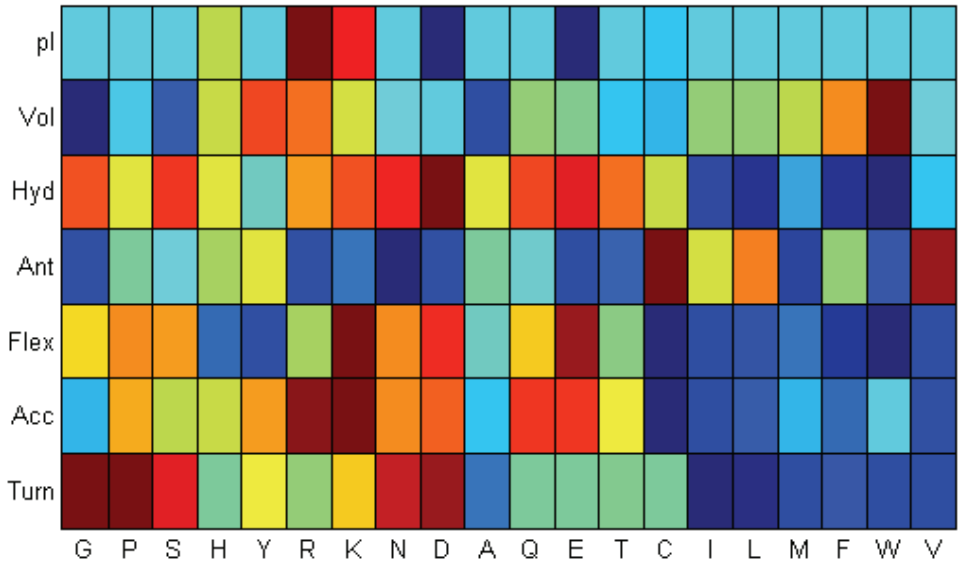


Figure C-1. Amino Acid Contributions to Peptide Properties. The score for each amino acid for each scale used in Chapter 4 is shown as a heatmap. The minimum value of each scale is shown as blue and the maximum is shown as dark red

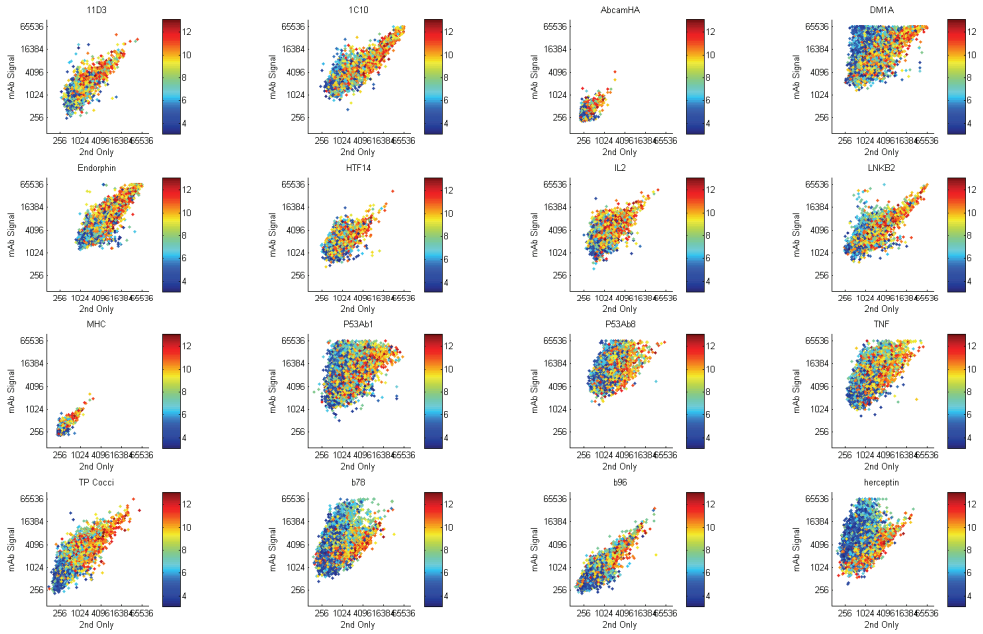


Figure C-2. Monoclonal Signals vs. Secondary Alone Colored by pI CIM10Kv1. The signal intensities for each monoclonal are plotted against the signal intensities of the secondary only control. The colors represent the isoelectric point of that peptide.

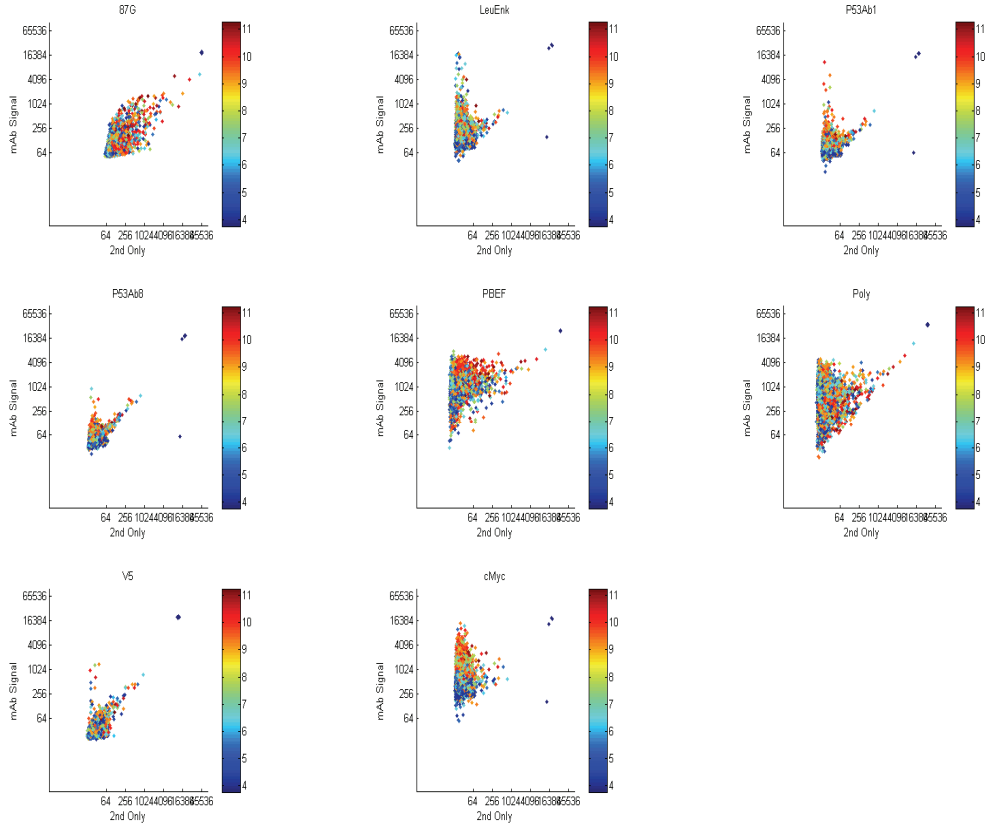


Figure C-3. Monoclonal Signals vs. Secondary Alone Colored by pI CIM10Kv2. The signal intensities for each monoclonal are plotted against the signal intensities of the secondary only control. The colors represent the isoelectric point of that peptide

.

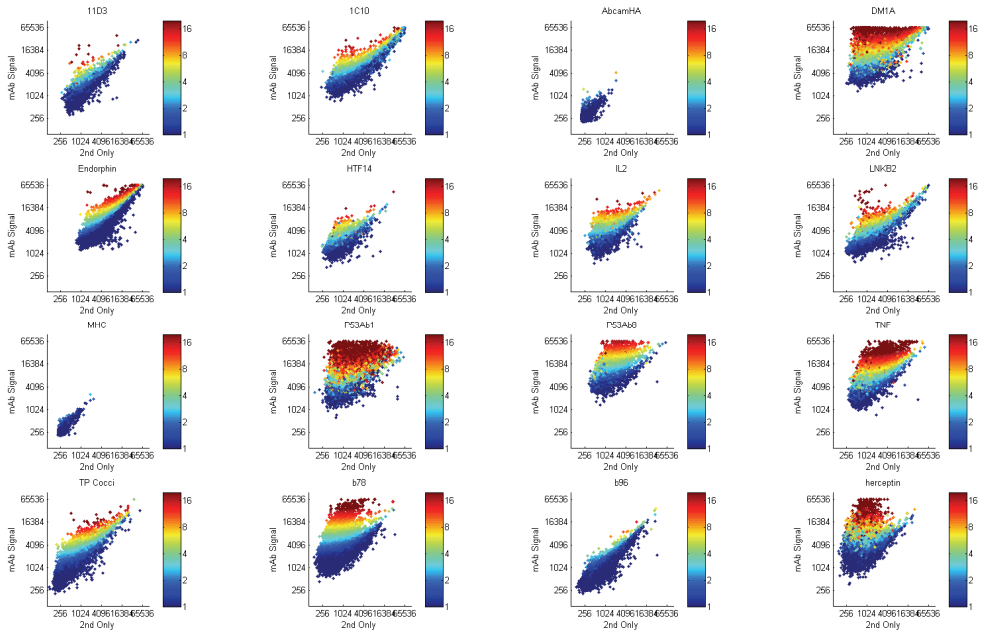


Figure C-4. Monoclonal Signals vs. Secondary Alone Colored by Normalized Signal CIM10Kv1. The signal intensities for each monoclonal are plotted against the signal intensities of the secondary only control. The peptides are colored by the normalized signal intensity.

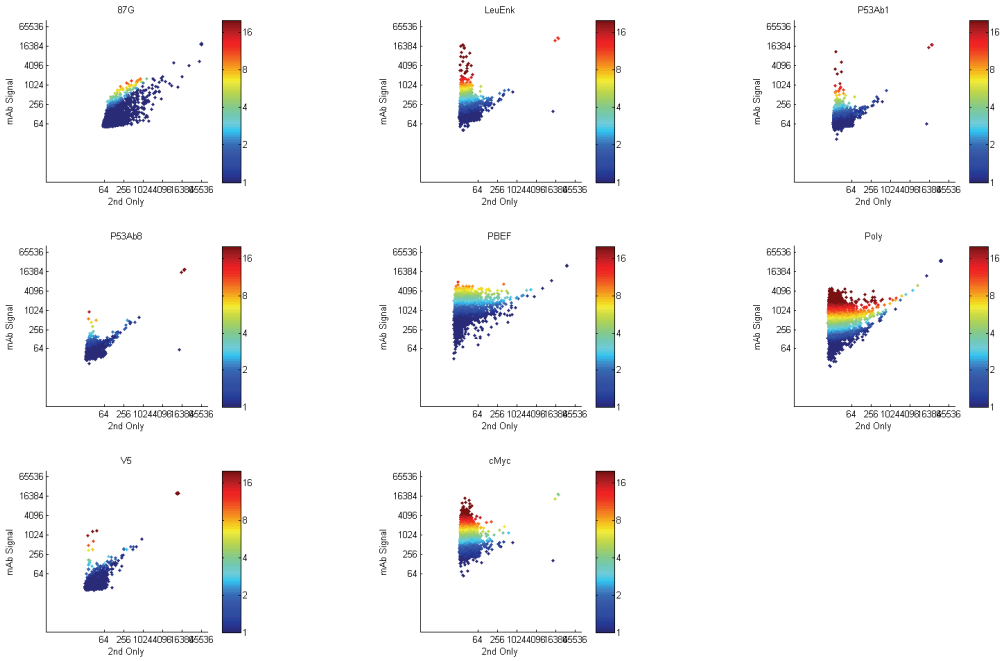


Figure C-5. Monoclonal Signals vs. Secondary Alone Colored by Normalized Signals CIM10Kv2. The signal intensities for each monoclonal are plotted against the signal intensities of the secondary only control. The peptides are colored by the normalized signal intensity.

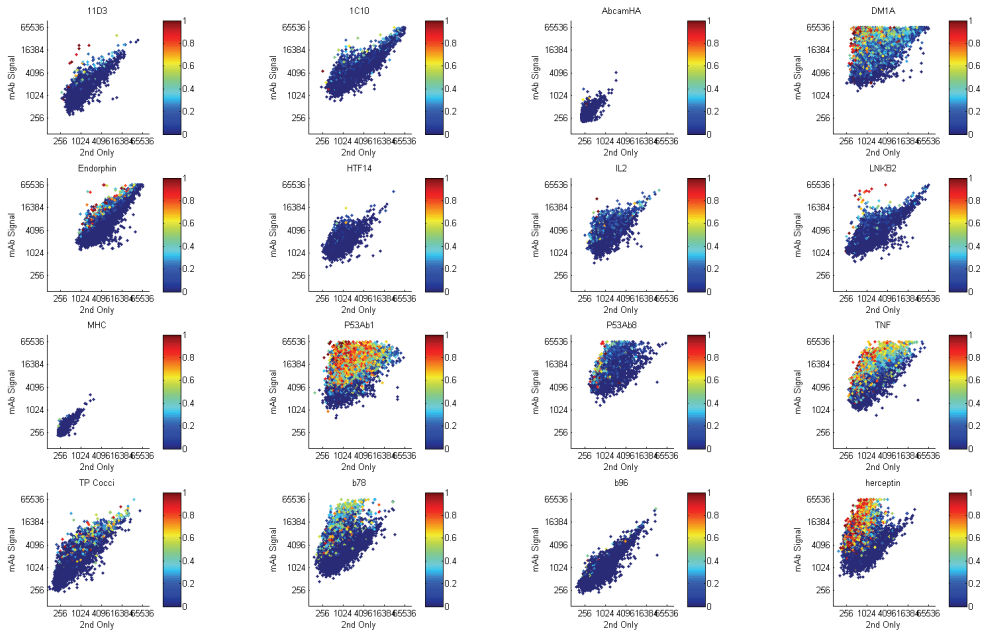


Figure C-6. Monoclonal Signals vs. Secondary Alone colored by Specificity Score CIM10Kv1. The signal intensities for each monoclonal are plotted against the signal intensities of the secondary only control. The peptides are colored by the specificity score.

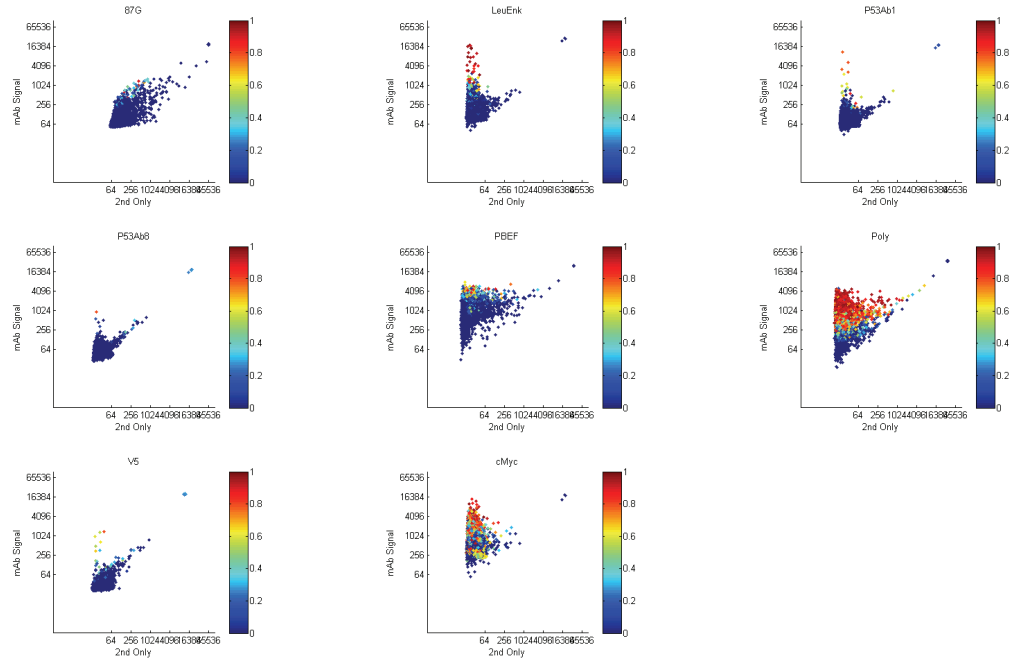


Figure C-7. Monoclonal Signals vs. Secondary Alone colored by Specificity Score 10Kv2. The signal intensities for each monoclonal are plotted against the signal intensities of the secondary only control. The peptides are colored by the specificity score.

APPENDIX D

GENETIC ALGORITHM FOR OPTIMIZATION OF SUBSTITUTION MATRIX FOR EPITOPE MAPPING

While modest prediction accuracy was found for predicting epitopes using the RELIC and GLAM2 programs, the accuracy was not sufficient to predict an epitope from a database (Chapter 5). I identified the substitution matrix as an important parameter for using an alignment approach such as was implemented in RELIC. However, RELIC did not publish their substitution matrix. Clearly, standard substitution matrices such as BLOSUM62 were not ideal for this task, as they were built with the underlying assumption that the amino acid frequencies in the sequences are the same as those occurring in most known proteomes. The peptide libraries I have been working with have equal frequencies of amino acids. Using a standard substitution matrix with these peptide libraries results in giving undue weight to those amino acids that are rare in nature, such as Trp. I have found that when I use the BLOSUM62 matrix to align selected peptides to protein sequence, I almost always find the alignments are anchored by Trp. Another potential problem with using standard substitution matrices is that they are based on the frequency with which amino acids substitutions occur in evolution. The requirements for antibody recognition may be different than the requirements for protein function in general.

I aimed to identify an optimal substitution matrix for epitope mapping from selected random sequence peptides using a genetic algorithm approach. A population of substitution matrices would be created, they would be evaluated for their epitope prediction performance, those that performed well would be mutated and recombined, and those that did not would be eliminated. The performance of the new population would be evaluated and the cycle would be repeated until there was no further improvement in the performance.

The first step was to develop a method to generate substitution matrices. Since a substitution matrix consists of 400 values, independently generating each value did not appear to be a feasible strategy because the search space would be so large. I decided to use amino acid indices to generate the matrices. Amino acid indices were downloaded from the amino acid index database (http://www.genome.jp/aaindex/) and each was transformed to a zero to one scale. To create each scoring matrix, several amino acid indices were randomly selected from the list. For each pair of amino acids the absolute value of the difference between each of the selected index values was found and that value was raised to an exponent. The sum of all of the amino acid indices differences was subtracted from ten. The sum of each row and column of this negative distance matrix were found. Each value was divided by the product of its row and column sum and log transformed to generate log odds like substitution matrix. The substitution matrix could be recombined and mutated by exchanging or varying the amino acid indices, the amino acid index weights, the exponent, or the alignment score cutoff.

The next step was to define the epitope scoring algorithm. First the maximal alignment of each peptide with the protein sequence would be found using the given substitution matrix. A windowing approach like was used in RELIC, was implemented but I allowed the window size to be one of the parameters to evolve. The scores of each peptide alignment at each amino acid position of the protein were added to give protein residue scores. The maximum residue score was used to rank the proteins. The area under the ROC curve was used as the fitness function to evaluate the substitution matrices.

The genetic algorithm was run using the same dataset as was used in Chapter 5 to evaluate Glam2 and RELIC. A leave one out approach was used where the genetic algorithm was trained on nine of the known epitope examples and tested on the example left out. The average training fitness was found to be 0.876, but the average test fitness was found to be only 0.533. I think that this result indicated that the algorithm was over fitting to the nine examples, so it did not have any predictive power on the one left out. Maybe a larger data set would allow this approach to succeed.

APPENDIX E

PEPTIDE QUANTITATION AND NORMALIZATION

Some batch to batch variation exists for the CIM10K arrays. It has also been shown that peptide microarray signal intensities correlate poorly with solution phase affinities (REF). I thought that both of these issues could be related to variability in peptide density. I sought to develop a method to measure the peptide density and an algorithm to normalize for these differences. All of the peptides on all three platforms have free N-terminal amines (though on the 10Kv2 the accessibility may be limited as the peptides are immobilized through the sulfhydryl group of the N-terminal Cys). The amine also an attractive labeling target because there are a number of well characterized and readily available chemical groups that react with primary amines. However, the amine labeling strategy is not ideal because it can also react with primary amines on Lys, which will require an extra normalization step. There are also free amines on the amino silane surface. While most of the amines should have reacted with the sulfo-SMCC linker, that reaction most likely does not go to completion leaving some available amines on the surface. Despite these shortcomings, the amine labeling was the strategy I choose to test.

The first labeling strategy that I attempted utilized NHS-Alexa-555. This strategy proved to have high background, was prone to uneven labeling across the slide surface, and was not able to detect a significant fraction of the peptides. Though I attempted to optimize the reactive dye concentration, reaction time, and pH of the reaction buffer, I was not able to completely eliminate these problems. After labeling a custom array where I had peptides spotted at different concentrations, I noticed that one peptide labeled better at lower concentrations. This peptide was highly negatively charged. Examining the entire 10K, I saw that

there was strong correlation between peptide charge and labeling efficiency. This effect is not surprising because the Alexa-555 dye is highly negatively charge.

Since most dyes are large charged molecules, I decided to try labeling the peptides with amine reactive biotin, and detecting with labeled streptavadin. I tested three different types of reactive biotin, NHS-LC-LC-biotin, TFP-PEG-biotin, and PFP-biotin. I found that the PFP-biotin had the best evenness of labeling across the slide surface and the most signals above background. The PFP biotin showed an increase in signal with peptide concentration, as expected. It also had only a slight correlation with charge which was removed after controlling for the number of Lys, and did not have any strong correlations with any other peptide properties.

After identifying the PFP-biotin as a suitable method to estimate the peptide density, I sought to use that data to normalize immunosignaturing experiments. I tried simply taking a ratio of the sera signal to the biotin signal, but that tending to result in exaggerated normalized signals for peptides that had low biotin signals. The relationship between the serum signals and biotin signals appeared to be non-linear, so I tried using non-linear regression, but that seemed to over correct where there where high biotin signals but low sera signals. I found a manuscript that described a method to normalize protein array data using the signal intensities obtained from detecting a tag on all of the recombinant protein. This method, called CDF, divided the peptides in windows based on the tag signal, and did a z-transformation within each window (REF). I tried this approach on the peptide data and found that it did not work well because the standard deviations varied between the windows, so the z-transformation distorted the distribution.

I sought to develop a method based on the windowing approach used in the CDF method, but that maintained the shape of the distribution. I call the method that I developed WINdowing Quantile normalization (WINQ). First I find the quantiles for the entire raw biotin signals. Then I find the quantiles for all of the peptides that have the same number of Lys. Then interpolate from the quantiles of the Lys groups to the corresponding signal intensity for the overall signals. I call this k-norm data for Lys normalized biotin data. I similarly use the k-norm signals to normalize the sera signals. After finding the quantiles for the raw signal intensities, I find the quantiles for the sera signal intensities for peptides within a window of k-norm values. I use overlapping windows and find the average quantile. I then interpolate from the window quantiles back to the original sera signal distribution.

I used the dataset where the replicates were run across print runs where TCEP was included or left out of the spotting buffer to optimize the parameters in the WINQ normalization. The best parameter set did not result in an improvement in the correlation coefficient but did improve the average CV across replicates. While the improvement in CV was small, it was consistent across several datasets that it was evaluated on.

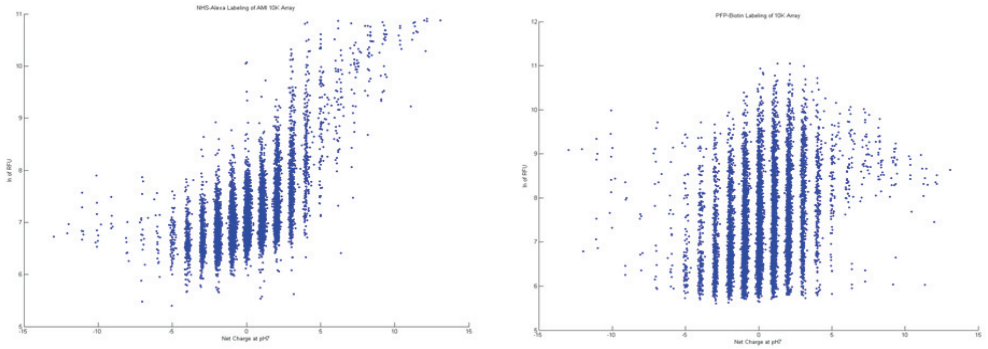
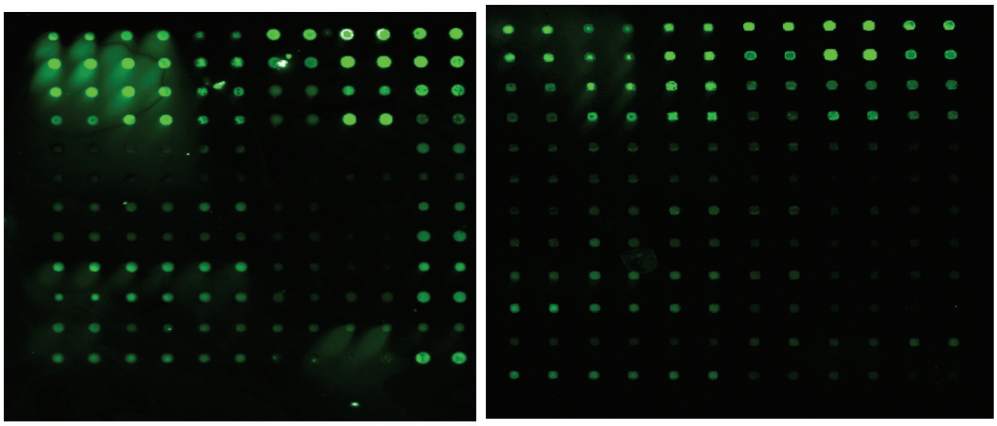


Figure E-1. Peptide Quantitation by Dye vs. Biotin. The peptide quantitation signal intensity is plotted against the peptide net charge at pH7. Direct labeling, shown on the left correlates strongly with peptide charge, while biotin labeling shown on the right, only shows a slight correlation with charge, which can be explained by labeling of Lys.



Biotinylation Subarray 4

Sera Subarray 4

Figure E-2. Images of subarrays for biotin quantitation and sera probing.

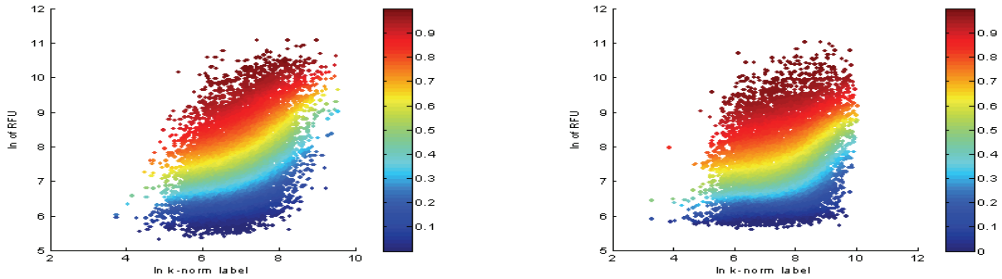


Figure E-3. WINQ (WINdowing Quantile Normalization). The signal intensities for a sera sample are plotted against the label signal that had been normalized for the number of Lys. Peptides are colored by their quantile within a normalized label signal window. The WINQ normalization aims to set the signal distribution to be similar within each window.

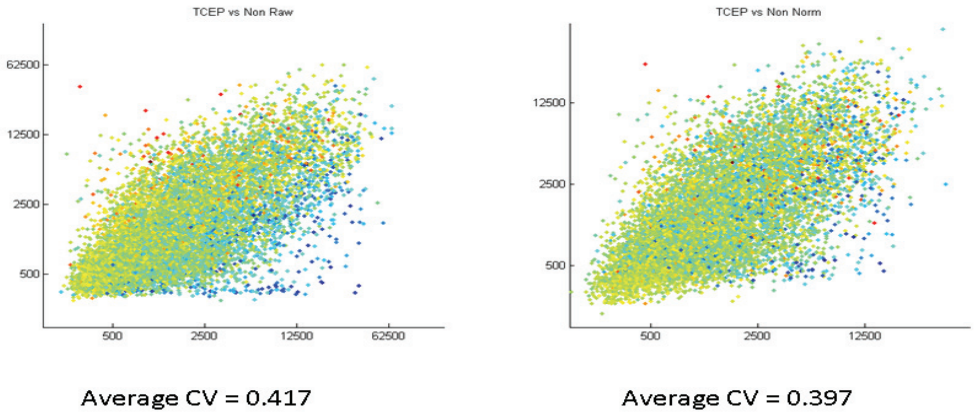


Figure E-4. Normalization Results. The plot on the left shows a scatterplot of the raw signal intensities for two replicate arrays on different print batches. The peptides are colored by the ratio of the biotin labeling signal between the two print batches. The plot on the right shows the WINQ normalized data for the same replicates. There was a small decrease in the average coefficient of variation with the normalization.

APPENDIX F

BIOMICROCHIP ANALYSIS

There have been two versions of the BioMicroChip that have been made by HealthTel so far. The first was a 100K peptides of which each were 10 amino acids long and a total of six different amino acids were used on the chip. We received two wafers of the 10mer6aa chips: the first had been labeled with tamra to measure the stepwise yield (W11), and the second had been labeled with fluorescein (W20). We found that both dyes were still present on the chip surface and appeared to block the binding of antibodies. Nidhi Gupta tested several methods of removing the dyes, but the best methods could only partially remove the dyes. Bart Legutki and I attempted to optimize the antibody binding protocol. While we were able to make some improvements to the dynamic range and reproducibility, there was still quite a bit of unevenness across the chip surface most likely due to residual dye. The reproducibility was not good enough to make biological comparisons.

The second version of the BioMicroChip (12mer8aa) was discussed extensively in chapter 4. Some additional figures are shown here.

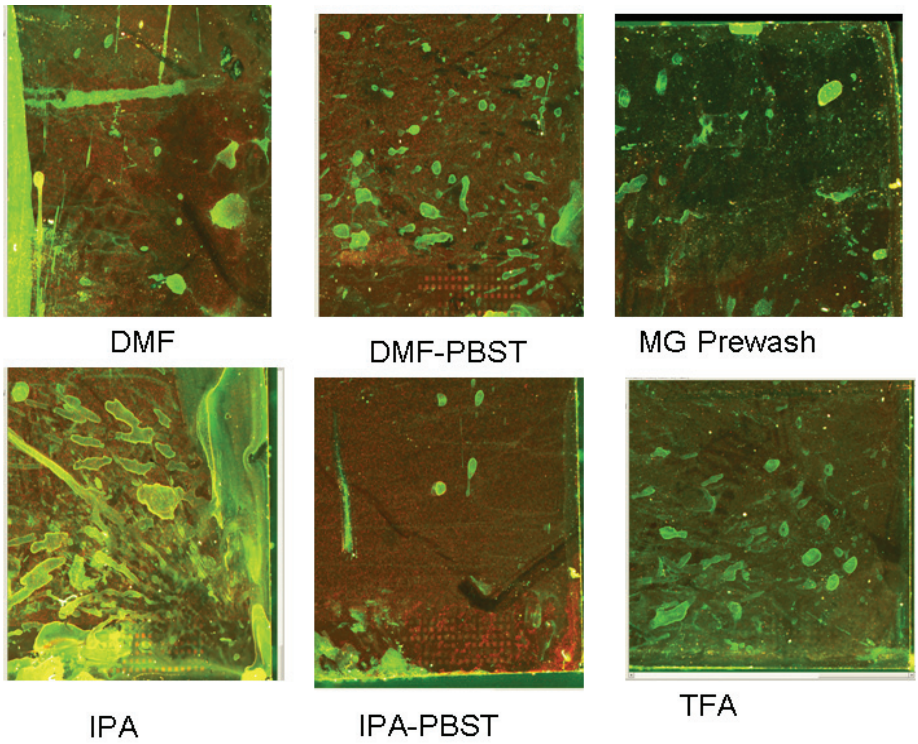


Figure F-1. Pretreatments on 10mer6aa. Six prewash protocols were tested to remove the fluorescein from the wafer 20. After the prewash, all six were probed with sera which was detected on the red channel. The IPA-PBST prewash seems to allow the most sera binding.

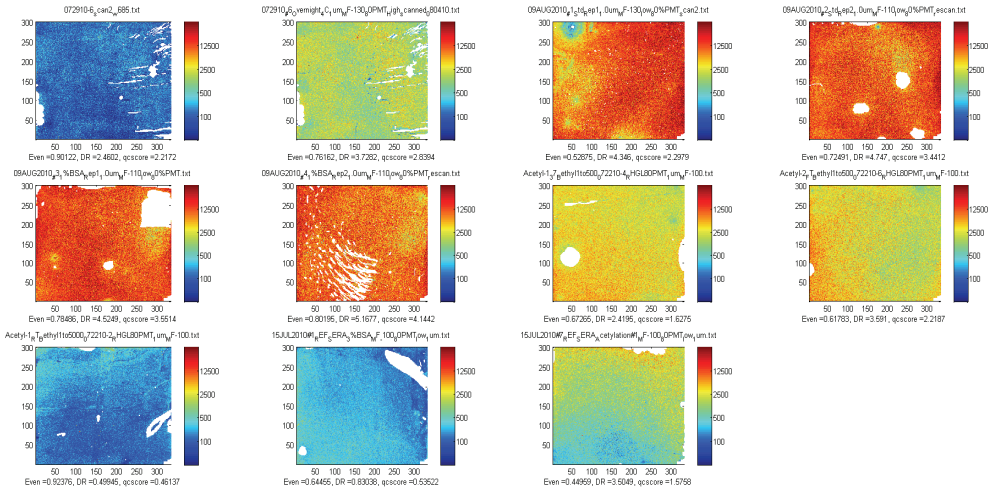
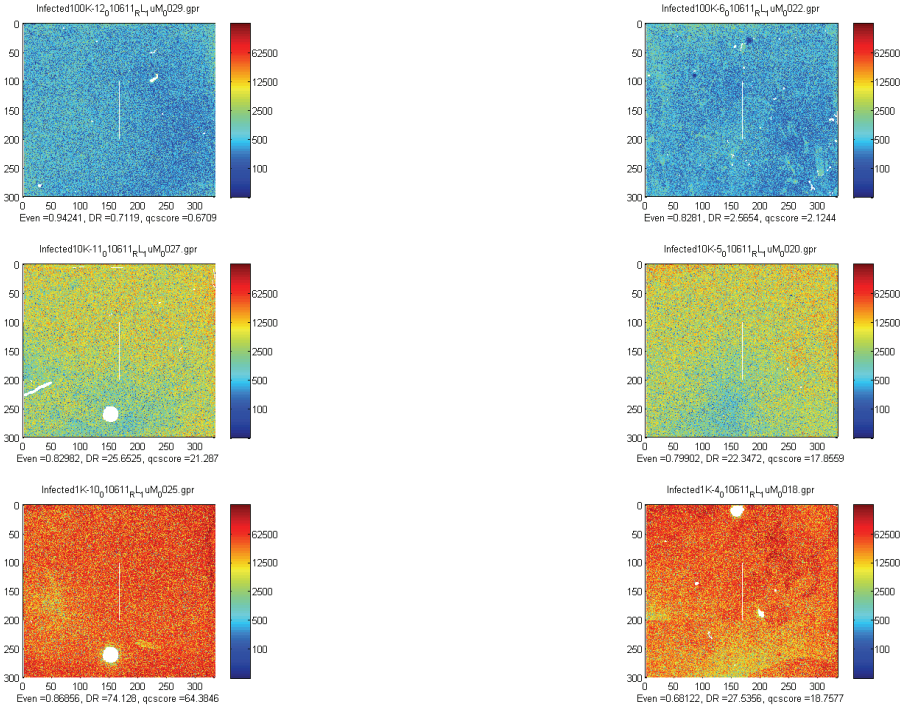


Figure F-2. Assay Optimization on 10mer6aa. Heatmaps of signal intensities spatially arranged as they are laid out on the chip. "Even" is a score indicating how much spatial variability there is on the chip (higher scores indicate less spatial variability), "DR" is dynamic range, and "qcscore" is the product of the evenness score and the dynamic range. The two chips on the left hand side of the second row have the best QC scores. These were acetylated, and included 1% BSA in the wash step



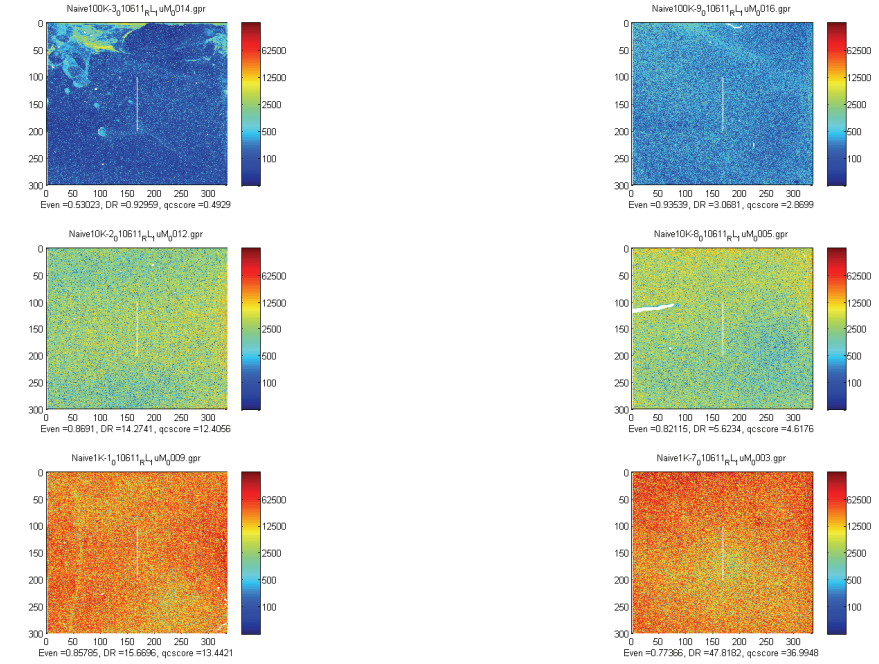


Figure F-3. Dilution Series of Infected vs. Naïve on 12mer8aa. Influenza infected mouse sera or naïve mouse sera was diluted 1:1000, 1:10000, or 1:100000 (labeled 1K, 10K or 100K respectively). The heatmaps above show the signal intensity spatial distributions on all chips.

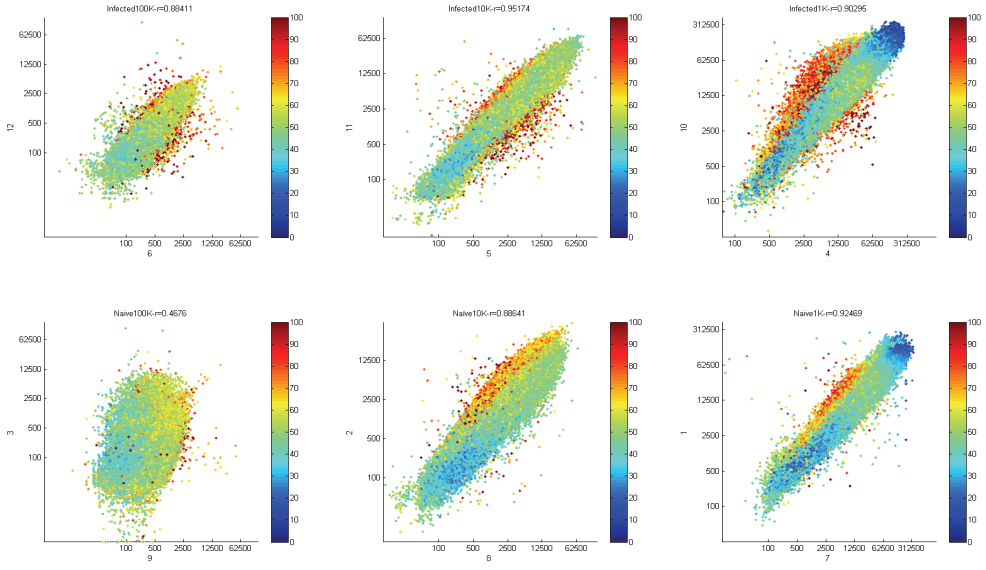


Figure F-4. Infected vs. Naïve Dilution Series Replicate Scatter Plots. Scatter plots of replicate arrays are shown above for the infected and naïve dilution series. Peptides are colored by the coefficient of variation (CV) for the pixels with each feature. Within spot CV appears to be a useful criteria for flagging bad data.

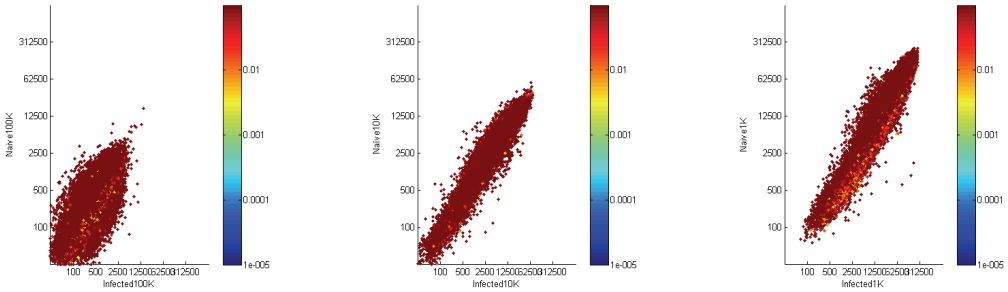


Figure F-5. Infected vs. Naïve Scatter Plots colored by t-test p-value. The signals for the infected sera were plotted against those of the naïve sera at each dilution. Peptides are colored by the t-test p-value between the two conditions. None of the peptides that appear to be different between infected and naïve have significant p-values.

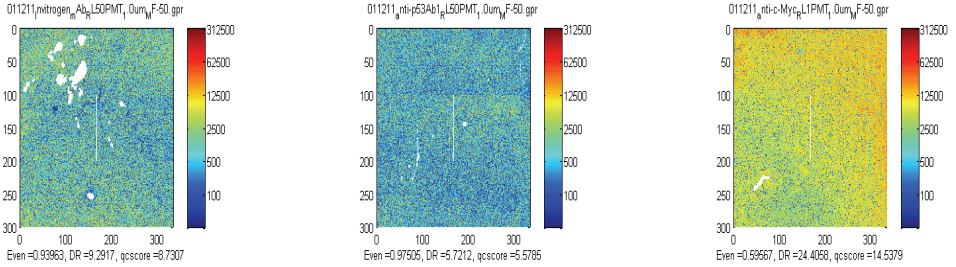


Figure F-6. Monoclonals on 12mer8aa. Heatmaps for the spatial distribution of signal intensities are shown above. Note that the two chips on the left (Invitrogen and P53Ab1) were scanned at 50PMT, but the third chip (cMyc) was scanned at 1PMT and still has higher overall signal intensities.

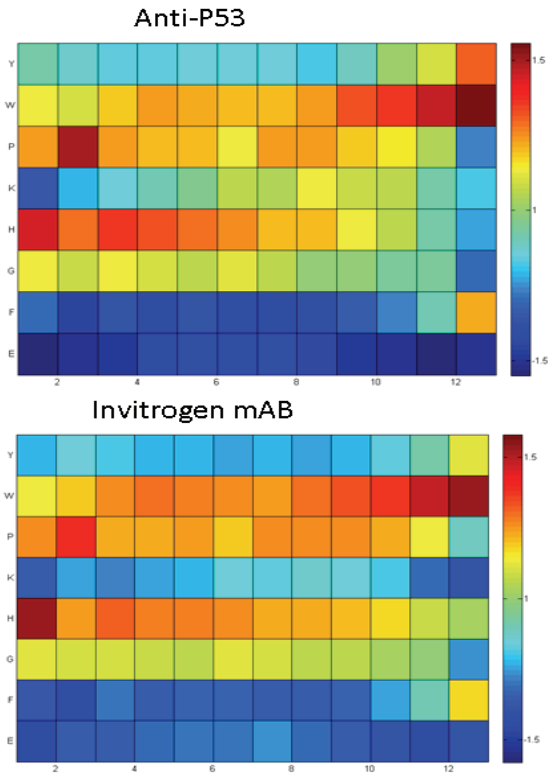


Figure F-7. Amino Acid Positional Bias. The average signal intensity of all of the peptides that have each amino acid at each position are plotted as a heat map above. It appears that it is more favorable to have Trp at the C-terminus and His at the N-terminus. These trends are strikingly similar for both antibodies.

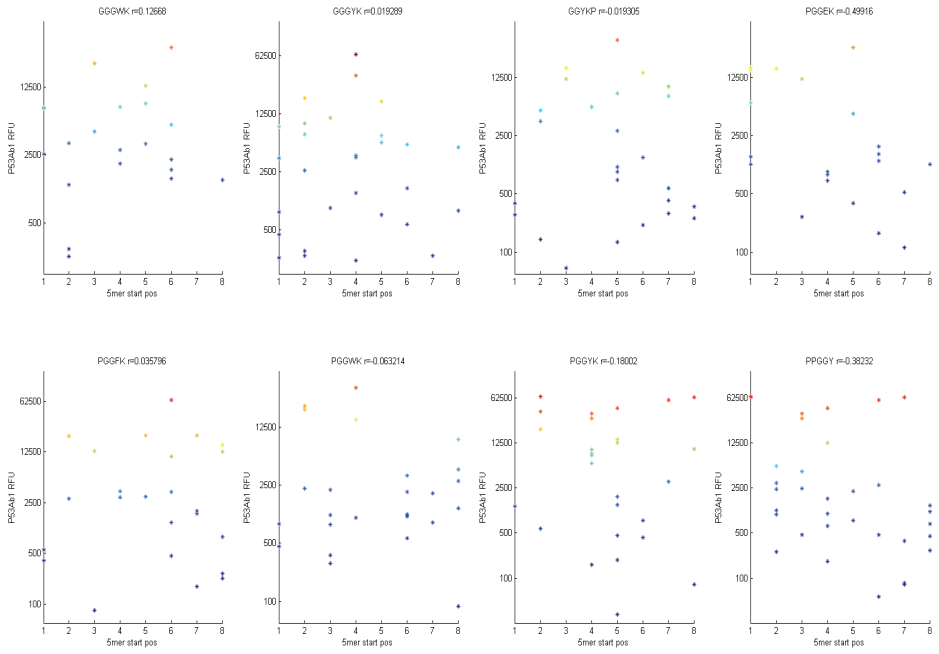


Figure F-8. 5mer Position Analysis. For each of the 5mers identified as binding favorably to P53Ab1 in chapter 4, the position of the 5mer with the peptide was noted. The signal intensity for P53Ab1 binding is plotted against the position of each 5mer. No preference for a particular position was observed.

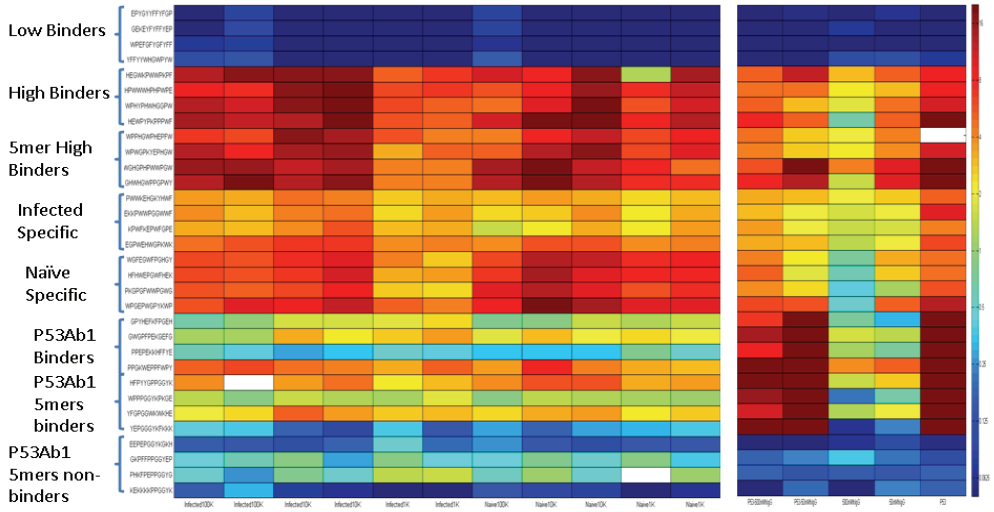


Figure F-9. Peptides Selected for Resynthesis from 12mer8aa. This set of peptides was selected from the 100K based on the criteria listed on the left. The heatmap shows the median normalized signal intensities of these peptides in the experiments listed on the bottom.

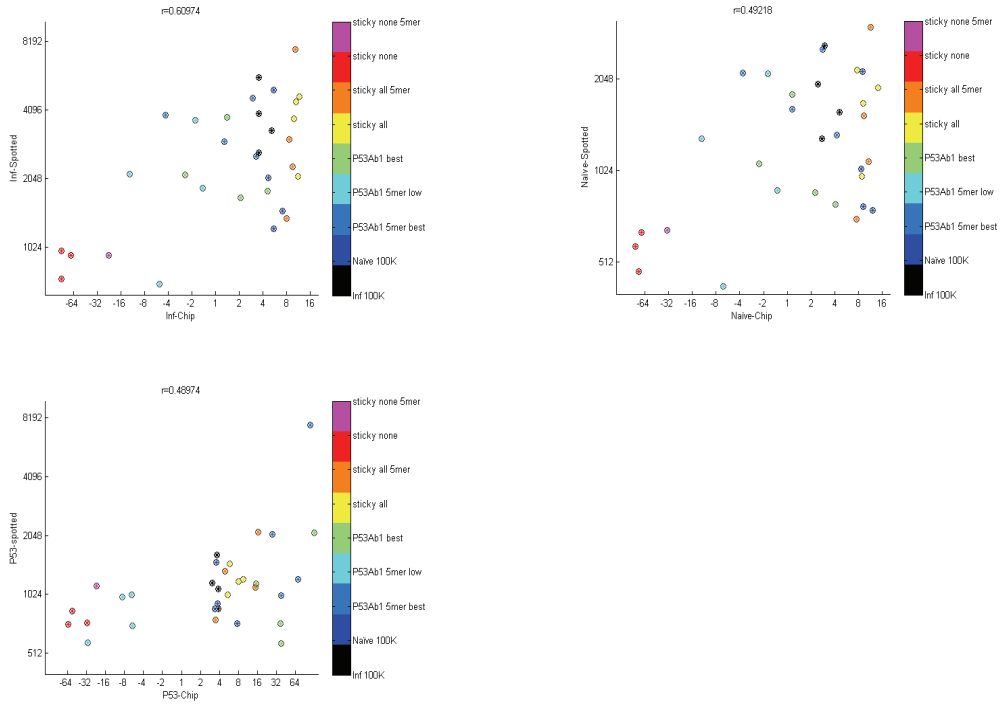
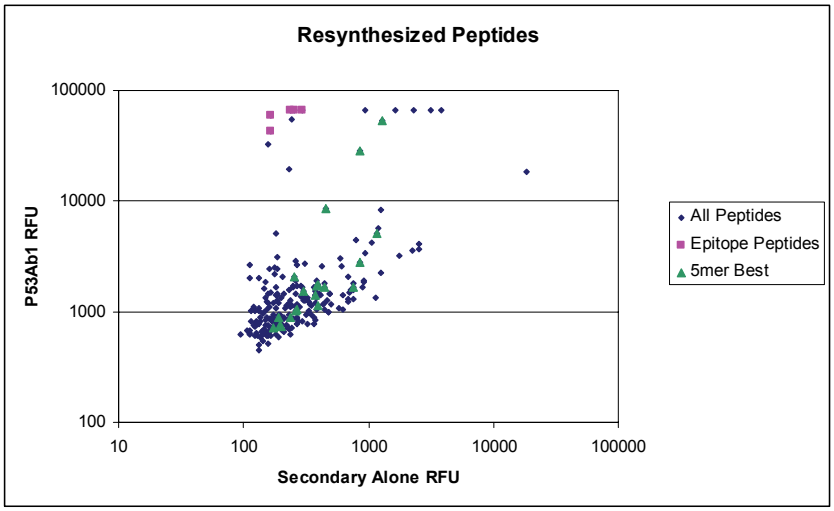


Figure F-10. 100K peptides vs. Resynthesized. The peptides selected in figure 9 were printed in a custom array. Here the normalized signals from the chip are plotted against the signals on the custom array.



FigureF-11. Scatterplot of P53Ab1 vs. Secondary Alone on Resynthesized Peptide Array. Peptides containing the P53Ab1 epitope sequence are shown in pink. The peptides containing the 5mers shown in F-9 are shown in green.

APPENDIX G

VIRUS N-MER ANALYSIS

Table 1. Counts of Sequences in Entrez for Selected Virus. "# Entrez names" indicates the number of virus strains are found in the database. "# Entrez Seq" indicates the number of protein sequences for all strains in the database.

Viruses Pathogen Proteome	**# Entrez Names**	**# Entrez Seq**
Coronavirus	1	30
Dengue	4	43
Ebola	2	17
Herpes Simplex Virus	3	275
HIV	2	34
Human Papillomavirus	28	198
Influenza	3	28
Japanese encephalitis	1	14
Nipah virus	1	8
Norovirus	1	9
Orthopox	1	197
Rotavirus	0	0
Varicella Zoster Virus	1	70
West Nile virus	1	14
Sum	**49**	**937**

Table 2. Counts of Unique 12mers per Virus Strain. The number of 12mers that only occur in each virus strain are shown. 12mers that occur in at least 2 virus strains are in the "common" entry.

Virus	**unique 12mers**
Dengue virus type 3	1
Human herpesvirus 3 (strain Dumas)	1
Human papillomavirus - 10	1
Human papillomavirus - 18	1
Human papillomavirus - 26	1
Human papillomavirus - 41	1
Human papillomavirus - 54	1
Human papillomavirus type 71	1
Human papillomavirus type 90	1
Human papillomavirus type 92	1
Influenza A virus (A-Hong Kong-1073-99(H9N2))	1
Influenza A virus (A-Puerto Rico-8-34(H1N1))	1
Human papillomavirus type 53	1517
Human papillomavirus - 34	1617
Human papillomavirus - 32	1656

Human papillomavirus - 61	1707
Human papillomavirus type 49	1726
Human papillomavirus RTRX7	1729
Human papillomavirus - 2	1768
Human papillomavirus type 48	1769
Human immunodeficiency virus 2	1770
Human papillomavirus type 16	1781
Human papillomavirus type 50	1798
Human papillomavirus - 63	1813
Human papillomavirus - cand96	1816
Human papillomavirus type 6b	1858
Human papillomavirus type 24	1860
Human papillomavirus type 60	1913
Human papillomavirus - 5	1937
Norwalk virus	2000
Human papillomavirus type 7	2093
Human papillomavirus type 9	2134
Human papillomavirus - 1	2176
Dengue virus type 4	2261
West Nile virus	2363
Human papillomavirus - 4	2368
Japanese encephalitis virus	2374
Dengue virus type 1	2678
Human immunodeficiency virus 1	2695
Dengue virus type 2	2871
Reston ebolavirus	3424
Zaire ebolavirus	3483
Influenza B virus	3843
Nipah virus	4672
SARS coronavirus	9285
Influenza A virus	12033
Human herpesvirus 1	27066
Human herpesvirus 2	27368
Human herpesvirus 3	33275
Variola virus	52288
common	**52962**
Human herpesvirus 6	63421
total	**349180**

Table 3. Number of 12mers required to cover selected viruses using a tiling strategy. Sequences of the proteins for these viruses were obtained from Expasy's ViralZone. The sequences were split into 12mers with a tiling step of six. Any 12mer occurring in more than one virus is listed in the "common" row.

virus	**peptide count**
common	*209*
Hepatitis B virus genotype C subtype ayr (HBV-C)	289
Adeno-associated virus 2 (AAV-2)	335
Norwalk virus (strain GI/Human/United States/Norwalk/1968) (Hu/NV/NV/1968/US)	418
Human immunodeficiency virus 1	440
Hepatitis C virus genotype 1a (isolate H) (HCV)	526
Rubella virus (strain Therien) (RUBV)	527
Human respiratory syncytial virus B (strain B1)	741
Influenza A virus (strain A/Puerto Rico/8/1934 H1N1)	742
Influenza B virus (strain B/Lee/1940)	771
Dengue virus type 1 (strain Nauru/West Pac/1974) (DENV-1)	564
Eastern equine encephalitis virus (EEEV) (Eastern equine encephalomyelitis virus)	617
Human hepatitis A virus genotype IB (isolate HM175) (HHAV) (Human hepatitis A virus (isolate Human/Australia/HM175/1976))	370
total	**6549**

APPENDIX H

LUPUS EPITOPE PREDICITIONS

Ab	Accession	ID	Protein Name	Peptide Name	Peptide Seq	Kyte Doolittle Hydrophobicity	Net Charge At pH7	Rank on Guitope Filtered Proteome	Rank on Glam2 and Guitope Combined
D1	Q2VPU4	MLXIP	MLX-interacting protein	MLXIP_248	DDMLYWHKHGDGWKTPVPME	-1.21	-1.86	3	2
D1	P48302	EDNRB	Endothelin B receptor	EDNRB_265	FMQFYKTAKDWWLFSFYFCL	0.36	0.96	2	1
D1	P56475	GBRR1	Gamma-aminobutyric acid receptor subunit rho-1	GBRR1_221	AYTEDDLMLYWKKGNDSLKT	-1.01	-1.09	14	40
D1	Q7TN73	CASD1	CAS1 domain-containing protein 1	CASD1_194	LLEKLAKTSDVYWVLQDPVY	0.06	-1.04	30	6
D1	Q60813	ADM1A	Disintegrin and metalloproteinase domain-containing pro	ADM1A_222	DDLLVLTDWWSHTKYVEMFV	0.12	-2.99	19	4
D1	Q6P5F7	TTYH3	Protein tweety homolog 3	TTYH3_196	EVLAEQVDLYDWYRWLGYLG	-0.16	-3.00	58	7
G4	Q9Z1S8	GAB2	GRB2-associated-binding protein 2	GAB2_15	LRKSPPEKKLRRYAWKKRWF	-1.80	8.00	3	1
G4	Q9R1C6	DGKE	Diacylglycerol kinase epsilon	DGKE_344	LDRWKVQVTNKGYYNLRKPK	-1.44	4.95	9	2
G4	Q9R0N4	SYT10	Synaptotagmin-10	SYT10_474	EGLGRDHWNEMLAYHRKPIT	-1.14	0.14	13	3
G4	Q8C2K1	DEFI6	Differentially expressed in FDCP 6	DEFI6_229	LRRNWAERWFQLQPSSLCYF	-0.63	2.00	30	11
G4	Q8BFU8	VGLU3	Vesicular glutamate transporter 3	VGLU3_4	KAFDTFKEKILKPGKEGVKN	-1.03	2.96	34	7
G4	Q91XQ5	CHSTF	Carbohydrate sulfotransferase 15	CHSTF_281	FSAIKEPHWWTRKRFGIVRL	-0.45	4.05	127	16
D9	P46467	VPS4B	Vacuolar protein sorting-associated protein 4B	VPS4B_31	LQLYQHAVQYFLHVVKYEAQ	-0.07	0.18	82	2
D9	Q6PHN7	TM164	Transmembrane protein 164	TM164_172	ELEIYYIQHAMLYVVPVYLL	1.00	-1.91	41	1
D9	Q9QX11	CYH1	Cytohesin-1	CYH1_119	RDEFSIQVLHAFVELHEFTD	-0.14	-3.86	179	5
D9	O88853	GALR3	Galanin receptor type 3	GALR3_248	WGPHHALILCFWYGRFAFSP	0.39	1.18	179	4
D9	Q8BLD9	DRD5	D(1B) dopamine receptor	DRD5_135	DRYWAISRPFRYERKMTQRV	-1.42	3.96	182	10
D9	Q8K097	FAIM2	Fas apoptotic inhibitory molecule 2	FAIM2_242	LAVLLPFQVVPWLHAVYAVL	1.53	0.09	193	6
F9	Q6ZQ29	TAOK2	Serine/threonine-protein kinase TAO2	TAOK2_272	SEVLLKHRFVLRERPPTVIM	0.05	2.05	14	2
F9	Q01279	EGFR	Epidermal growth factor receptor	EGFR_874	EGGKVPIKWMALESILHRIY	0.06	1.09	26	6
F9	Q01098	NMDE3	Glutamate [NMDA] receptor subunit epsilon-3	NMDE3_761	IAMQKDSHWKRAIDLALLQF	-0.04	1.09	22	14
F9	Q99P88	NU155	Nuclear pore complex protein Nup155	NU155_647	ASMSGLTGPEIVYSGKHNGI	-0.05	0.05	57	1
F9	Q9D2G5	SYNJ2	Synaptojanin-2	SYNJ2_799	AWTDRVLWWRKKHPYDKTAG	-1.34	3.00	73	16
F9	Q61137	ASTN1	Astrotactin-1	ASTN1_240	DGYEYDITDLRHHLQRECMN	-1.36	-2.86	133	3
G10	Q68FE2	ATG9A	Autophagy-related protein 9A	ATG9A_150	FIYNICCYWEIHSFYLHALR	0.50	0.18	4	5
G10	Q8C031	LRC4C	Leucine-rich repeat-containing protein 4C	LRC4C_305	CNCDILWLSWWIRDMAPSNT	0.01	-1.04	22	8
G10	P53690	MMP14	Matrix metalloproteinase-14	MMP14_472	EVFTYFYKGNKYWKFNNQKL	-1.13	2.96	16	2
G10	Q812G0	MGT4A	Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosa	MGT4A_233	RWRTKQNLDYCFLMMYAQEK	-1.00	1.96	40	4
G10	Q80VA0	GALT7	N-acetylgalactosaminyltransferase 7	GALT7_468	VVEVWWDEYKDYFYASRPES	-0.88	-2.99	40	4
G10	P39688	FYN	Tyrosine-protein kinase Fyn	FYN_283	GEVWLGTWNGNTKVAIKTLK	-0.25	1.87	456	6

Figure H-1. Peptides Selected by GuiTope and Glam2 Analysis for Lupus Monoclonals. Stephanie Williams (Steven Hoffman's lab) created five monoclonal antibodies from a lupus mouse model, and selected them for reactivity against the membrane fraction of brain lysate. She found the molecular weight of the monoclonal targets by western blotting, and used them to probe the CIM10Kv2 array. Specific peptides were selected for each. I used in GuiTope to search against membrane proteins in the mouse proteome that had the corresponding molecular weight. I also identified motifs using glam2 and used them to search the filtered proteome using glam2scan. The predicted epitopes above are based on the averaged GuiTope and glam2 scores. Those known not to be expressed in the brain were excluded. Those highlighted in green were identified by Stephanie as having biological connection to lupus.

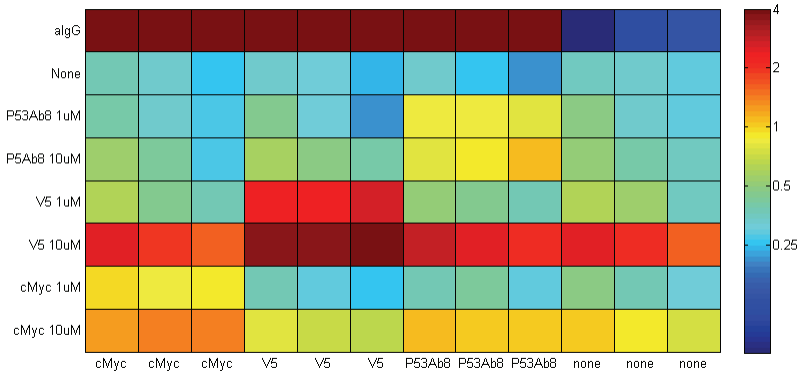


Figure H-2. Control Peptides ELISA. The plate was coated with the antibodies indicated at the bottom of the columns. After blocking, it was probed with the biotinylated peptide corresponding to the cognate epitope sequence as labeled in the rows. Binding was detected with HRP-streptavadin. The top row was probed with HRP labeled anti-IgG to detect antibody coating. At high peptide concentrations, two of the peptides bind to the uncoated wells, but at the low peptide concentration, only the cognate interaction is observed.

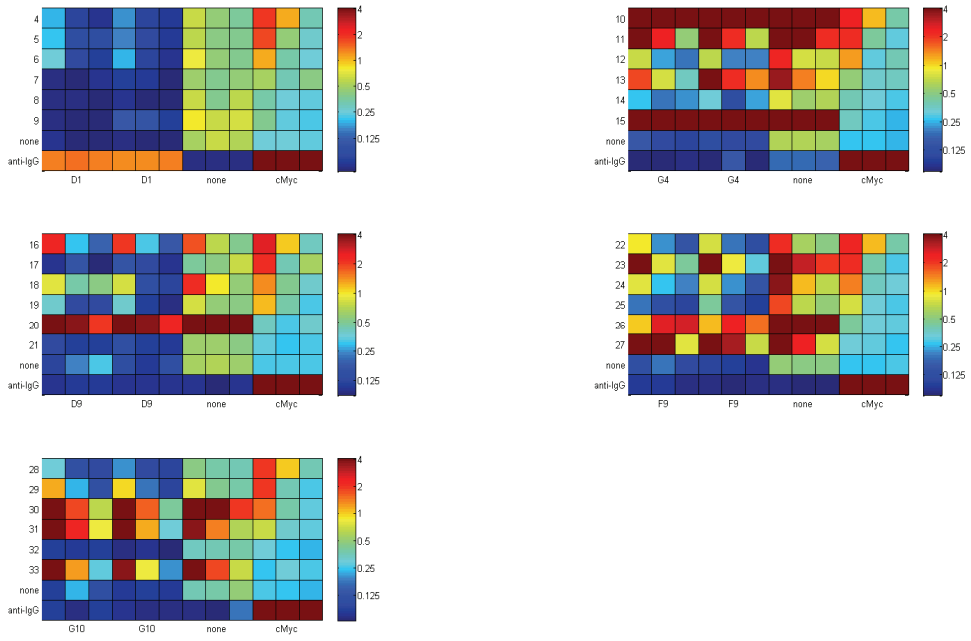
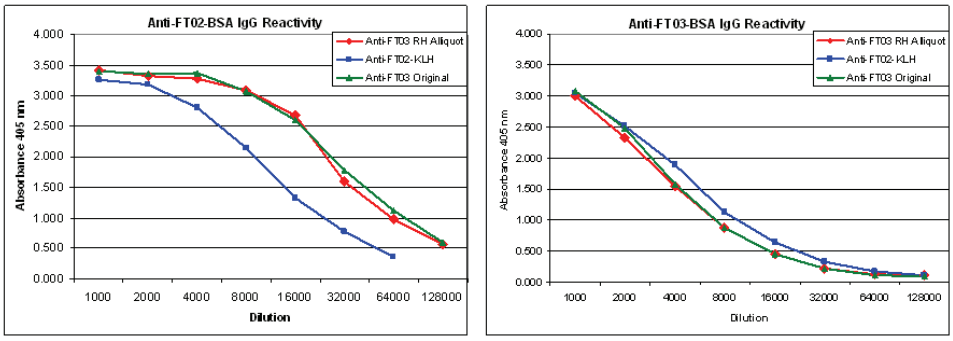


Figure H-3. Lupus Peptides ELISA. The ELISA was performed as described for figure 2, except peptides were diluted 1:50 in columns 1, 4, and 7, 1:500 in 2, 5, and 8, and 1:5000 in 3, 6, and 9. Columns 10, 11, and 12 contain serial dilutions of the cMyc, P53Ab8, and V5 peptides respectively. Antibody coating was only detected for D1. A sandwich ELISA later confirmed that no detectable IgG was present for the other four antibodies. Peptides all bound better to the uncoated wells than to the antibody coated wells.

APPENDIX I

DIPEPTIDE INVERSIONS



FT02 -ialFDp**tk**wpehhqyfagsc
FT03 kanwFDf**kt**fnqmtqvwgsc-

Figure I-1. Crossreactivity of anti-Peptide Sera with Dipeptide Inversion. Mouse sera collected after immunization with the FT02 or FT03 peptides shown above as KLH conjugates was used in ELISA to check reactivity to both peptides. Sera from both peptide immunizations reacted similar with both peptides. A region of similar sequence (underlined) was identified that contained a dipeptide inversion, shown in bold. This data was provided by Bart Legutki.

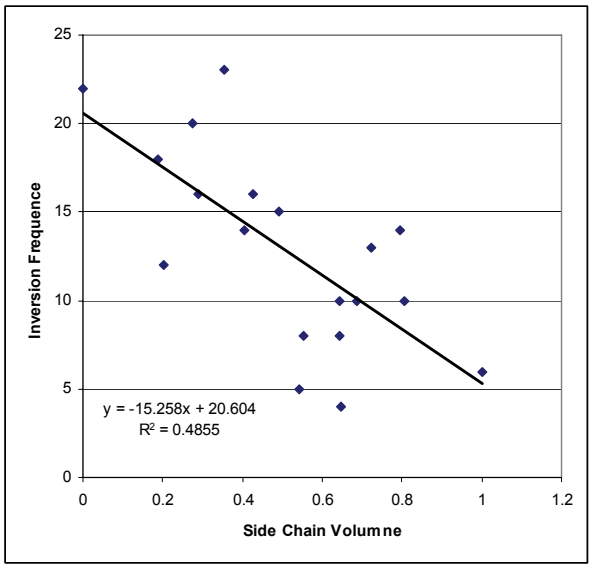


Figure I-2. Side Chain Volume vs. Inversion Frequency. Random-sequence peptides found to separate Valley Fever from normals were aligned to Valley Fever proteins in GuiTope using default paramaters including dipeptide inversion weight as one. The number of times each amino acid was found in a dipeptide inversion is plotted against the residues side chain volume. A line was fit to the data and the equation and RSQ value are shown on the plot.

APPENDIX J

CITATIONS AND PERMISSIONS

Chapter 2
This research was submitted to Molecular and Cellular Proteomics on Sept. 18, 2011.
Phillip Stafford, Rebecca F. Halperin, J. Bart Legutki, D. Mitch Magee, John Galgiani and Stephen Albert Johnston. Antibodies as Biomarkers of Health Status.
All co-authors have granted permission for the inclusion of this work in this dissertation.

Chapter 3
This research is included in a manuscript in preparation.
Rebecca F. Halperin, Phillip Stafford, Ying Xiong, Abner L. Notkins, and Stephen Albert Johnston. A Polyreactive Antibody Recognizes a Large Proportion of both Random and Protein Derived Peptide Sequences.
All co-authors have granted permission for the inclusion of this work in this dissertation.

Chapter 5
This research was origninally pubilished in Molecular and Cellular Proteomics.
Rebecca F. Halperin, Phillip Stafford, and Stephen Albert Johnston. Exploring antibody recognition of sequence space through random-sequence peptide microarrays. *Molecular and Cellular Proteomics*. 2011; 10(3):M110.000786.

All co-authors have granted permission for the inclusion of this work in this dissertation.

Chapter 6
This research was submitted to BMC Bioinformatics on Sept. 19, 2011
Rebecca F. Halperin, Phillip Stafford, Jack S. Emery, Krupa Arun Navalkar, and Stephen Albert Johnston. GuiTope: An Application for Mapping Random-Sequence Peptides to Protein Sequences.
All co-authors have granted permission for the inclusion of this work in this dissertation.