Exploring the Nature of Protein-Peptide Interactions on Surfaces

by

Wei Wang

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

Approved June 2014 by the
Graduate Supervisory Committee:

Neal Woodbury, Chair
Yan Liu
John Chaput

ARIZONA STATE UNIVERSITY

August 2014

ABSTRACT

Protein-surface interactions, no matter structured or unstructured, are important in both biological and man-made systems. Unstructured interactions are more difficult to study with conventional techniques due to the lack of a specific binding structure. In this dissertation, a novel approach is employed to study the unstructured interactions between proteins and heterogonous surfaces, by looking at a large number of different binding partners at surfaces and using the binding information to understand the chemistry of binding. In this regard, surface-bound peptide arrays are used as a model for the study. Specifically, in Chapter 2, the effects of charge, hydrophobicity and length of surface-bound peptides on binding affinity for specific globular proteins (β-galactosidase and α1-antitrypsin) and relative binding of different proteins were examined with LC Sciences peptide array platform. While the general charge and hydrophobicity of the peptides are certainly important, more surprising is that β-galactosidase affinity for the surface does not simply increase with the length of the peptide. Another interesting observation that leads to the next part of the study is that even very short surface-bound peptides can have both strong and selective interactions with proteins. Hence, in Chapter 3, selected tetrapeptide sequences with known binding characteristics to β-galactosidase are used as building blocks to create longer sequences to see if the binding function can be added together. The conclusion is that while adding two component sequences together can either greatly increase or decrease overall binding and specificity, the contribution to the binding affinity and specificity of the individual binding components is strongly dependent on

their position in the peptide. Finally, in Chapter 4, another array platform is utilized to overcome the limitations associated with LC Sciences. It is found that effects of peptide sequence properties on IgG binding with HealthTell array are quiet similar to what was observed with β-galactosidase on LC Science array surface. In summary, the approach presented in this dissertation can provide binding information for both structured and unstructured interactions taking place at complex surfaces and has the potential to help develop surfaces covered with specific short peptide sequences with relatively specific protein interaction profiles.

DEDICATION

To my parents, Lianyuan Wang and Lilian Song, for raising, advising, encouraging and loving me

To my husband, Franklin Liu, for his unconditional love and support

To my son, Leon, for bringing joy to my life

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Neal Woodbury, for his guidance, support and constant encouragement through my entire Ph.D. study. Dr. Woodbury always encourages me to think independently, to design experiments and to arrive at new concepts on my own. It was quite difficult for me at the very beginning but Dr. Woodbury was always there to help me with great patience. He has provided direction and support for almost everything I have done in the past 6 years. In addition, because of his understanding and supports, I had the chance to participate in many extracurricular activities to explore different potential career paths beyond academia.

I would like to thank the rest of my committee, Dr. John Chaput and Dr. Yan Liu, for their mentorship. Dr. Chaput and Dr. Liu spent considerable time with me working on my presentation and conceptualization skills. They also provided important feedback to my dissertation.

I also need to thank Drs. Phillip Stafford, Jinglin Fu, Zhan-Gong Zhao and Matt Greving for helpful discussions on data analysis and experiment design; Dr. Su Lin, Dr. Zhi Guo, Ms. Carole Flores and Ms. Rui Liu for helping me and encouraging me whenever I have problems in my study or life. I would also like to thank all my colleagues and friends that supported me through this experience: Pallav Kumar, Jie Pan, Albert Pedroza, Anil Prasad, Minghui Liu, Xixi Wei, Bing Jiang, Hu Duan and all my current and former colleagues in the Woodbury Lab. I also owe many thanks to the people in the Center for Innovations in Medicine: Dr. Stephen Johnston; Dr. Rebecca Halperin, Dr. Krupa Navalkar and Dr. Bart

Legutki for help in microarray assays and data analysis; Pattie Madjidi and Penny Gwynne for helping to keep the lab running. Finally, thanks to the NSF and the Biodesign Institute for the financial support on the projects in my dissertation.

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

CHAPTER 1: GENERAL INTRODUCTION AND BACKGROUND INFORMATION

Protein-Surface Interactions

Protein-surface interactions are important in both biological and artificial systems[1]. Biological systems are organized by complex networks of molecular interactions with dynamic behavior, and proteins are one of the key components in such networks[2, 3]. The cell surface is one of the most common interfaces in biological systems. Some examples of biological processes involving proteins interacting with molecular components on cell surfaces include cell adhesion[4], cell-cell interactions[5], hormone cellular signaling pathways[6] and information/nutrient exchange. Artificial systems are usually designed and synthesized to mimic the functional properties of biological systems to achieve similar or even better results with manufacturing advantages. Interactions between proteins and synthetic material surfaces are important in developing and manipulating artificial systems[7]. For example, medical implants and artificial tissues developed to help patients regain body function are exposed to proteins and interact with them to trigger defined biological reactions[8]; biomolecular functional units such as enzymes, receptors and antibodies are utilized in various types of biosensor and bioelectronics to increase device performance[9]; DNA, peptide and protein based diagnostic and sensor arrays that incorporate protein functions are used clinically and/or at the research level in biomedical industries and academia[10, 11].

Specific versus Nonspecific Interactions

The interactions governing biological processes can be sorted into two general categories: specific interactions and nonspecific interactions. Specific interactions are

normally associated with well-defined recognition/binding structures and high free energy changes upon binding[12]. Some examples of specific interactions in nature include DNA-protein bindings, antigen-antibody interactions, protein-ligand bindings and enzyme-substrates complexes. Specific interactions are usually well-studied and more thoroughly researched in comparison to nonspecific interactions. In contrast, nonspecific interactions generally lack a unique structure for the final complex formed, and for this reason, they may also be referred to as unstructured interactions. Although the term "nonspecific" may refer to a weak interaction that is universal to all interacting molecules, in some cases, the situation is not that simple. Some nonspecific interactions can greatly favor some protein interactions over others and they can contribute significantly to the overall binding affinity. In fact in many cases, many kinds of interactions can take place at the complex binding surfaces of proteins and the strength of an interaction largely comes from nonspecific interactions such as general charge-charge interactions. The nonspecific interactions result in a set of shallow potential wells on the potential energy surface with multiple possible unstable interacting complexes. What makes a reaction specific is that a certain metastable structure could be formed by aid of specific short-range forces, which can make one potential well deep enough so that it is the one predominantly occupied. Hence, the concepts of specific and nonspecific interactions are better thought of as two ends of a more continuous spectrum.

Importance of Nonspecific Interactions

Nonspecific interactions, or unstructured interactions, are driven by some combination of general weak interactions including electrostatic interactions, hydrophobic interactions (entropic effects) and hydrogen bonding interactions (dipolar

interactions). They are themselves the sum of a variety interactions that do not stabilize one particular structure. In the crowded cellular environment, the majority of interactions proteins may have with each other or with the surfaces of cells are nonspecific. If one only considers protein-protein interactions, roughly, for a cell with N proteins, there are on the order of N specific protein-protein interactions, but almost N^2 nonspecific protein-protein interactions[13]. Although these nonspecific interactions are ubiquitous in biology, they are generally poorly studied in terms of their role in biology and thus there is a lack of fundamental understanding of their chemical and structural nature. Many biological processes cannot function properly without the participation of nonspecific interactions. In fact, in many cases, the distinction between "specific" and "nonspecific" interactions is not clear cut. For example, the interaction between a DNA-binding protein (i.e. RNA polymerase, endonuclease and methyltransferase) and its target DNA is initially a nonspecific (or sequence-independent) interaction at a random DNA site. This interaction helps the DNA-binding protein find its specific binding site rapidly and efficiently by "facilitated diffusion". It can then initiate sequence-dependent interactions at the correct binding location. The initial interaction is considered as a nonspecific interaction because it is not associated with a particular DNA sequence; however, it is quiet specific to double stranded, B-form DNA in general and is an important intermediate step in the process of sequence-dependent DNA recognition and binding[14, 15].

In an extraordinary crowded cellular environment, it is very important that a protein not only specifically bind to its target molecules, but also selectively not adsorb to other non-target molecules. The failure of protein to maintain a strong specificity to its

functional partners can result in protein misfolding and aggregation, which may lead to a variety of diseases including neurodegenerative diseases such as Parkinson's disease[16] and Alzheimer's disease[13, 17]. In this regard, a better understanding of how proteins interact with non-target molecules is the key to understanding how proteins remain stable and avoid aggregation in complex environment.

Nonspecific or unstructured interactions between proteins and complex surfaces are of particular significance to our understanding of fundamental biology and development of related applications. A protein interacting with surface bound molecules differs from its interaction with its binding partner in solution, where one protein only interacts with one binding partner. Due to the high local concentration, a protein may interact with multiple molecules or parts of a molecule on a surface. Thus, the interaction between a protein and surface bound molecules can be considered as one surface interacting with another. Nature uses selective, unstructured interactions between protein surfaces and complex surfaces in their surroundings to its advantage and mimicking that kind of selectivity is important in developing biomaterials or biosensors with surfaces that avoid background binding and interference, but facilitate selected interactions[18, 19]. Another application of utilizing unstructured interactions between proteins and complex surfaces is the immunosignature technology developed in the Center for Innovation in Medicine in the Biodesign Institute at Arizona State University. In that case, a surface covered with relatively short peptides is used to discriminate between the binding profile of antibodies in the serum of healthy individuals and people with specific diseases. The general characteristics of a surface can result in very large discrimination between binding partners, even without forming specific binding structures[20].

Analytical Difficulties in Characterizing Nonspecific Interactions

A large number of methods have been developed to study interactions between proteins and their binding partners, either in solution or at surface [21-25]. Table 1.1 shows selected examples of different genetic, biochemical and biophysical methods for analyzing protein interactions. However, most methods focus on identifying and/or characterizing specific interactions rather than nonspecific interactions. In fact, most nonspecific interactions are generally considered as noise in experiments. The reason they are a problem in terms of noise is that they are not that low in total. The issue is that particular structures are formed with low affinity, but there are many possible structures. Most techniques listed below in Table 1.1, including crystallography, immunoprecipitation, phage display and affinity chromatography, force the system to search for the most favored single structure, which is defined by specific interactions. Techniques like NMR only show the specific interactions because the rest average out to zero given the highly specific signals. For either technique used, a very nonlinear process that strongly favors one specific interaction is imposed during the study.

Table 1.1. Selected Examples of Different Methods Studying Protein Interactions [21-25]

Method	**Description**	**Living Cell Assay**
Co-immunoprecipitation	Identify protein-protein interactions with a specific antibody	*In-vitro*

Affinity purification–MS	Isolate interacting protein targets with affinity chromatography and identify with MS analysis subsequently	*In-vitro*
Yeast two hybrid system	Screen interactions between fusion proteins inside the nucleus of yeast and identify binding partners of a protein	*In-vivo*
DNA/protein/peptide microarrays	Screen and identify interactions between molecules	*In-vitro*
Phage display, RNA display	High-throughput screening of protein interactions	*In-vitro*
Synthetic lethality	Genetic method for verify protein interactions	*In-vivo*
X-ray crystallography, NMR spectroscopy	Structural study for interactions with specific binding structure	*In-vitro*
Fluorescence resonance energy transfer (FRET)	Biological characterization for interactions between single molecules	*In-vivo*
Surface plasmon resonance (SPR)	Characterize kinetic constant for binding processes	*In-vitro*
Electron microscopy	Structural and biological characterization	*In-vitro*

Moreover, there are many techniques intentionally designed to eliminate nonspecific interactions. Examples include affinity purification methods such as tandem affinity purification (TAP) and “large library” selection approaches such as phage

display. Affinity purification methods are generally biased toward proteins with high affinity and slow kinetics of dissociation[22]. TAP method is a type of affinity purification using a TAP tag and two purification steps (Figure 1.1)[26, 27]. In combination with mass-spectroscopy, it can identify interacting partners with high affinity[27]. Weakly interacting partners without a stable binding structure are easily lost during the series of purification steps[22, 28] and thus it is not suitable for identifying nonspecific/unstructured binding partners.

"Large library" selection approaches are widely used for the selection of affinity reagents. Figure 1.2 shows the general scheme for *in vitro* biological selection approaches. Just like the natural selection process, a selection pressure is applied to the large library of random molecules. Molecules that successfully pass the selection process will be amplified for further analysis or served as the library for the next selection cycle and molecules failed to pass will be left out. Many approaches can be used to generate the desired libraries with capability of selection and amplification. For example, a random double stranded DNA pool can be directly synthesized with a DNA synthesizer and amplified through PCR; a single stranded DNA/RNA pool can be separated and transcribed from a double stranded DNA pool and amplified through reverse transcription and PCR[29]; peptide and proteins can be linked to their coding sequences and amplified through these sequences by molecular display techniques such as phage display, mRNA display and ribosome display[30-34]. These types of approaches may be great for identifying affinity reagents but they cannot serve as a tool to systematically study nonspecific interactions because the identity of the molecules in the library remain unknown until the end of the selection cycles and information about moderate or weak

interactions, which is critical for understanding general rules governing nonspecific interactions, is permanently lost. The use of chemical selection systems is another way to create large libraries for selection of desired molecules, especially for peptides[33]. The difference between chemical selection and in vitro biological selection is that the component cannot be easily amplified and reselected after the selection. Although chemical selection systems are very effective in creating large libraries that contain large number of nonspecific/unstructured binding candidates, only a few candidates survive the screening protocol and reach the final characterization step, which is required to obtain the candidate identity.

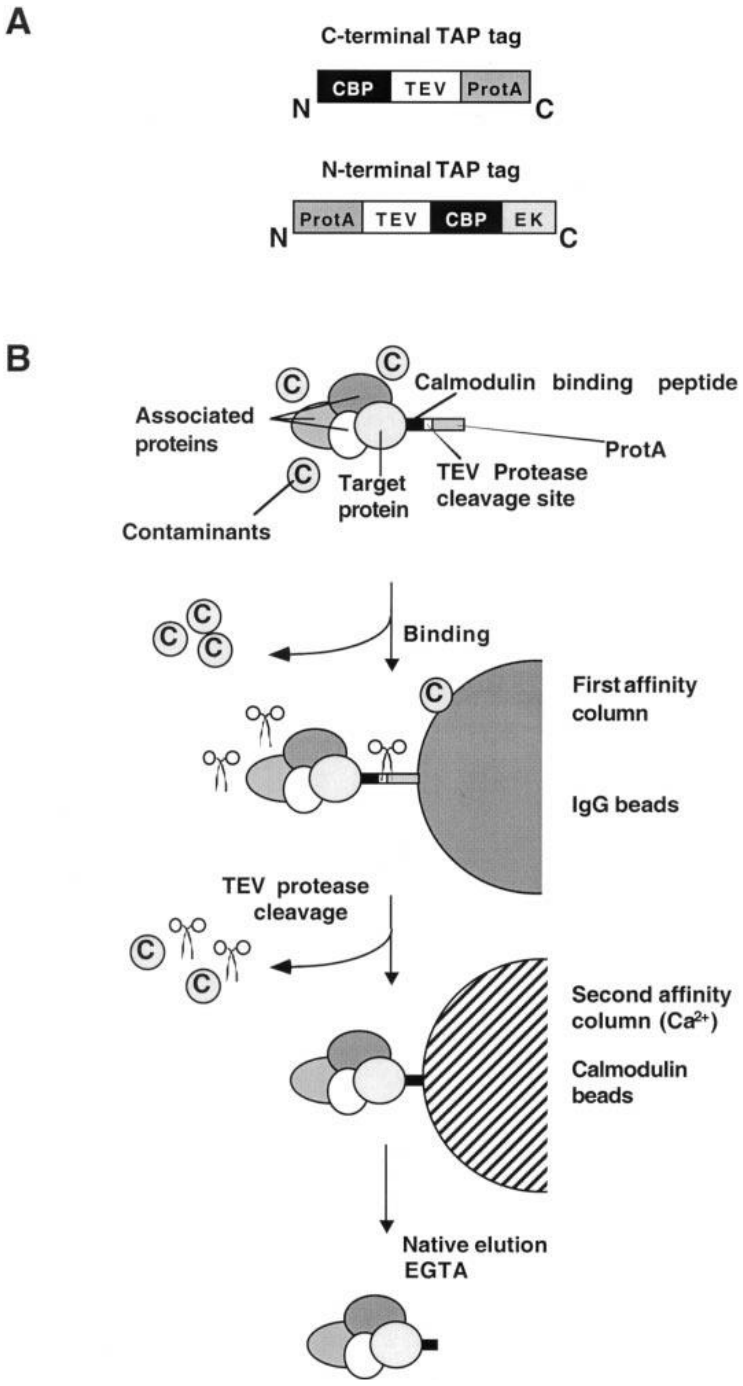

Figure 1.1. Schematic representation of TAP tags and the TAP purification strategy[27].

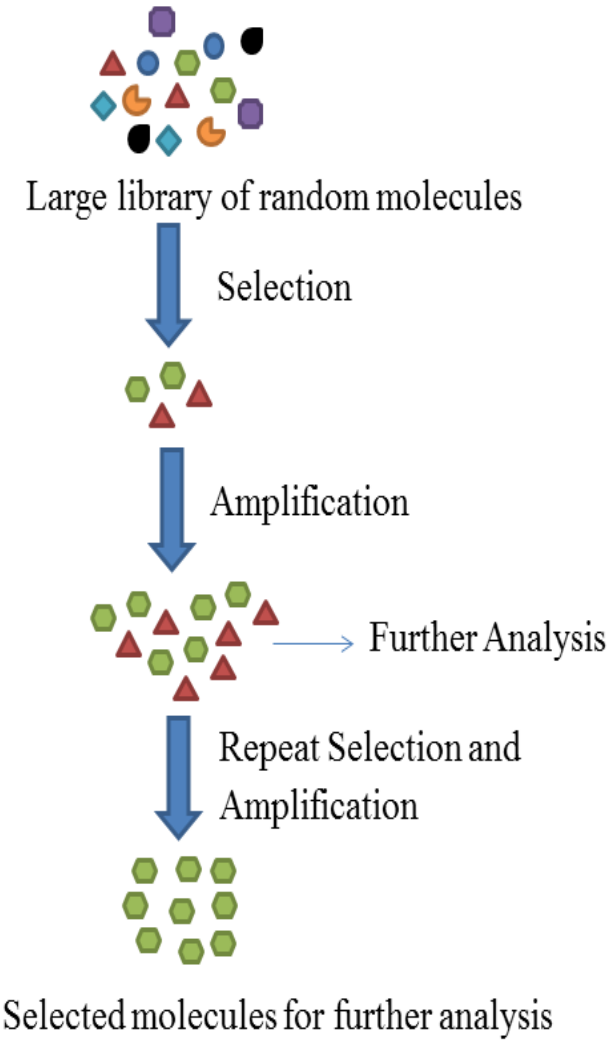

Figure 1.2. General scheme for *in vitro* selection approaches

Another difficulty in characterizing nonspecific interactions results from their lack of particular stable structures. For interactions that result in formation of a specific structure, conventional biophysical approaches such as X-ray crystallography, NMR spectroscopy and electron microscopy can be employed to characterize the structure of the complex and then that structural information can be used to understand the mechanism of these interactions. These approaches cannot be applied to the study of

nonspecific interactions because the binding complexes resulting in nonspecific interactions do not form a single final structure.

Approaches for Studying Nonspecific Interactions

A limited number of approaches have been developed to study nonspecific interactions. One chemical way to identify nonspecific protein interactions is using cross-linking reagents such as formaldehyde to crosslink molecules in close vicinity to the target protein[35, 36]. Although chemical crosslinking can help identify interactions that do not form stable binding structures, it can also result in complexes with molecules that are not in direct contact with the target protein, as long as they are in close proximity[23]. Bioinformatics methods have been used by Jiang's group at University of Washington to separate nonspecific effects from the many specific functions of proteins and further study the nonspecific interactions in molecular chaperones[18, 19]. There they define nonspecific interactions as interactions that lack specific protein functions. DNA and peptide microarrays allow high-throughput screening of compounds for affinity reagents and exploration of the diversity of interactions between proteins and surface bound molecules. Ordered array-based molecular interaction measurements have the advantage that one measures all strengths of interactions, rather than just seeing the strongest binder. Both DNA and peptide microarrays present a set of sequences with great chemical complexity that can be precisely controlled. Peptide microarrays, in particular, have great potential because the broad chemical diversity that can be represented in relatively modest length peptide sequences[37].

Different Peptide Microarray Platforms

Surface-bound peptide arrays can serve as a great model to explore the range of protein-surface interactions and to better understand the kinds of unstructured interactions that take place at chemically complex surfaces. Peptide arrays consist of immobilized peptides spatially addressed on substrates such as glass or silicon surfaces. They possess several advantages in protein-surface interactions research: (1) peptides are short fragments of protein molecules, which are key surface components of biological interfaces; (2) short peptides are relatively easy to make by chemical synthesis; (3) synthesis is not limited to using only the 20 naturally occurring amino acids; (4) modest size peptide libraries can present high chemical complexity and can be arranged in an addressable format; (5) Interactions can be studied systematically as interacting events between all binding partners, weak or strong, will be revealed. Currently, there are two major ways to prepare peptides on surfaces. One is pre-synthesizing functionalized peptides and then covalently immobilizing them to the surface via linker molecules; and the other one is *in situ* synthesis of peptide microarrays[38, 39].

There are several peptide immobilization approaches available to attach pre-synthesized peptides to surface. For example, peptides can be synthesized with a C-terminal linker containing a cysteine and covalently attached to a sulfo-SMCC treated surface via disulfide bond[40, 41]. This immobilization strategy has been employed by researchers at Center of Innovation in Medicine in Biodesign Institute at ASU to synthesize CIM10K peptide arrays, which has been used in several different applications including immunosignaturing and ligand discovery.

Some commercially available peptide arrays are synthesized by a variety of *in-situ* peptide synthesis approaches including SPOT synthesis, light-directed synthesis and particle-based synthesis. The SPOT synthesis technology delivers C-terminally activated amino acid derivatives and reactions reagents within liquid droplets onto defined areas at surface. Usually, a standard Fmoc solid phase peptide synthesis reaction is followed to couple these amino acid building blocks to surface[42]. SPOT technology based synthesis can be performed either manually or automatically and it is by far the most common and frequently used method for *in situ* synthesis of peptide arrays[39]. Particle-based synthesis is commercialized and utilized by PEPperPRINT to fabricate their PEPperCHIP. In this method, Fmoc-amino acids with C-terminal OPfp-ester activation group are embedded within a solid microparticle and printed onto the surface with a custom 24-color laser printer. This is followed by simultaneous particle melting, which releases the amino acids and enables the coupling reactions. As in conventional Fmoc solid phase synthesis, a cycle of synthesis is finished after washing off excessive materials and Fmoc deprotection[42]. The idea of light-directed peptide array synthesis, which utilizes photolabile protected amino acids and lithographic masks, was brought up by Fodor et al. in 1991[43]. The original technology requires the use of photolabile protecting groups, which did not perform well in term of repetitive coupling yield when compare to the conventional t-boc or Fmoc protecting groups, for a long time[42]. A couple of variant photolithographic synthesis methods were developed to overcome this problem by using photo acids to deprotect conventional t-boc protected amino acids at the designated positions that are exposed to light. However, these techniques are still associated with large numbers of peptide-specific coupling cycles as only one single kind

of amino acid can be coupled at a time. Two commercially available peptide array platforms are used in the work presented in subsequent chapters and the platforms will be discussed below.

LC Sciences PepArray™

A type of photolithographic technique utilizing photogenerated acids (PGA), digital photolithographic masks and conventional t-boc solid phase synthesis has been commercialized by LC Sciences to produce PepArray™, which is used in chapter 2 and chapter 3 to explore the nature of unstructured interactions between proteins and surface bound peptides. The technique was developed based on proprietary µParaflo® microfluidics technology, which allows direct synthesis of peptides on a high density microfluidic chip as features in specific locations on the chip using a PGA to deprotect the t-boc protected amines of nascent peptides[44, 45]. Peptide libraries synthesized on the chips are made to order and can be completely customized. Figure 1.3 shows the ordering process of LC Science PepArray chips. Two formats of chips are available through LC Sciences, 4000 or 30,000 features on arrays with ~1.5cm^2 surface area. The chip is actually an enclosed microfluidic system, which contains fluid distribution channels and picoliter scale reaction chambers with physical isolation from each other (Figure 1.4)[44]. Each chamber contains one specific peptide sequence. The in-situ peptide synthesis consists of the following steps (Figure 1.5): (1) derivatizing the surface with a protected NH_2-linker, with a density on the surface of less than 1pmol per 1mm^2 area; (2) applying a solution of the PGA precursor (PGA-P) in dichloromethane; (3) deprotecting the protected amine in the desired reaction chamber with digital photolithographic masks, (4) coupling the amino acid, capping any unreacted linkers and

deprotecting the side chain using standard peptide synthesis procedure, (5) repeat step 2, 3 and 4 until all amino acids to be incorporated is added for coupling[45-48].

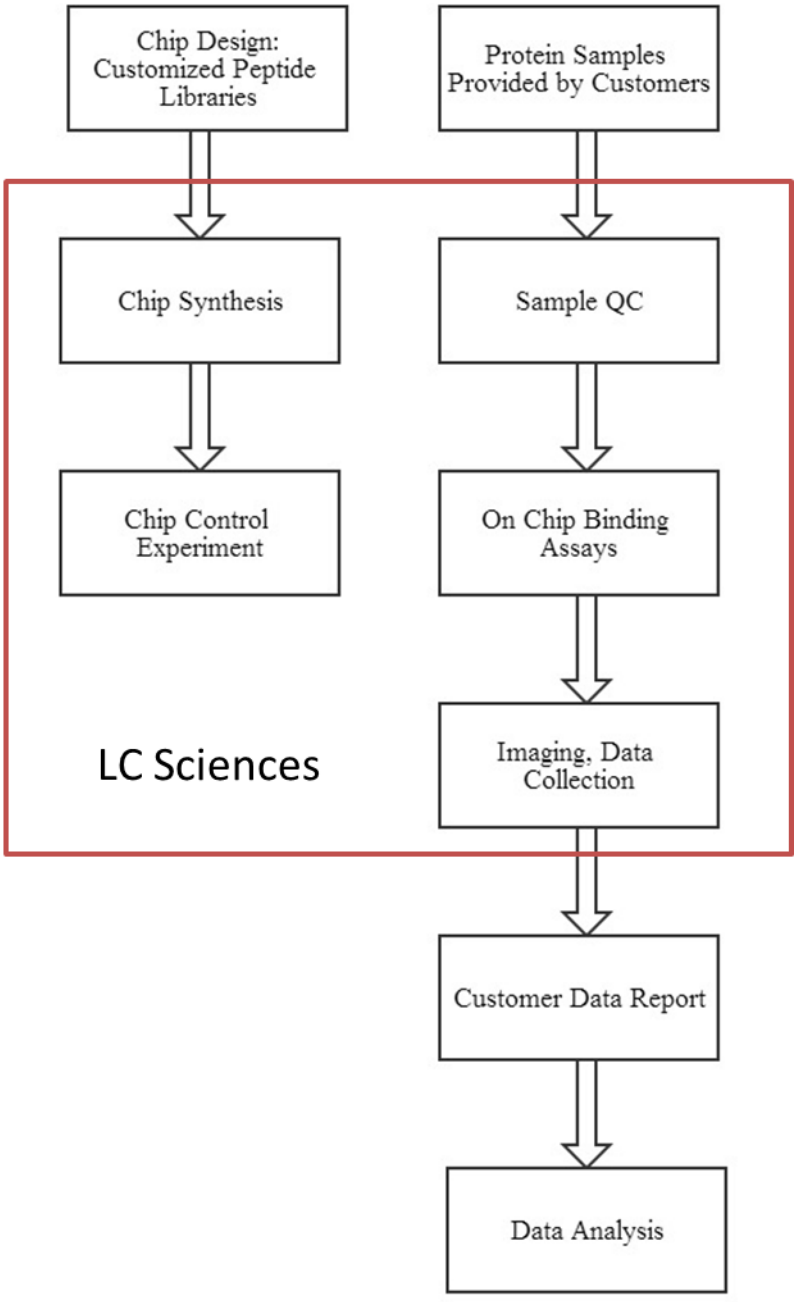

Figure 1.3. Flowchart of LC Science PepArray process. The red square indicates which steps were performed by LC Sciences.

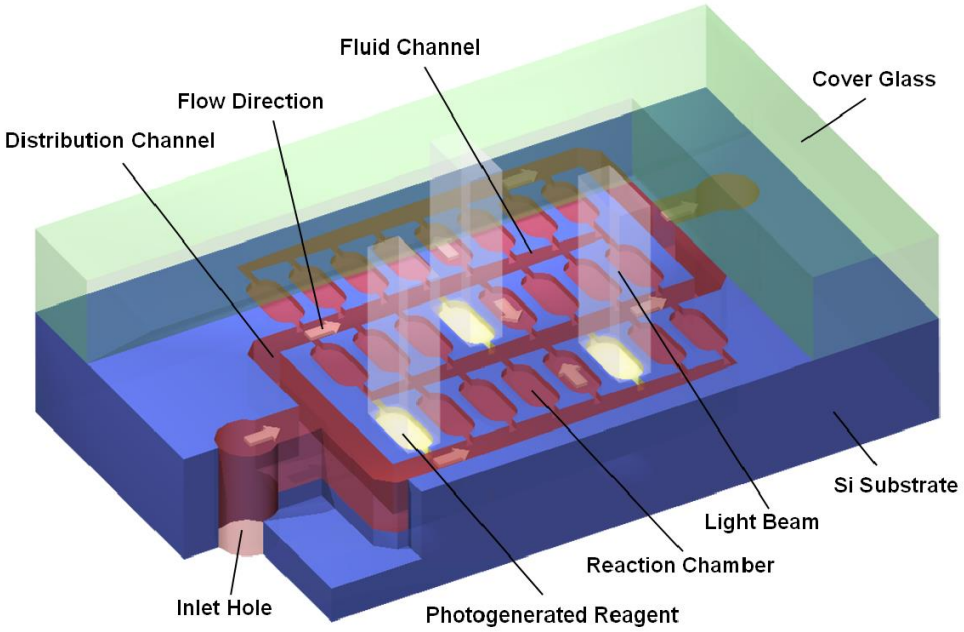

Figure 1.4. LC Science PepArray™ microfluidic system with fluid channels and reaction chambers[44].

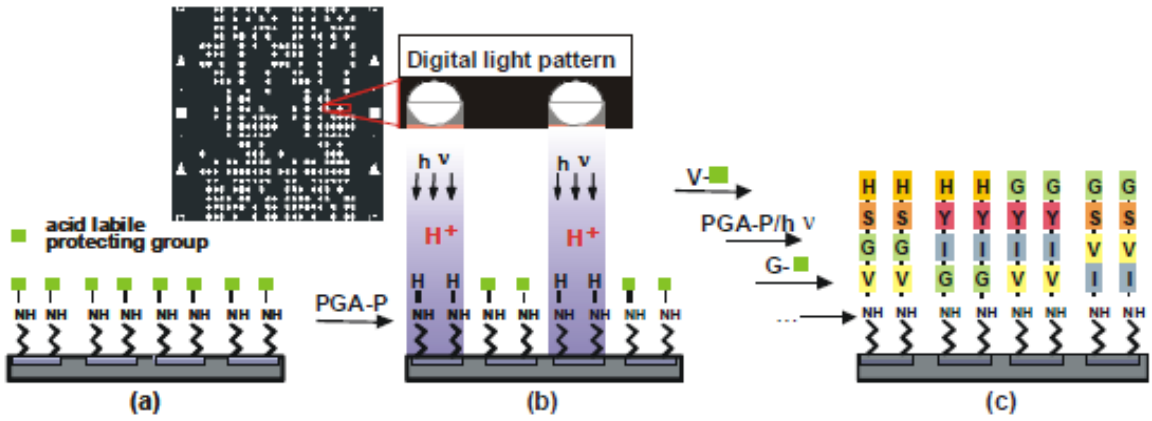

Figure 1.5. *In situ* peptide synthesis on the PepArray surface[44, 49].

HealthTell 330K Peptide Array

Another photolithographic mask-based patterned synthesis for peptide array production was developed and commercialized by HealthTell to fabricate very high density peptide arrays. Thus far, there are two types of peptide libraries available for the HealthTell arrays, one with ~330,000 peptide features on the array surface and the other with ~350,000. The work present in Chapter 4 utilizes the format of 330K peptide array to study the fundamental questions about unstructured interactions between immunoglobulins and complex, peptide covered surfaces. These 330K HealthTell peptide arrays consist of ~330,000 peptide sequences with an average length of ~12 amino acids (plus a three amino acid linker, GSG) and a length range from 3 to 17 amino acids. Sixteen of the twenty natural amino acids were used in the synthesis (cysteine, methionine, isoleucine and threonine were excluded). The ~330,000 feature array covers an area of ~0.5 cm^2 and each feature is 8 microns in diameter and contains a different peptide sequence. The peptide sequences were generated using a pseudo-random algorithm designed to minimize the number of synthesis steps while sampling combinatorial sequence space fairly evenly. The procedure for synthesis of the HealthTell peptide arrays is described below. First, Boc-glycine is uniformly attached to the thermal oxide-coated silicon wafer surface derivitized with monolayer of amnio silane; second, a photoresist with photoacid generator is spun onto the wafer surface and exposed through a mask to 365nm light; third, Boc-protected amines are deprotected at desired feature locations according to the pattern of the mask used; and last, a coupling solution with an amino acid is spun onto the wafer surface and the coupling reaction takes place at the

deprotected features. The process is repeated to create the desired peptide sequences at each feature (adapted from a procedure from Legutki et al., in press).

Proteins Used in the Experiments

A few individual proteins, including β-galactosidase, α-1 antitrypsin and immunoglobulins, as well as mixtures of protein and serum have been used in the next three chapters of this thesis to profile the interactions between proteins and surface bound peptides. An overview of these proteins and their known binding partners is presented below.

β-Galactosidase: Function, Structure and Modulation

β-galactosidase is a well-studied enzyme that catalyzes hydrolysis of lactose and other β-galactosides into monosaccharides[50]. It is encoded by the *LacZ* gene, which is one of the three adjacent genes in the *lac* operon model developed by Jacob and Monod[51]. *E.coli* β-galactosidase is a tetramer with four identical polypeptide chains, each of 1023 amino acids[52]. Each monomer is composed of five domains and domain 3 has an α/β or 'TIM' barrel structure with the active site located at the C-terminal end of the barrel (Figure 1.6)[50, 53, 54]. Furthermore, it has been found that certain residues near amino-terminal contribute significantly to the function of the active site by stabilizing each of the dimers in the overall tetrameric structure[55, 56]. This phenomenon, known as α-complementation, is critical to maintain the enzyme activity and served as the basis for the common blue/white screening (with X-gal) used in cloning[57].

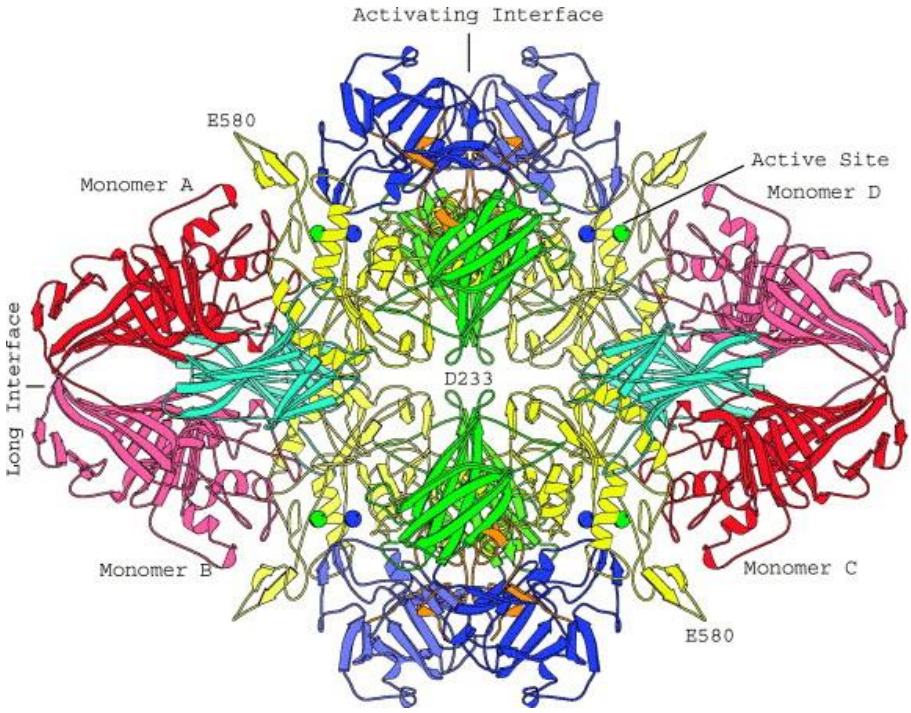

Figure 1.6. The backbone structure of the β-galactosidase tetramer[50, 54]. Coloring is by domain: complementation peptide, orange; Domain 1, blue; Domain 2, green; Domain 3, yellow; Domain 4, cyan; Domain 5, red. Lighter and darker shades of a given color are used to distinguish the same domain in different subunits. The metal cations in each of the four active sites are shown as spheres: Na^{+}, green; Mg^{2+}, blue[50].

It has been long known that the enzymatic activity of β-galactosidase can be either activated or inhibited through interactions with antibodies[58, 59]. A more recent study conducted by Fu et.al showed that the enzymatic activity can also be regulated by relatively short peptides[60, 61]. Proteolytic mapping of peptide binding to tetrameric β-galactosidase showed that the binding region is located at the activating interface[61].

α1-Antitrypsin: A Serine Protease Inhibitor (Serpin)

α1- antitrypsin is a plasma glycoprotein that belongs to serpin (serine protease inhibitor) superfamily. It is the major physiologic inhibitor of human neutrophil elastase,

which can break down elastin, a major contributor to the elasticity of the lungs[62]. α1-antitrypsin deficiency, which causes chronic obstructive pulmonary disease (COPD) and chronic liver disease, is a genetic disorder that is caused by a variety of mutations (i.e. S mutant (Glu264Val), Z mutant (Glu342Lys)) in the gene encoding the protein[63]. The mature α1-antitripsin protein in human plasma is a 394 amino acids monomer with carbohydrate side chains linked to the protein via asparagine[62, 64]. It consists of nine α-helices (A–I), three β-pleated sheets (A, B and C), and a mobile reactive center loop, which can be cleaved and inserted into the β-pleated sheet A when it interacts with the target proteases[65-67] (Figure 1.7). The reactive center loop in the Z mutant of one α1-antitrypsin can interact with β-sheet A in a second due to the aberrant opening of the β-sheet A resulting from the Glu342Lys mutation at the end of the reactive loop and the top of the strand 5 of β-sheet A[68, 69]. This can cause an abnormal loop-sheet polymerization in α1-antitrypsin and a polymer of α1-antitrypsin will form and accumulate in the hepatocyte and lead to a plasma deficiency[70, 71].

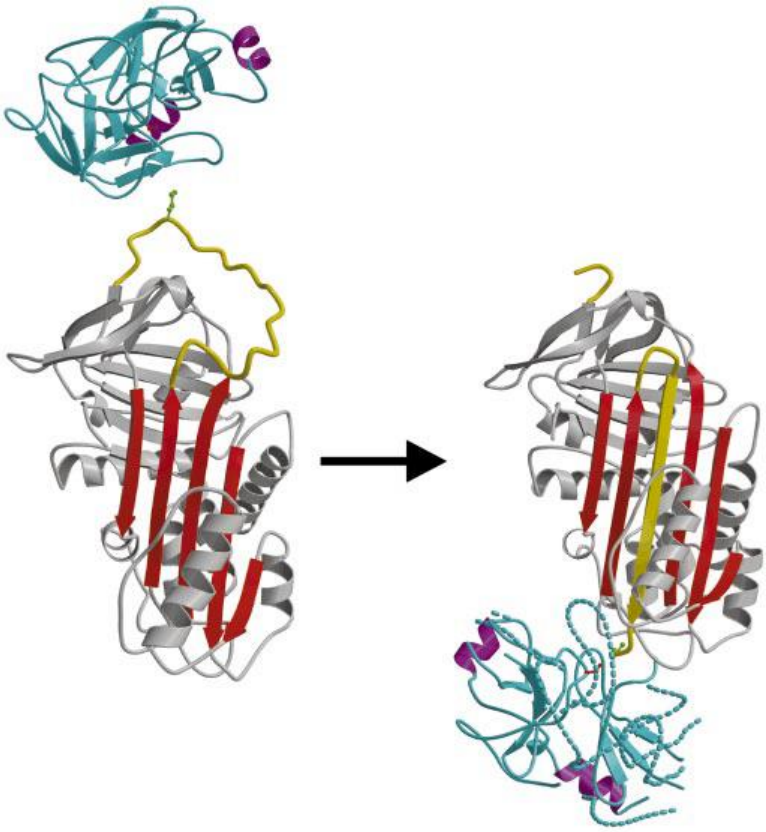

Figure 1.7. Formation of the α1-antitrypsin-trypsin complex[65]. Red, α1-antitrypsin β-sheet A; yellow, mobile reactive center loop.

A number of synthetic peptides have been found to bind to the mutant form of α1-antitrypsin to stop the polymerization[69]. This involves the binding of the reactive center loop analogue peptides to the loop insertion site in β-sheet A. Polymerization may be blocked because the space for insertion is filled with the synthetic peptide. A 6-mer reactive loop peptide (FLGAIG) has demonstrated high specificity and affinity to Z α1-antitrypsin and other similar peptides (i.e. FLAAIG, FLEAAG, FLAA, TTAI etc) were later confirmed possessing similar binding propensity[71-73]. Another way to prevent polymerization is to fill the cavity, a surface hydrophobic pocket which is present in the native α1-antitrypsin but filled when the β-sheet A accepts a reactive loop peptide during

polymerization[67, 69]. While the space-filling mutation has demonstrated the possibility of blocking polymerization, there is no effective synthetic peptide that mimics this effect[66].

Antibodies: Monoclonal Antibody and Serum

Antibodies are "Y"shaped protein molecules produced by plasma B-cells to identify and bind foreign objects. They are produced when anything recognized as foreign is found by the immune system. There are five classes of antibodies: Immunoglobulin G (IgG), IgM, IgA, IgD, and IgE. These immunoglobulins, or antibodies, all have the basic four-chain antibody structure, which is composed of two identical heavy chains and two identical light chains[74]. The antibody structure contains two antigen-binding (Fab) regions, which help the binding of antibody to antigen, and one Fc region, which mediates interaction with effector molecules[74]. The class of antibody is determined by its heavy chain while all classes have the same light chains. Light chains are composed of one constant domain and one variable domain, whereas heavy chains have either three or four constant domains and one variable domain. The variable domains of heavy and light chains form the antigen-binding site, which is located at the tip of the Fab regions[75].

Monoclonal antibodies are commercially available monospecific antibodies that are produced by a clonal cell line and that bind to a specific epitope. They are good affinity reagents, which are useful in developing tools in therapy, diagnostics and purification. It is well known that monoclonal antibodies bind their cognate sequences with high affinity and specificity; however, some recent studies have shown that they can also have strong and unique interactions with noncognate peptide sequences[20, 76, 77].

Serum from blood contains a heterogeneous mixture of many biomolecules including a large diversity of different antibodies. The major type of antibody in the serum is IgG. Serum is used in many diagnostic tests to detect different type of diseases including infectious diseases and cancer. A novel approach, an immunosignature, utilizes peptide microarrays to profile the humoral immune response by monitoring the interaction between serum and a large array of surface bound peptides[41]. It is clear that serum from people with certain diseases present different binding patterns in comparison with the binding patterns observed in healthy individuals.

Project Overview

In this dissertation, surface-bound peptide microarrays are used as a model to explore the range of protein-peptide interactions that take place at chemically complex surfaces. A new approach for studying interactions between protein and surface bound peptides is described and utilized in the subsequent chapters. The approach is statistical, simply looking at a large number of different potential binding partners on surfaces and using that to understand the chemistry of binding between proteins and chemically complex surfaces. Chapter 2 addresses how proteins behave on the arrays and to what extent proteins bind to the short random-sequence peptides made using a subset of 7 amino acids out of 20 natural amino acids at surfaces. The effect of charge, hydrophobicity and length of surface-bound peptides on surface affinity of β-galactosidase and α1-antitrypsin and the degree of differential binding will be discussed in detail. Chapter 3 tests a set of hypotheses of practical importance in the development of complex surfaces that interact with individual proteins (β-galactosidase as an example) and protein mixtures. These hypotheses center around fundamental questions of to what

extent the properties of short peptide sequences can be added together to create longer sequences with composite binding function and how one might build such surfaces from smaller components of known binding characteristics. Chapter 4 addresses how monoclonal antibodies and serum interact with diverse-sequence peptides on the surface of high density peptide arrays. How the affinity of monoclonal antibodies and serum to surface-bound peptides correlate with the physical properties of such peptides will be discussed in detail.

References

[1] Ostuni E, Yan L, Whitesides GM. The interaction of proteins and cells with self-assembled monolayers of alkanethiolates on gold and silver. Colloid Surface B 1999;15:3-30.

[2] Volkmer R, Tapia V, Landgraf C. Synthetic peptide arrays for investigating protein interaction domains. FEBS Lett 2012;586:2780-6.

[3] Zhu X, Gerstein M, Snyder M. Getting connected: analysis and principles of biological networks. Genes & development 2007;21:1010-24.

[4] Mager MD, LaPointe V, Stevens MM. Exploring and exploiting chemistry at the cell surface. Nature chemistry 2011;3:582-9.

[5] Zaia J. At last, functional glycomics. Nature methods 2011;8:55-7.

[6] Taniguchi CM, Emanuelli B, Kahn CR. Critical nodes in signalling pathways: insights into insulin action. Nature reviews Molecular cell biology 2006;7:85-96.

[7] Kasemo B. Biological surface science. Surface science 2002;500:656-77.

[8] Ratner BD. The engineering of biomaterials exhibiting recognition and specificity. J Mol Recognit 1996;9:617-25.

[9] Gopel W. Bioelectronics and nanotechnologies. Biosens Bioelectron 1998;13:723-8.

[10] Andresen H, Grotzinger C. Deciphering the Antibodyome - Peptide Arrays for Serum Antibody Biomarker Diagnostics. Curr Proteomics 2009;6:1-12.

[11] Graves DJ. Powerful tools for genetic analysis come of age. Trends in Biotechnology 1999;17:127-34.

[12] Ball V, Maechling C. Isothermal microcalorimetry to investigate non specific interactions in biophysical chemistry. Int J Mol Sci 2009;10:3283-315.

[13] Johnson ME, Hummer G. Nonspecific binding limits the number of proteins in a cell and shapes their interaction networks. Proc Natl Acad Sci U S A 2011;108:603-8.

[14] Kalodimos CG, Biris N, Bonvin AM, Levandoski MM, Guennuegues M, Boelens R, et al. Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. Science 2004;305:386-9.

[15] Halford SE, Marko JF. How do site-specific DNA-binding proteins find their targets? Nucleic Acids Res 2004;32:3040-52.

[16] Tan JM, Wong ES, Lim KL. Protein misfolding and aggregation in Parkinson's disease. Antioxidants & redox signaling 2009;11:2119-34.

[17] Bucciantini M, Giannoni E, Chiti F, Baroni F, Formigli L, Zurdo J, et al. Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. Nature 2002;416:507-11.

[18] White AD, Nowinski AK, Huang WJ, Keefe AJ, Sun F, Jiang SY. Decoding nonspecific interactions from nature. Chem Sci 2012;3:3488-94.

[19] White AD, Huang WJ, Jiang SY. Role of Nonspecific Interactions in Molecular Chaperones through Model-Based Bioinformatics. Biophys J 2012;103:2484-91.

[20] Stafford P, Halperin R, Legutki JB, Magee DM, Galgiani J, Johnston SA. Physical characterization of the "immunosignaturing effect". Mol Cell Proteomics 2012;11:M111 011593.

[21] Dey B, Thukral S, Krishnan S, Chakrobarty M, Gupta S, Manghani C, et al. DNA-protein interactions: methods for detection and analysis. Mol Cell Biochem 2012;365:279-99.

[22] Berggard T, Linse S, James P. Methods for the detection and analysis of protein-protein interactions. Proteomics 2007;7:2833-42.

[23] Phizicky EM, Fields S. Protein-protein interactions: methods for detection and analysis. Microbiological reviews 1995;59:94-123.

[24] Miernyk JA, Thelen JJ. Biochemical approaches for discovering protein-protein interactions. The Plant journal : for cell and molecular biology 2008;53:597-609.

[25] Shoemaker BA, Panchenko AR. Deciphering protein-protein interactions. Part I. Experimental techniques and databases. Plos Comput Biol 2007;3:e42.

[26] Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol 1999;17:1030-2.

[27] Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, et al. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods 2001;24:218-29.

[28] Bauch A, Superti-Furga G. Charting protein complexes, signaling pathways, and networks in the immune system. Immunological reviews 2006;210:187-207.

[29] Wilson DS, Szostak JW. In vitro selection of functional nucleic acids. Annu Rev Biochem 1999;68:611-47.

[30] Smith GP. Filamentous Fusion Phage - Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. Science 1985;228:1315-7.

[31] Keefe AD, Szostak JW. Functional proteins from a random-sequence library. Nature 2001;410:715-8.

[32] Hanes J, Pluckthun A. In vitro selection and evolution of functional proteins by using ribosome display. P Natl Acad Sci USA 1997;94:4937-42.

[33] Clackson T, Wells JA. In vitro selection from protein and peptide libraries. Trends Biotechnol 1994;12:173-84.

[34] Zahnd C, Amstutz P, Pluckthun A. Ribosome display: selecting and evolving proteins in vitro that specifically bind to a target. Nature methods 2007;4:269-79.

[35] Sutherland BW, Toews J, Kast J. Utility of formaldehyde cross-linking and mass spectrometry in the study of protein-protein interactions. J Mass Spectrom 2008;43:699-715.

[36] Fu JL, Reinhold J, Woodbury NW. Peptide-Modified Surfaces for Enzyme Immobilization. Plos One 2011;6.

[37] Woodson M, Liu J. Functional nanostructures from surface chemistry patterning. Phys Chem Chem Phys 2007;9:207-25.

[38] Gao X, Pellois JP, Na Y, Kim Y, Gulari E, Zhou X. High density peptide microarrays. In situ synthesis and applications. Molecular diversity 2004;8:177-87.

[39] Katz C, Levy-Beladev L, Rotem-Bamberger S, Rito T, Rudiger SG, Friedler A. Studying protein-protein interactions using peptide arrays. Chemical Society reviews 2011;40:2131-45.

[40] Morales Betanzos C, Gonzalez-Moa MJ, Boltz KW, Vander Werf BD, Johnston SA, Svarovsky SA. Bacterial glycoprofiling by using random sequence peptide microarrays. Chembiochem : a European journal of chemical biology 2009;10:877-88.

[41] Legutki JB, Magee DM, Stafford P, Johnston SA. A general method for characterization of humoral immunity induced by a vaccine or infection. Vaccine 2010;28:4529-37.

[42] Breitling F, Nesterov A, Stadler V, Felgenhauer T, Bischoff FR. High-density peptide arrays. Mol Biosyst 2009;5:224-34.

[43] Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. Light-directed, spatially addressable parallel chemical synthesis. Science 1991;251:767-73.

[44] http://www.lcsciences.com/applications/proteomics/.

[45] Gao XL, Zhou XC, Gulari E. Light directed massively parallel on-chip synthesis of peptide arrays with t-Boc chemistry. Proteomics 2003;3:2135-41.

[46] Pellois JP, Zhou X, Srivannavit O, Zhou T, Gulari E, Gao X. Individually addressable parallel peptide synthesis on microchips. Nat Biotechnol 2002;20:922-6.

[47] Pellois JP, Wang W, Gao X. Peptide synthesis based on t-Boc chemistry and solution photogenerated acids. Journal of combinatorial chemistry 2000;2:355-60.

[48] Stewart JM, Young,J.D. Solid Phase Peptide Synthesis. 2nd edition ed: Pierce Chemical Co; 1984.

[49] Gao X, Zhou X, Gulari E. Light directed massively parallel on-chip synthesis of peptide arrays with t-Boc chemistry. Proteomics 2003;3:2135-41.

[50] Matthews BW. The structure of E. coli beta-galactosidase. Comptes rendus biologies 2005;328:549-56.

[51] Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 1961;3:318-56.

[52] Kalnins A, Otto K, Ruther U, Muller-Hill B. Sequence of the lacZ gene of Escherichia coli. The EMBO journal 1983;2:593-7.

[53] Jacobson RH, Zhang XJ, DuBose RF, Matthews BW. Three-dimensional structure of beta-galactosidase from E. coli. Nature 1994;369:761-6.

[54] Juers DH, Jacobson RH, Wigley D, Zhang XJ, Huber RE, Tronrud DE, et al. High resolution refinement of beta-galactosidase in a new crystal form reveals multiple metal-binding sites and provides a structural basis for alpha-complementation. Protein Sci 2000;9:1685-99.

[55] Welply JK, Fowler AV, Zabin I. Beta-Galactosidase Alpha-Complementation - Effect of Single Amino-Acid Substitutions. J Biol Chem 1981;256:6811-6.

[56] Welply JK, Fowler AV, Zabin I. Beta-Galactosidase Alpha-Complementation - Overlapping Sequences. J Biol Chem 1981;256:6804-10.

[57] Juers DH, Matthews BW, Huber RE. LacZ beta-galactosidase: structure and function of an enzyme of historical and molecular biological importance. Protein Sci 2012;21:1792-807.

[58] Accolla RS, Cina R, Montesoro E, Celada F. Antibody-mediated activation of genetically defective Escherichia coli beta-galactosidases by monoclonal antibodies produced by somatic cell hybrids. Proc Natl Acad Sci U S A 1981;78:2478-82.

[59] Roth RA, Rotman B. Inactivation of normal beta-D-galactosidase by antibodies to defective forms of the enzyme. J Biol Chem 1975;250:7759-65.

[60] Fu J, Cai K, Johnston SA, Woodbury NW. Exploring peptide space for enzyme modulators. J Am Chem Soc 2010;132:6419-24.

[61] Fu J, Reinhold J, Woodbury NW. Peptide-modified surfaces for enzyme immobilization. Plos One 2011;6:e18692.

[62] Kalsheker N. Alpha 1-antitrypsin: structure, function and molecular biology of the gene. Biosci Rep 1989;9:129-38.

[63] Crystal RG. The alpha 1-antitrypsin gene and its deficiency states. Trends in genetics : TIG 1989;5:411-7.

[64] Carrell RW, Jeppsson JO, Vaughan L, Brennan SO, Owen MC, Boswell DR. Human alpha 1-antitrypsin: carbohydrate attachment and sequence homology. FEBS Lett 1981;135:301-3.

[65] Huntington JA, Read RJ, Carrell RW. Structure of a serpin-protease complex shows inhibition by deformation. Nature 2000;407:923-6.

[66] Patschull AOM, Segu L, Nyon MP, Lomas DA, Nobeli I, Barrett TE, et al. Therapeutic target-site variability in a1-antitrypsin characterized at high resolution. Acta Crystallogr F 2011;67:1492-7.

[67] Elliott PR, Pei XY, Dafforn TR, Lomas DA. Topography of a 2.0 angstrom structure of alpha(1)-antitrypsin reveals targets for rational drug design to prevent conformational disease. Protein Sci 2000;9:1274-81.

[68] Parfrey H, Mahadeva R, Ravenhill NA, Zhou AW, Dafforn TR, Foreman RC, et al. Targeting a surface cavity of alpha(1)-antitrypsin to prevent conformational disease. J Biol Chem 2003;278:33060-6.

[69] Gooptu B, Lomas DA. Conformational Pathology of the Serpins: Themes, Variations, and Therapeutic Strategies. Annu Rev Biochem 2009;78:147-76.

[70] Lomas DA, Evans DL, Finch JT, Carrell RW. The Mechanism of Z-Alpha-1-Antitrypsin Accumulation in the Liver. Nature 1992;357:605-7.

[71] Chang YP, Mahadeva R, Chang WSW, Lin SC, Chu YH. Small-molecule peptides inhibit Z alpha(1)-antitrypsin polymerization. J Cell Mol Med 2009;13:2304-16.

[72] Mahadeva R, Dafforn TR, Carrelll RW, Lomas DA. 6-mer peptide selectively anneals to a pathogenic serpin conformation and blocks polymerization - Implications for the prevention of Z alpha(1)-antitrypsin-related cirrhosis. J Biol Chem 2002;277:6771-4.

[73] Chang YP, Mahadeva R, Chang WSW, Shukla A, Dafforn TR, Chu YH. Identification of a 4-mer peptide inhibitor that effectively blocks the polymerization of pathogenic Z alpha(1)-antitrypsin. Am J Resp Cell Mol 2006;35:540-8.

[74] Delves PJ, Martin SJ, Burton DR. Essentials : Roitt's Essential Immunology (12th Edition). Hoboken, NJ, USA: Wiley-Blackwell; 2011.

[75] Woof JM, Burton DR. Human antibody-Fc receptor interactions illuminated by crystal structures. Nature reviews Immunology 2004;4:89-99.

[76] Halperin RF, Stafford P, Johnston SA. Exploring antibody recognition of sequence space through random-sequence peptide microarrays. Mol Cell Proteomics 2011;10:M110 000786.

[77] Reineke U, Ivascu C, Schlief M, Landgraf C, Gericke S, Zahn G, et al. Identification of distinct antibody epitopes and mimotopes from a peptide array of 5520 randomly generated sequences. Journal of immunological methods 2002;267:37-51.

CHAPTER 2: SELECTIVE PROTEIN-PEPTIDE INTERACTIONS AT SURFACES

ABSTRACT

Protein-surface interactions are of critical significance in both biological and man-made systems. While the term "specific binding" is normally reserved for the description of well-structured interactions, it is often the case in biology that there are unstructured interactions that greatly favor some protein interactions over others, a necessity in the highly crowded environment of the cell. In this study, surface-bound peptide-arrays were used as a model to explore the range of protein-surface interactions and to better understand the kinds of "nonspecific" or unstructured interactions that take place at chemically complex surfaces. Three samples, β-galactosidase, α1-antitrypsin and a mixture of 9 different proteins, were bound to arrays of nearly 5000 different peptides with a wide range of hydrophobicity, charge and peptide length. All three protein samples show higher binding affinity to positively charged peptides. While β-galactosidase binds poorly to very hydrophobic peptides, either in terms of absolute binding or relative to the mixture of proteins, α1-antitrypsin binds with higher affinity to more hydrophobic peptides. More surprising is the observation that β-galactosidase affinity for the surface does not simply increase with the length of the peptide, as one might expect, even when only the best binders are considered. Instead, its affinity (both absolute and relative to the protein mixture) peaks in the 4-9 amino acid residue range and then decreases substantially by 12 amino acids. In contrast, α1-antitrypsin increases nearly monotonically with peptide length, both in terms of apparent affinity and binding relative to other proteins. Of particular significance in a practical sense, it was possible to obtain

quite specific binding; the identity of the 100 peptides that showed the best apparent affinity for each of the three protein samples overlapped very little. Thus, using this approach it would be straightforward to develop surfaces covered with specific short peptide sequences with relatively specific protein interaction profiles. This work has been published on *Acta Biomaterialia*. Wang W, Woodbury NW. Selective protein-peptide interactions at surfaces. *Acta Biomater* 2014;10:761-8.

INTRODUCTION

The interaction between proteins and chemically heterogeneous surfaces is important in both biological and man-made systems[1]. Cell surfaces are involved in myriad activities in which proteins in solution interact with cellular components including cell-cell interactions, cell-surface adhesion, protein hormone sensing and exchange of nutrients or other small molecules that are carried by proteins. The need to avoid nonspecific binding of proteins to surfaces and other proteins in this environment likely limits the total number and surface properties of proteins in the cell[2]. Interactions between proteins and man-made material surfaces have become increasingly important in the development of implants, artificial tissues, and diagnostic assays, as well as in the incorporation of protein functions (e.g. enzymes, antibodies) in a variety of commercial products and devices[1, 3]. In this regard, some fundamental questions remain unanswered about the relationship between the structural nature and complexity of molecules on surfaces and their interactions with proteins. The concept of avidity, interactions between multiple closely spaced ligands on a surface and a protein, is not well understood in terms of the various kinds of interactions that take place and the

interplay between general chemical properties such as charge and hydrophobicity vs. structurally dependent interactions.

Affinity between a surface and a complex biological sample is often thought of in terms of specific and nonspecific interactions. The specific interactions are normally described in terms of well-defined, uniform contacts that result in a particular structure for the final complex formed. Non-specific interactions generally do not result in a unique structure for the complex, but instead are driven some combination of general interactions (e.g., charge, hydrophobicity) and a multiplicity of structurally dependent weak interactions between particular groups. However, non-specific interactions are not always non-selective or necessarily weak, and they can contribute significantly to the overall binding affinity, particularly where complex chemical surfaces are involved. For example, the non-specific adhesion of cells even to a relatively homogeneous surface such as a self-assembled monolayer can vary significantly with the details of the surface and the cells and its magnitude can dominate any specific interactions[4]. In fact in many cases the nonspecific interactions make up the bulk of the binding affinity with the specific interactions serving largely to limit the orientational entropy of the system. In such cases, the concepts of specific and nonspecific binding are probably better thought of as two ends of a more continuous spectrum. For example, in the interaction between an endonuclease and DNA is initially a charge-charge interaction and indeed this is a large fraction of the binding force[5]. This interaction allows the endonuclease to stay in the correct general orientation relative to the DNA but to diffuse along it without forming a specific structure until the site of action is reached [5-7]. The initial interaction is, in fact, quite specific for DNA. It is just not associated with a particular DNA structure

(sequence). Likewise, the general characteristics of a surface can result in very large discrimination between binding partners, even though no specific structures are formed.

A number of different methods have been used to characterize and to alter the interaction of biological macromolecules with surfaces. Metal or nanoparticle patterning has been used to create surfaces that differentially bind components of complex mixtures[8]. For example, Gilles et al. developed an approach to control cell adhesion and guide neurite outgrowth by patterned deposition of gold nanoparticles[9]. Yu et al. chemically modified hydrogels with nitrilotriacetic acid to enable protein immobilization at selected sites[10]. DNA and peptide microarrays are useful for exploring the diversity of interaction with proteins and other macromolecules because they present a set of sequences with great chemical complexity that can be precisely controlled. Surface-bound peptides, in particular, have considerable potential because of the broad chemical diversity that can be represented in relatively modest length sequences[8]. Another application that utilizes a library of relatively short peptides for discrimination of complex mixtures is immunosignaturing[11]. In that case, surface bound peptides are exposed to blood, binding to circulating antibodies and providing a profile of the antibody repertoire. This antibody profile is strongly correlated with disease state or immunological response to a vaccine [11, 12].

Here, the nature of nonspecific binding between proteins and surfaces covered with particular peptide sequences will be explored. This will be accomplished by systematically studying the binding of a number of different proteins to modest sized libraries of short (2-12 amino acid residue) peptides in ordered arrays on surfaces. Specifically, the affinity of specific proteins and the relative binding of different proteins

(the degree of differential binding) will be monitored as a function of the hydrophobicity, charge and length of peptides in the ordered arrays. While the general role of charge and hydrophobicity are relatively intuitive, it is less obvious how changing the length of the peptide bound to the surface will affect binding. First, increased length corresponds to increased chemical complexity; the more residues in each peptide, the greater the number of possible peptides that can be made. In the work described below, the choice of sequences is purely random and the libraries are too small to expect to create highly specific contacts. Second, longer peptides of random sequence tend to be more nonpolar on average. Third, as length is increased, the total number of possible interactions between a specific peptide and a protein increases; in principle, a greater number of interaction points between the peptide and protein should result in greater affinity, though there are both steric issues and entropy penalties involved. Finally, the possible number of peptide structures increases as the length is increased; this could result in partially specific structural pairing between peptide and protein, though again, conformational entropy comes into play.

MATERIALS AND METHODS

Peptide Microarrays

The peptide microarrays used were ordered as custom arrays from LC Sciences (Houston, TX)[13]. This company fabricates peptide arrays comprised of 30,000 peptides, up to 12 amino acids in length (below referred to as the 'variable' region). The array was divided in 6 identical sub-arrays (5000 features on each sub-array) and each was bound under same conditions, but with different protein samples. A sub-set of 7 of

the 20 natural amino acids {E, L, S, R, Q, W, Y} was used for peptide synthesis. Using such a sub-set of amino acids allows the study of a larger percentage or full sequence space for short peptides. These particular residues were chosen because they span the hydropathic index[14, 15] and produce a net neutral charge between them, while reducing the complexity of the peptide array synthesis by not using the entire set of 20 amino acids.

Each sub-array contains 4876 custom peptides (the rest are control peptides provided by the company). The peptide length was varied from 2 to 12 residues. All possible dipeptides (49), tripeptides (343) and an equal distribution of randomly selected tetrapeptides (896), pentapeptides (897), heptapeptides (897), nonapeptides (897) and dodecapeptides (896) were included in the library. A GSG tri-peptide linker was added to the C terminus of each sequence to maintain a uniform distance between the peptides and the array surfaces in addition to the proprietary surface linker[16] used by LC Sciences.

Protein Binding

Alexa Fluor®-555 (AF555, Invitrogen) and Alexa Fluor®-647 (AF647, Invitrogen) labeled proteins were used for the protein-peptide binding assay. The labeling was done using a kit (Invitrogen) and the manufacture's protocol was followed in the protein labeling process. β-galactosidase, α1-antitrypsin and a protein mixture included Fetuin, Horseradish peroxidase, Bovine serum albumin, Carbonic anhydrase Haptoglobin, Transferrin, Amylase, Pyruvate Kinase and Glucose Oxidase were used in the binding assays. Table 2.1 lists the size, pI, accessible solvent surface area, source, company and the estimated surface area relative to β-galactosidase for each protein. The

concentration of protein used in each case was adjusted to give a constant total surface area calculated according to footnote [e] of Table 2.1. Dye/ α1-antitrypsin labeling ratios were kept near 1 due to the small size of this protein and the increased possibility of dominant dye interactions. β-galactosidase, being a much larger protein is somewhat less sensitive to covalent addition of labeled dyes and thus for this protein labeling ratios were allowed to go as high as 3. The proteins in the mixture were labeled separately and then mixed together. The labeling ratios for each protein in the mixture were kept below 1. The final dye/protein ratio in the mixture was 0.5 (unlabeled proteins were added to the corresponding labeled proteins whose ratios were over 0.5 before the mixing to control the overall fluorescence intensity).

Table 2.1: Summary of proteins used in the binding assays

Protein	Size (KDa)	pI[a]	Accessible surface ($Å^2$)[c]	Accessible hydrophobic ($Å^2$)[d]	Source	Company	Rel. surface area[e]
β-galactosidase[17, 18]	464	4.61	3.00×10^5	3.49×10^4	*Escherichia coli*	Sigma G6008	1.000
α1-antitrypsin[19]	50-56	5.37[b]	2.34×10^4	3.20×10^3	Human plasma	Sigma A9024	0.226
Fetuin (Type III)[20]	48.4	3.3	N/A[c]	N/A[c]	Fetal calf serum	Sigma F2379	0.222
Horseradish peroxidase (Type VI)[21, 22]	44	7.20	2.76×10^4	3.07×10^3	Horseradish	Sigma P8375	0.208
Bovine serum albumin[23, 24]	66.4	5.3	6.70×10^4	1.25×10^4	Bovine serum	Sigma A7906	0.274
Carbonic anhydrase(Iso	31	5.9	4.66×10^4	3.98×10^3	Bovine	Sigma C2522	0.165

zyme II)[25]					erythrocytes		
Haptoglobin (Mixture type)	~86	6.13[b]	8.60×10^4	1.31×10^4	Human plasma	EMD4Biosciences 372022	0.325
Transferrin[26]	77	6.90	3.20×10^4	5.78×10^3	Human blood	Sigma T3309	0.302
α-Amylase (Type II-A)[27]	50	7.00	2.00×10^4	2.54×10^3	*Bacillus sp.*	Sigma A6380	0.226
Pyruvate Kinase (Type III)	237	7.60	7.25×10^4	1.69×10^4	Rabbit Muscle	Sigma P9136	0.639
Glucose Oxidase[28, 29]	160	4.20	8.10×10^4	1.07×10^4	*Aspergillus niger*	Sigma G7141	0.492

[a] pI is the abbreviation for the isoelectric point, which is the pH at which the particular protein carries no net electrical charge
[b] These values were theoretical values calculated based on the peptide sequences
[c]These values were calculated in PyMol using the corresponding protein crystal structure information from PDB database
[d] Crystal structure information for bovine fetuin is not available in major databases
[e] The relative surface area was calculated assuming that each protein could be considered as a perfect sphere using the equation [(molecular weight of sample protein)/(molecular weight of β-galactosidase)]$^{2/3}$

In order to test directly for effects of the dye on binding, proteins were labeled with two different dyes and all experiments were done with both protein samples. AF555 and AF647 labeled proteins were applied on identical sub-arrays respectively under the same experimental conditions. Peptides exhibiting significantly different binding behaviors between the two dyes were excluded from further data analysis, except as indicated.

The actual protein binding assays were performed by LC Sciences (LC Sciences both synthesized the arrays and performed the binding experiments). The following

procedures and conditions for binding were used. First, the chips were treated with blocking buffer (super block, a proprietary blocking solution, with 0.05% Tween-20 and 0.05% Triton X-100, pH7.4) overnight to minimize non-specific binding and then washed in washing buffer (1×PBS, pH 7.0). Second, the six different protein samples (AF555-β-galactosidase, AF647-β-galactosidase, AF555-α1-antitrypsin, AF647-α1-antitrypsin, AF555-protein mixture and AF647-protein mixture) were applied to 6 sub-arrays respectively and incubated for 1 hour in binding buffer (1×PBS, pH7.4) at 25°C. Third, the array was washed with washing buffer and imaged using either 635 nm excitation and 655-695 nm emission filter or 532 nm excitation and 550-600 nm emission filter, depending on which dye was used. Note that blank features on the array, left as a control, showed binding levels that were essentially the same as the lowest levels of protein binding.

Additionally, to verify that the relationships between binding and peptide sequence on the LC science arrays were not simply unique to that platform, 14 peptides that had been tested on the array were resynthesized by Sigma-Aldrich PepScreen® service and bound to the surface of microwell plates. These peptides covered the full range of binding strengths to β-galactosidase. Binding with AF555 and AF647 labeled β-galactosidase was repeated and the binding results compared (in terms of the binding rank) between the two platforms (Table S2.1). The order was very similar, implying that both the sequences from the two sources were, indeed, the same and the exact nature of the surface and attachment is not critical.

Data Analysis

Binding signals were normalized to the median intensity for each respective sub-array[30]. All subsequent analysis utilized the median-normalized data. For β-galactosidase and α1-antitrypsin, data from arrays bound with same protein but using different dye labels was compared, and peptides with binding values that were more than 2-fold different between the samples were excluded in further analyses (see Results). There were 4876 total peptides in the array, for β-galactosidase 897 were excluded and for α1-antitrypsin 1490 were excluded. For the protein mixture, binding signals from arrays bound with same mixture but using different dye labels were averaged and the mean of the binding signals was used in further analyses. Locally written scripts in either Matlab or Bash Shell environment were used to sort the peptides with different lengths and to perform most of the statistical studies. An online bioinformatics tool was used to generate the Venn Diagram[31].

RESULTS

Two proteins, β-galactosidase and α-1 antitrypsin were incubated with the LC-sciences peptide arrays, as described in Methods. The binding of these particular proteins was then compared to that of a mixture of 9 other proteins of various sizes and pI values (Table 1). These proteins were chosen to cover a broad range of sizes and isoelectric points. For both β-galactosidase and α-1 antitrypsin, the binding intensities for proteins labeled with two different fluorescent dyes (AF555 and AF647, see Methods) were compared and peptides that were suspected of binding dye primarily, rather than the protein, were removed from further analysis. The peptide arrays used for this analysis consisted of peptides ranging from 2 to 12 amino acids in length of random sequence (the

sequence of each peptide feature is homogeneous but selected with a random number generator) (see Methods for more details). This array composition makes it possible to explore various dimensions of protein interactions with complex surfaces. Even with very modest sized peptide libraries such as this, the dynamic range of the protein association with the surface is very high. For example, using β-galactosidase, the highest signals are roughly 60-fold higher than the lowest signals (Table 2.2). With small libraries, it is unlikely that the binding is strongly dependent on rigid structural interactions. Instead, long range interactions (charge) and small numbers of specific contacts enabled by the flexibility of the peptide chains are more likely. In order to explore this in more detail, the peptide set and associated binding was parsed in a number of ways: total charge, hydrophobicity and length.

Table 2.2: Peptides-protein binding dynamic range for each protein sample

Sample	Average Binding of the top 10 peptides	Average Binding of the bottom 10 peptides	Fold change[a]
β-galactosidase	18.54	0.29	63.9
α1-antitrypsin	2.97	0.50	5.92
Mixture with 9 proteins	4.78	0.45	10.6

[a] These values represent the fold change between binding of strongest binding peptides and weakest binding peptides. The equation used for calculating these values is: (Average binding of the top 10 peptides) / (Average binding of the bottom 10 peptides)

Effect of charge and hydrophobicity on surface affinity

The charge of both a protein and the surface it interacts with will play a major role in differential protein binding. β-galactosidase has a pI of 4.61 and α1-antitrypsin has a

pI of 5.37 (Table 2.1), thus both are negatively charged at neutral pH. In fact most of the proteins used in the protein mixture are negatively charged under the binding conditions used here. This is reflected in Figure 2.1 which shows how each of the protein samples tested interacts with peptides on the surface as a function of net peptide charge at neutral pH. All of the protein samples show higher affinity to surface features consisting of peptides that have more positive charges. The affect is particularly pronounced for β-galactosidase where the average binding intensity for positively charged peptides is 6.8 fold higher than that of negatively charged peptides and 2.9 fold higher than neutral peptides. α1-antitrypsin and the mixture of nine proteins show a considerably less pronounced increase (1.5 fold for α1-antitrypsin and 2.5 fold for the protein mixture relative to negatively charged peptides, Figure 2.1).

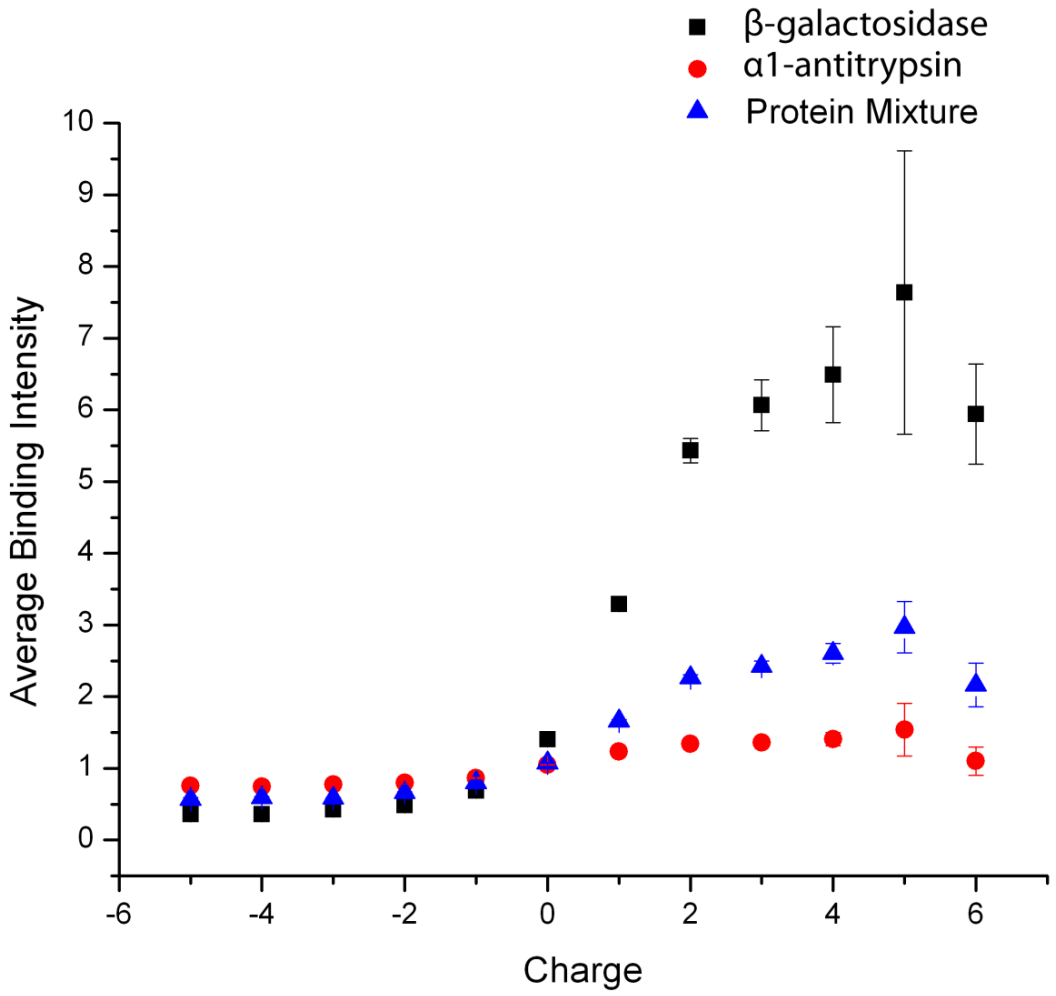

Figure 2.1. Average binding intensity as a function of net charge per peptide . The net charge of the peptides at neutral pH varies from -5 to +6. Intensity values shown are normalized to the median for each sample. The error bars shown represent the standard error in the mean for peptides with that net charge.

Charge is not the only physical parameter of the peptides that can strongly correlate with binding. As seen in Figure 2.2, the overall hydrophobicity of the peptide on each feature in the array also substantially affects the affinity of proteins for that

feature. In this case, however, the effects are different for the three protein samples. The hydrophobicity of each peptide was estimated from the sum of the hydrophobicity values for each amino acid [32]. Based on this, the peptides were sorted into four groups (very hydrophobic, hydrophobic, neutral and hydrophilic) according to their hydrophobicity value. At the highest hydrophobicity, β-galactosidase binding dropped by 2/3rds of its binding value for more polar peptides (Figure 2.2). In contrast, α1-antitrypsin binding increases by nearly 1/3rd as hydrophobicity increases. The mixture of nine proteins gives intermediate results, with a small increase in average binding for peptides with higher hydrophobicity values (about 15%).

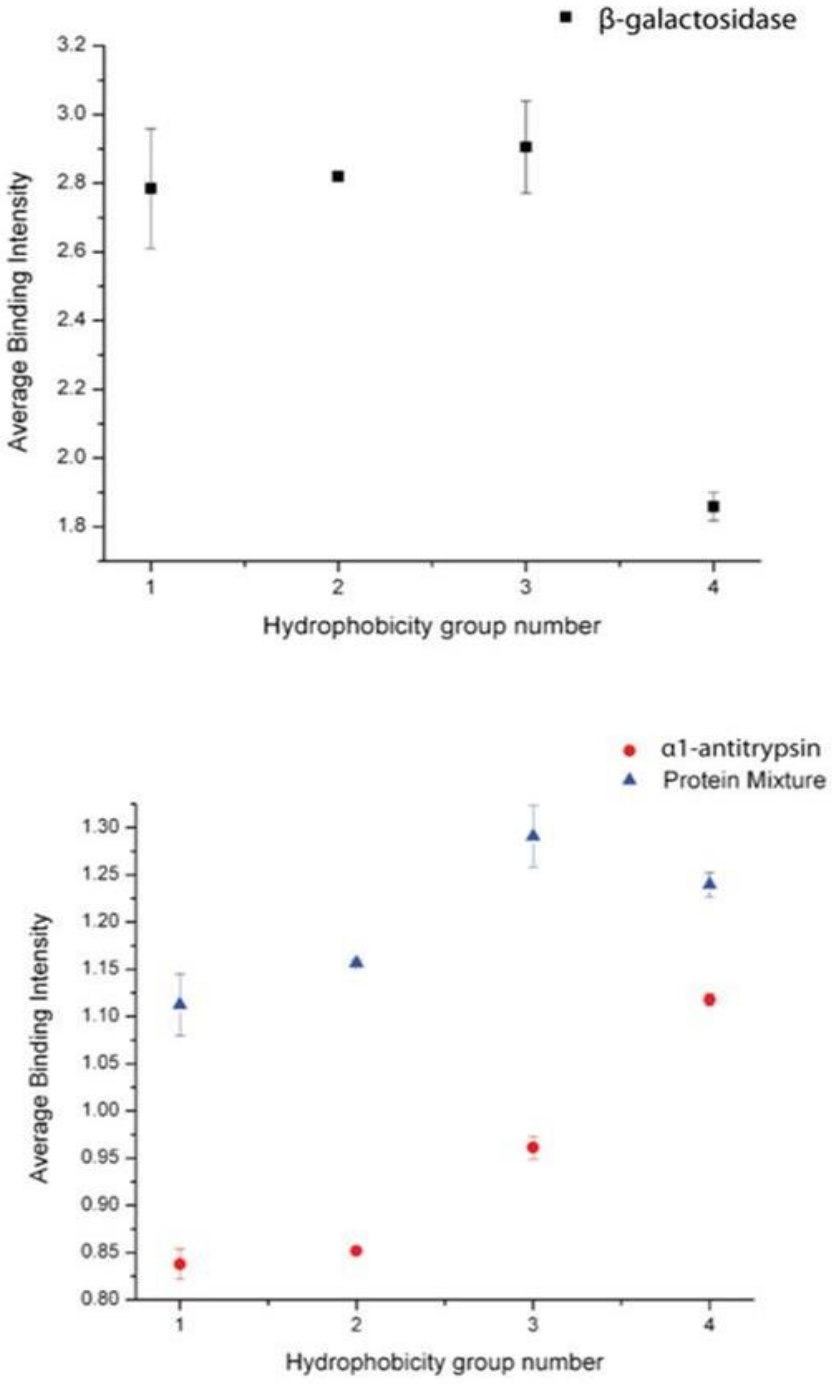

Figure 2.2. Binding intensity as a function of peptide hydrophobicity. The binding intensities from peptides of each hydrophobicity group were averaged together. There were four hydrophobicity groups. Group 1 (hydrophilic): hydrophobicity < -10; group 2

(neutral): hydrophobicity 20 to -10; group 3 (hydrophobic): hydrophobicity 20 to 70; group 4 (very hydrophobic) hydrophobicity > 70. Intensity values shown are normalized to the global median. The error bars shown represent the standard error in the mean for that peptide hydrophobicity group.

Effect of peptide length on surface affinity

Features in the array with longer peptides have potentially more complex chemical properties, by virtue of the fact that more chemically different amino acids are present and the number of possible sequences becomes exponentially larger with peptide length. One might think that binding would simple become stronger with length due to the larger number of possible interactions, however, previous work from the authors' laboratory suggested that the dependence of binding on peptide length was more complex than this[14], at least in some cases, prompting a more detailed study. The average binding intensity as a function of length for the three protein samples is shown in Figure2.3. For β-galactosidase, there appear to be two observable maxima, though the statistical validity is borderline. In any case, there is a ~2-fold increase in average binding observed as the peptide length is increased from 2 to 4 amino acids, but then the average binding signal seems to peak ~4 amino acids. As the length is increased beyond ~7 amino acids, the binding drops so that by 12 amino acids the average binding is similar to the binding to amino acid dipeptides. The situation for α1-antitrypsin is different. In contrast to β-galactosidase, the average binding affinity of α1-antitrypsin to the peptide array increases monotonically, or nearly so, with peptide length. The protein mixture shows essentially no significant length dependence of binding besides a small apparent decrease in binding signal at a peptide length of 7 amino acids.

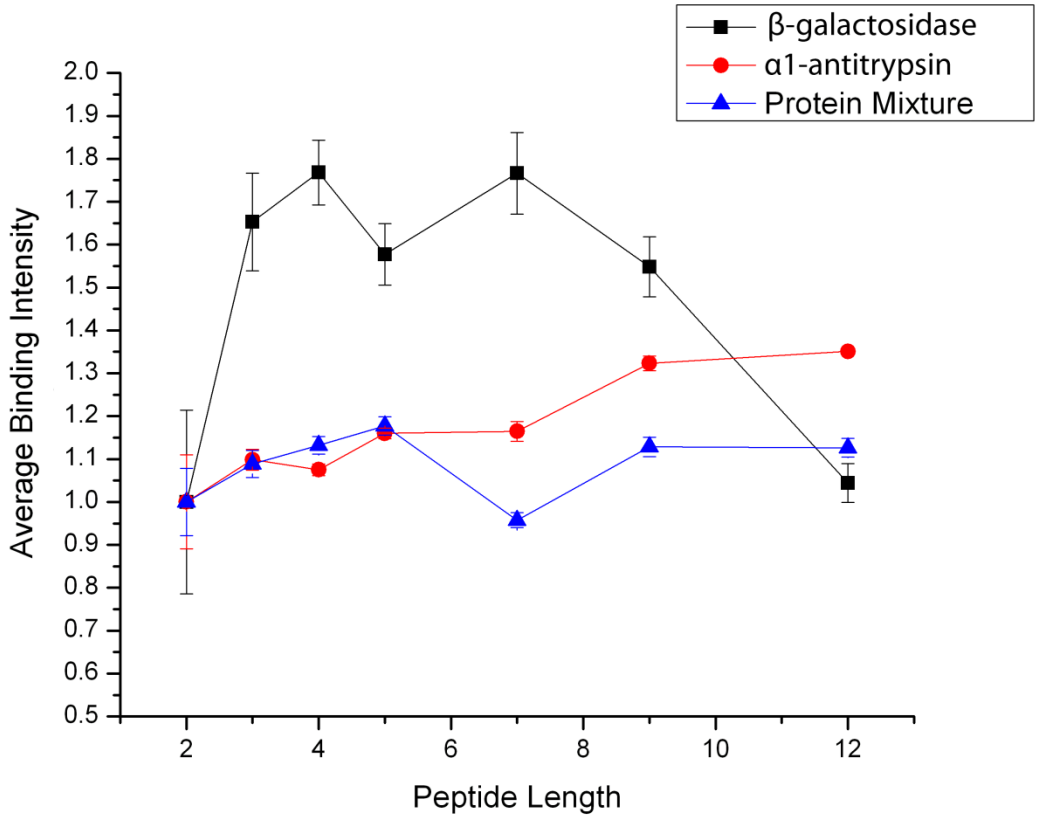

Figure 2.3. Average binding affinity versus length for each protein sample. The binding intensities from peptides of each length were averaged together. All values are normalized to the average binding intensity of the corresponding dimer peptide for that sample. The error bars shown represent the standard error of the mean for that peptide length.

More can be learned from looking at the distributions of peptide binding as a function of length. This is shown in Figure 2.4 for β-galactosidase. The binding distributions reflect the population of peptides that are bound to the labeled proteins at each length (each length is normalized so that the intensities shown for each length of peptide are given as a fraction of the total fluorescence). As might be expected, most

peptides of all lengths show relatively weak binding to β-galactosidase, and the number of peptides that bind at successively higher intensities falls off rapidly. There are a small number of peptides however that bind with quite high intensity. Similar distributions are seen for both α1-antitrypsin and the protein mixture (see supplementary material).

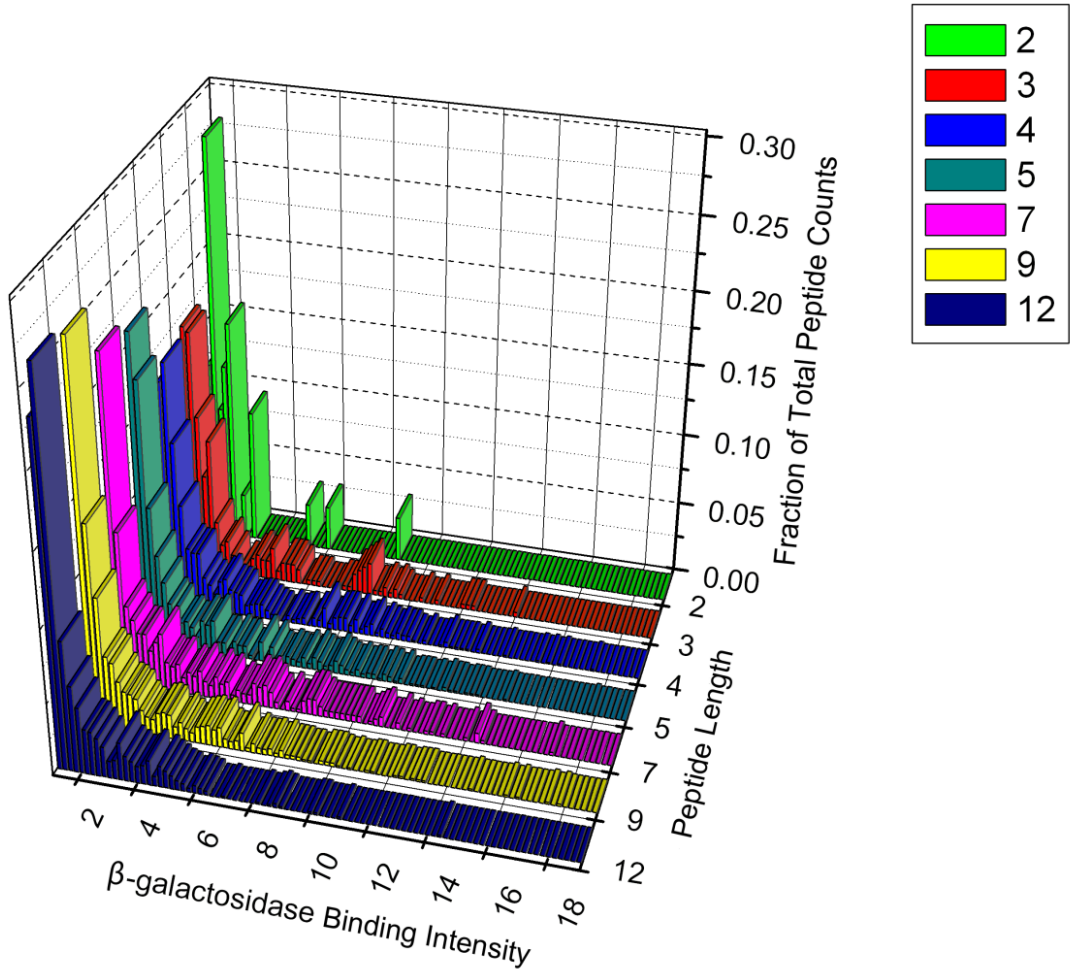

Figure 2.4: Histogram of β-galactosidase binding distribution as a function of peptide length. Fluorescence intensities shown are normalized to the total fluorescence intensity for the peptides of a particular length.

This brings up the question of whether the high binding signals have the same length dependence as does the average binding shown in Figure 2.3. One might think that the best binders would always be the longest ones, as these would simply have more to offer in terms of possible interactions. When the average signal from only the top 30 binding peptides at each length is plotted, the length dependence is generally similar to that seen for the average of all peptides in Figure 2.3 (Figure S2.3). However, when just the top binders are considered, binding for β-galactosidase does not drop as dramatically for longer peptides as it does when all peptide features are averaged. In addition, the peak of the binding shifts to longer peptide lengths, from of 7 to 9 amino acids. Finally, in the case of the protein mixture, the best binders are clearly the longer peptides, whereas the average of all peptides showed little length dependence. Thus, in general, it appears that the best binders are biased towards the longer peptides.

The relationship between length, charge and hydrophobicity

Peptide length is not entirely independent of charge and hydrophobicity; for randomly chosen sequences, longer peptides tend to be more hydrophobic and are statistically more likely to have more balanced charge characteristics. The relationship between binding, charge and length is shown for β-galactosidase in Figure 2.5. Here the length dependence of binding is plotted for positive, negative and neutral peptides. Negative peptides show less dependence on length (though still bind statistically better for mid-length than long peptides) while positive, and to a lesser extent neutral, peptides show a strong length dependence. This is particularly evident with regard to the drop in binding for longer, positive peptides; in this case, the binding to 12-mers is 2-fold lower than that to peptides that are 2-4 amino acids long. Very similar studies were performed

on α1-antitrypsin and the protein mixture (see supplementary material). In those two cases, charge had little effect on the length dependence.

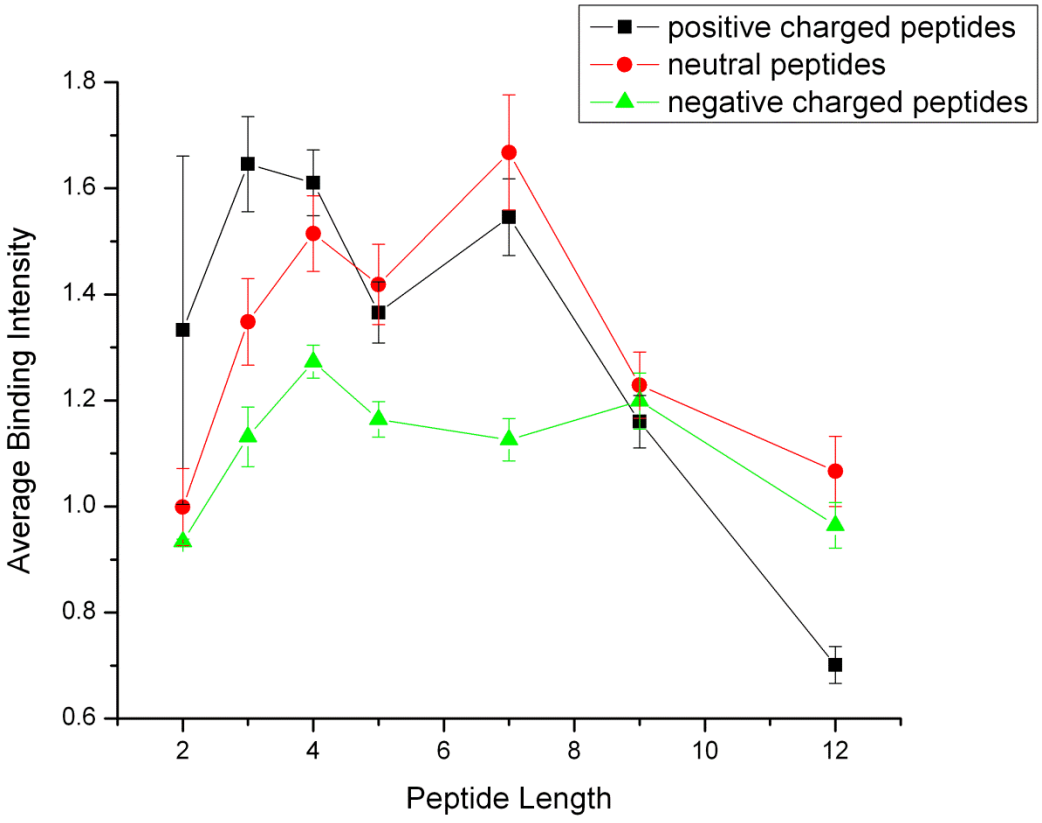

Figure 2.5. β-galactosidase average binding intensity versus peptide length for each charge group. The binding intensities from peptides of a particular net charge and length were averaged together. In this study, there were 1387 positively charged peptides (black squares), 1286 neutral peptides (red circles) and 1305 negatively charged peptides (green triangles). The intensity values shown are normalized to the median for each charge group. The error bars shown represent the standard error of the mean for that peptide length and net charge.

One might think that the observed drop in β-galactosidase binding at longer lengths may arise from the tendency of longer peptides to be more nonpolar, given the fact that more hydrophobic peptides tend to bind β-galactosidase more weakly (Figure 2.2). However, when the peptides were separated into hydrophilic, neutral, hydrophobic and very hydrophobic, all four classes showed lower binding at long lengths (Figure S2.6). The only significant difference was the behavior at shorter lengths, where the hydrophilic short peptides bound more strongly than the others.

The average binding intensity for α1-antitrypsin and protein mixture versus peptide length for four hydrophobicity groups was also studied but no significant difference in the overall binding trends was observed. (Figures not shown).

Discrimination between proteins

To what extent does a modest library of different peptides provide differential binding of one protein vs. another? Figure 2.6 shows a Venn diagram of the top 100 peptides binding to the three samples tested. There is generally very little overlap in which peptides bind best to each sample. By random chance, one would have expected approximately 2 peptides to be common between each sample. What one sees is that α1-antitrypsin shows about this expected overlap with the protein mixture, but both α1-antitrypsin and the protein mixture show a somewhat greater than random overlap with β-galactosidase. This is presumably because of the general propensity for positively charged peptides, particularly in the case of β-galactosidase, limiting the set of peptides from which the top 100 were selected. Still, even among highly positive peptides, there is an ability to distinguish easily between bindings to the three different protein samples.

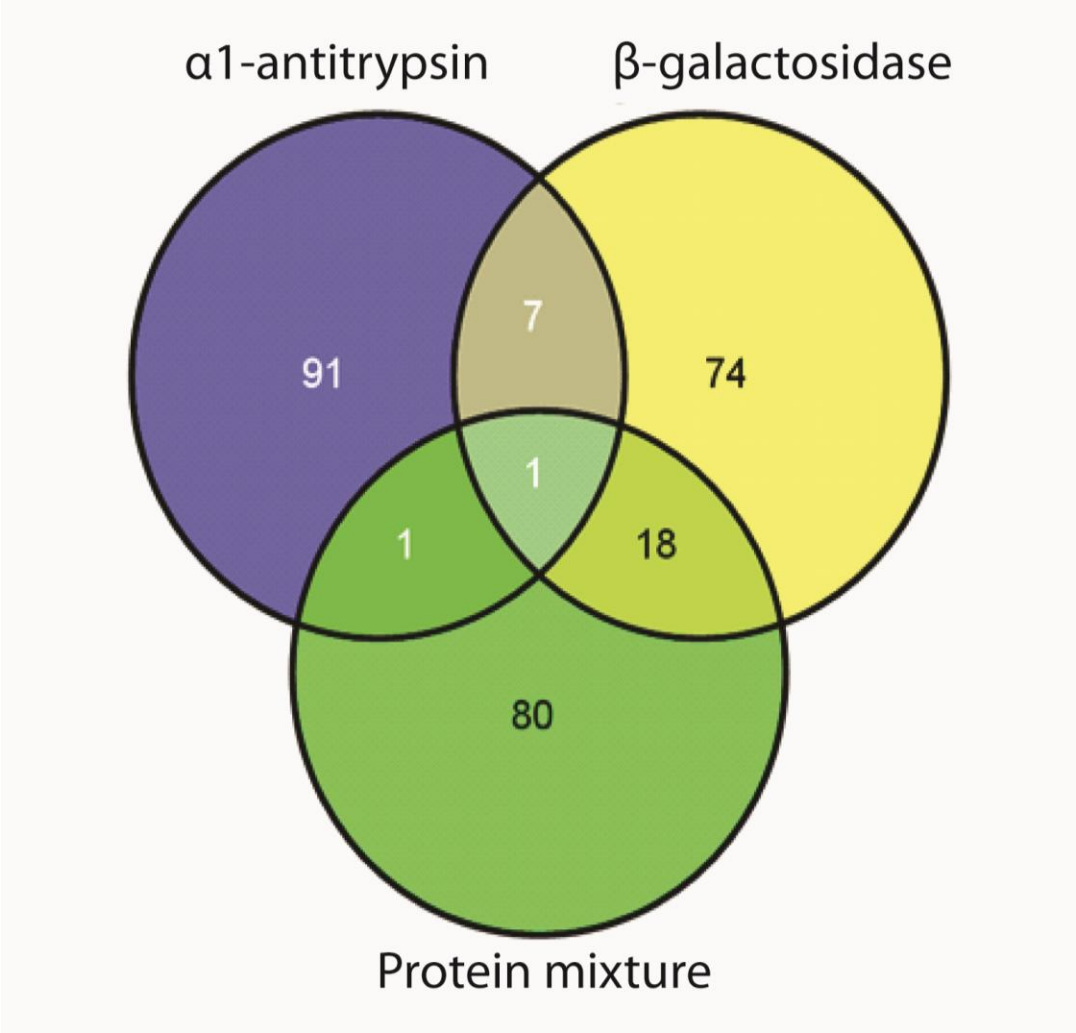

Figure 2.6. Venn diagram[31] of the highest affinity 100 peptides that bind to β-galactosidase, α1-antitrypsin and the mix of 9 proteins. There were 4876 total peptides in the array, but, 377 were excluded due to obvious dye effects on binding to both β-galactosidase and α1-antitrypsin.

Another way to consider the level of specificity of binding is to compare the way binding depends on charge, hydrophobicity and length in the two specific proteins β-galactosidase and α1-antitrypsin) relative to the mixture of 9 proteins. One can see from Figure 2.3, that the mixture shows little length dependence on average binding, suggesting that the forces that result in changes in affinity as a function of length also

discriminate different proteins from each other. More enlightening are the curves depicting the ratio between binding of each of the two individual proteins (β-galactosidase and α1-antitrypsin) and the mixture as a function of charge and hydrophobicity. As a function of charge (Figure 2.7), the specificity ratio for β-galactosidase increases in essentially the same way as does the affinity in Fig. 2.1; larger positive charge results in more binding of β-galactosidase compared to a mixture of nine proteins. However, for α1-antitrypsin, while binding overall increases somewhat for positively charged peptides, it does so less than the aggregate of the other nine proteins. Thus positively charged peptides in general are less selective for α1-antitrypsin than are negatively charged peptides.

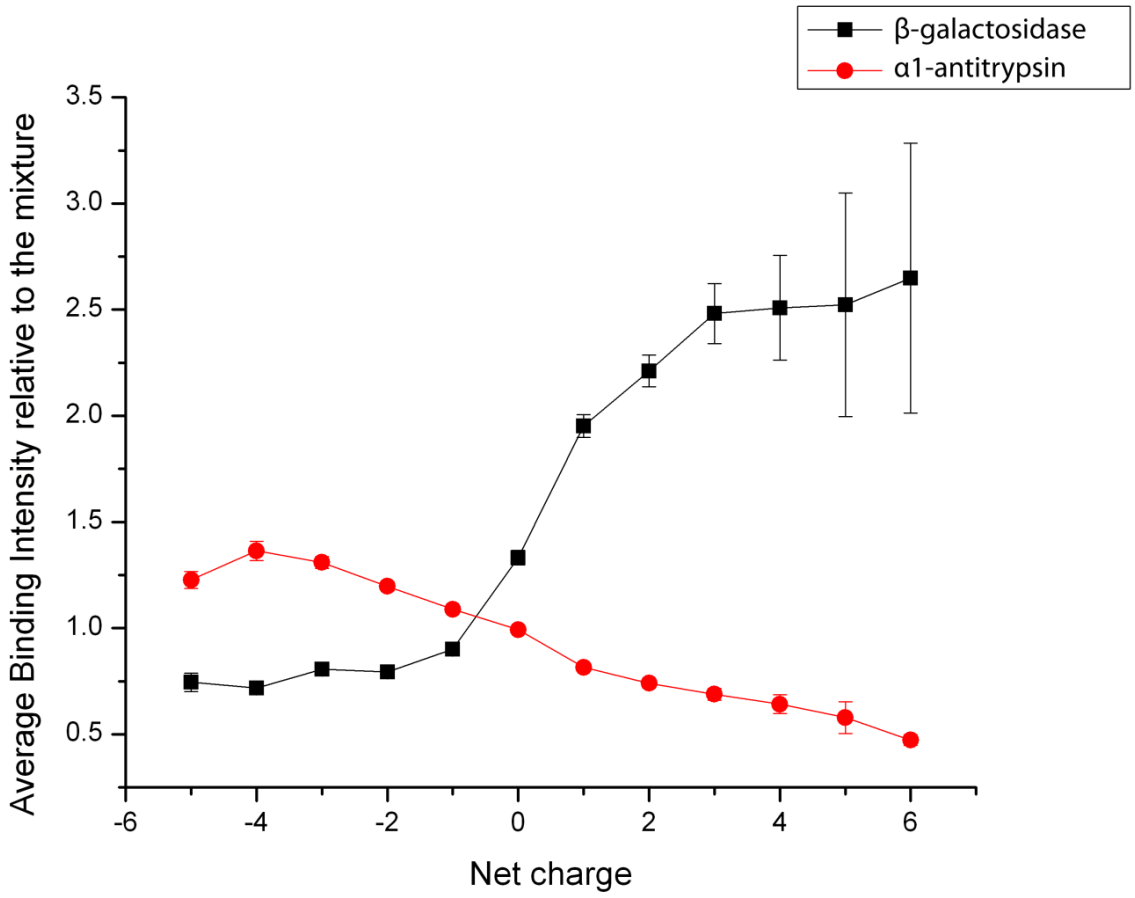

Figure 2.7. Relative binding discrimination as a function of net charge per peptide. The relative average binding discrimination (the ratio of average binding of β-galactosidase or α1-antitrypsin the average binding of the mixture of 9 proteins) is shown for peptides of each net charge group. The values are normalized to the global median before the ratio calculation. The error bars shown represent standard errors of the mean for each value.

Perhaps not surprisingly, given the charge effects described above and the hydrophobicity results of Figure 2, β-galactosidase binds most selectively to hydrophillic peptides while α1-antitrypsin binds most selectively to hydrophobic peptides (Figure S7).

DISCUSSION

The interaction of proteins with surfaces, and particularly with chemically complex surfaces such as the ones described here, is of considerable interest both in terms of what it potentially tells us about protein-surface interactions in cells and in terms of its potential for utility in developing surfaces with selective binding or nonbinding properties for practical use. In this context, there are several important observations about the dependence of protein/surface interactions on the properties of peptides tethered to a surface.

One observation is that even short peptides can have both strong and diverse interactions with proteins. Some of the strongest binding peptides are comprised of just a few amino acids (Table 2.2, Figure 2.3) and dynamic range among short peptides was almost as great as among long ones. The binding of short peptides is strongly dictated by charge, as evidenced by the fact that highly positively charged 2-4 amino acid peptides show some of the strongest binding, particularly for β-galactosidase (Figure 2.5).

More importantly, the peptides that interact with one protein are distinct from those that interact with others. The Venn diagram of Figure 6 shows that the overlap between the peptide features that bind the best to the three samples is minimal. The somewhat larger degree of overlap between β-galactosidase and the protein mixture arises because β-galactosidase is a large tetrameric protein with low pI (around 4.7) and therefore binds strongly to most positively charged peptides (all of the peptides that are in

common between β-galactosidase and the protein mixture are positive charged with two or more Arginine residues). Not only are the peptides that bind well different for different proteins, but their dependence on peptide charge, hydrophobicity and length is also unique. β-galactosidase is selectively favored by highly charged peptides between 4 and 9 amino acids in length when compared to the binding of a mixture of 9 proteins and disfavored by long, hydrophobic peptides; α1-antitrypsin behaves almost in the opposite fashion, preferring the longer, more hydrophobic peptides.

Interestingly, the isolectric point and the fraction of the surface area consisting of solvent accessible hydrophobic amino acid side chains (Table 2.1) are not that different between β-galactosidase and α1-antitrypsin. Both are strongly negative proteins at the pH tested and both have between 11 and 14% of the surface area composed of hydrophobic amino acids. Yet, in terms of their interactions with the surface, they behave very differently. This may have to do simply with the size difference; because of its larger size, there is more opportunity for weak, long-range interactions, like columbic interactions, to have an effect, whereas for a smaller protein, the shorter range interactions between hydrophobic regions excluding water may be more important.

With the growing commercial availability of both peptide arrays and the ability to create libraries of peptides on a wide variety of surfaces, it has become practical and cost effective to empirically tune the chemical properties of a surface by selecting appropriate molecules from surface-based molecular libraries of peptides or related heteropolymers. The number of companies that will generate peptide libraries on surfaces either via synthesis and printing or *in situ* synthesis is growing. While selection of peptides from small libraries is unlikely to provide antibody-level affinity and specificity, even very

simple sequences can provide quite selective chemical properties and interactions among multiple components, opening the possibility for using this approach to very simply and quickly create surfaces with specific properties relative to their interactions with complex mixtures such as blood as well as spatially patterned interactions. These experiments were designed to be interpreted at a high level, without granular dissection of amino acid sequences, but rather at the level of general peptide characteristics. Protein:peptide interactions can be mapped and described with a small number of peptides because the design of those peptides can be optimized to take advantage of this interaction level, rather than attempting to create peptides requiring a highly granular level of detail and sequence-specific analyses. Additionally, because these are entirely synthetic systems, a wide variety of non-natural chemical variations can be easily incorporated, providing surfaces with both a wider diversity of chemical characteristics as well as resistance to biological degradation, adding flexibility to a measurement system that is well-suited for generalizing binding interactions.

ACKNOWLEDGEMENTS

The authors would like to thank Drs. Matthew Greving, Phillip Stafford and Stephen Johnston for helpful discussions and input. This material is based in part upon work supported by the National Science Foundation under Grant No. (MCB-1243082).Supporting Information is available online from Wiley InterScience or from the author.

SUPPORTING INFORMATION

Distribution of peptide binding as a function of length

The peptide binding distribution as a function of length for α1-antitrpsin and the protein mixture are shown in figures S2.1 and figure S2.2. The distributions reflect the relative population of peptides that are bound to the labeled proteins at each length. As was the case for β-galactosidase, most peptides of all lengths show relatively weak binding but there are a small number that bind with high intensity.

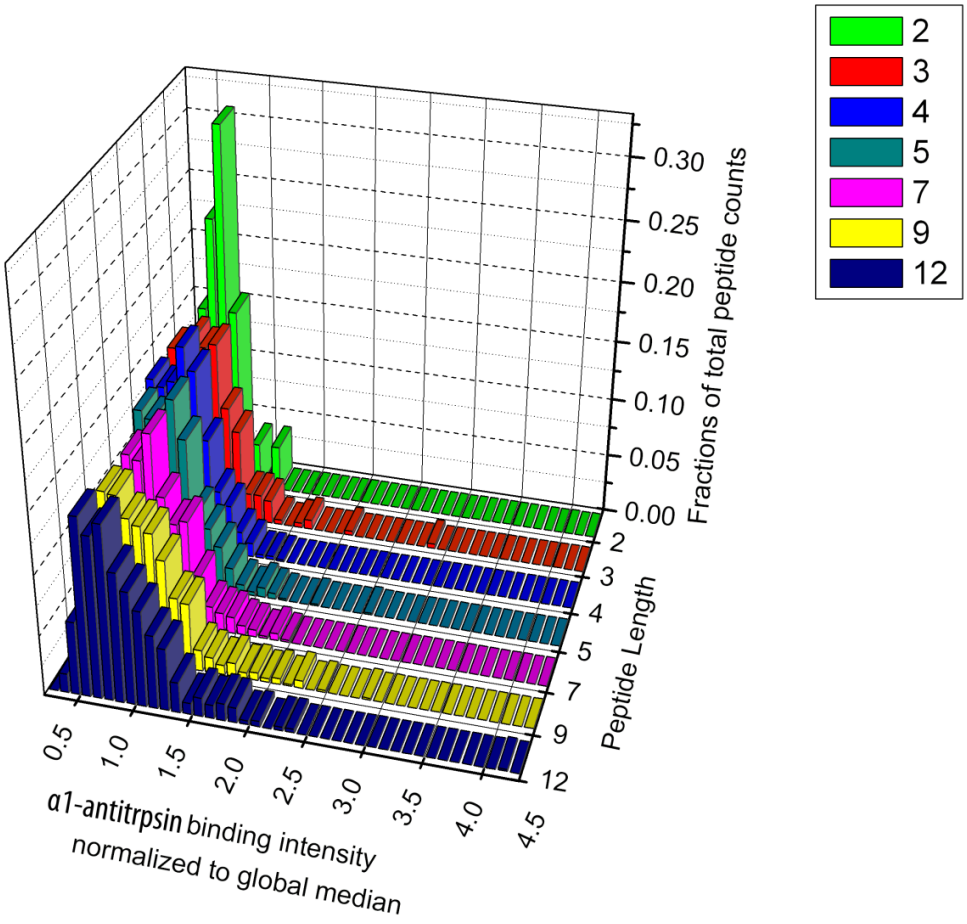

Figure S2.1 **Three dimensional histogram of α1-antitrpsin binding distribution as a function of peptide length.** Details as in Figure 4.

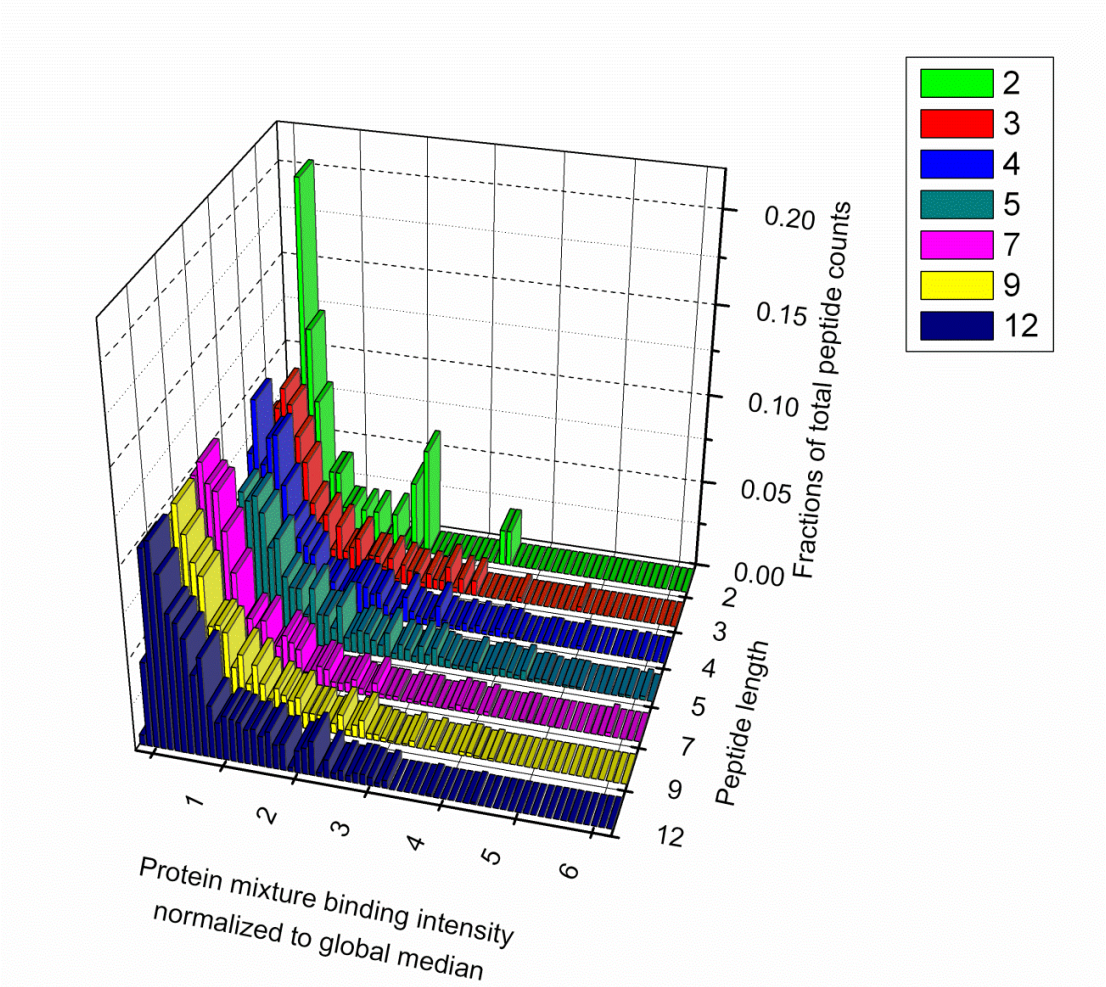

Figure S2.2 **Three dimensional histogram of protein mixture binding distribution as a function of peptide length.** Details as in Figure 2.4.

Length dependence for peptide with high binding signals

Figure S2.3 shows the average signal from only the top 30 binding peptides at each length, the length dependence is generally similar to that seen for the average of all peptides in Figure 2.3.

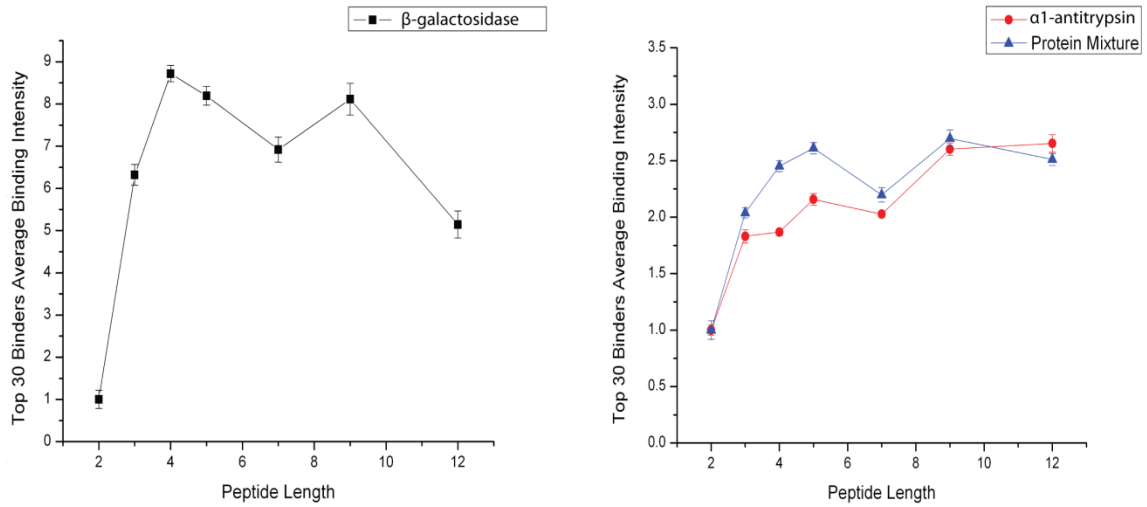

Figure S2.3. The average fluorescence intensity from the 30 peptides with the greatest affinity for a particular sample as a function of peptide length. The values shown are averages of the median-normalized binding intensity of peptides of a particular length. The error bars shown are calculated as standard error of the mean.

The relationship between length and charge

The relationship between peptide length and charge was studied by plotting the average binding intensity versus peptide length for each charge group. Figure S2.4 and Figure S2.5 show the length dependence of binding for positive, negative and neutral peptides for α1-antitrypsin and the protein mixture, respectively. The binding curves for both α1-antitrypsin and the protein mixture do not depend strongly on charge.

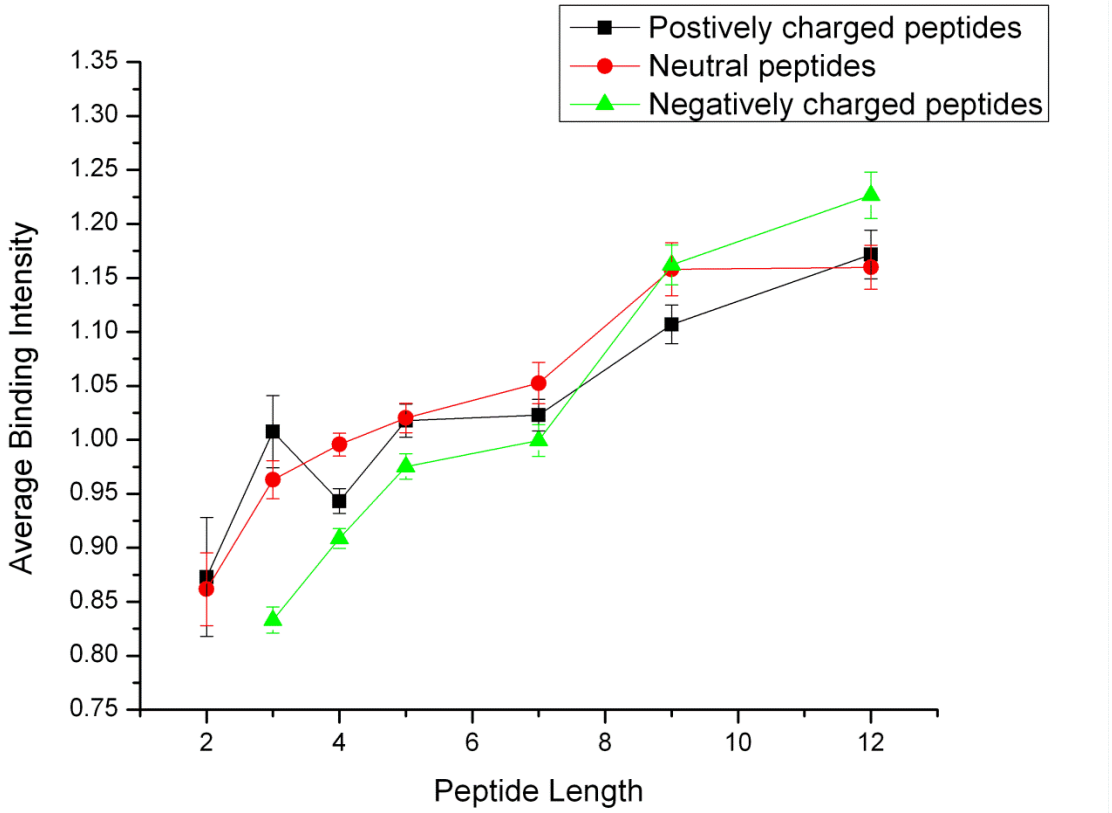

Figure S2.4 α1-antitrypsin average binding intensity versus peptide length for each charge group. Details are as in Figure 5 except that there were 1183 positively charged peptides (black squares), 1231 neutral peptides (red circles) and 971 negatively charged peptides (green triangles). The error bars shown represent the standard error in the mean for that peptide length.

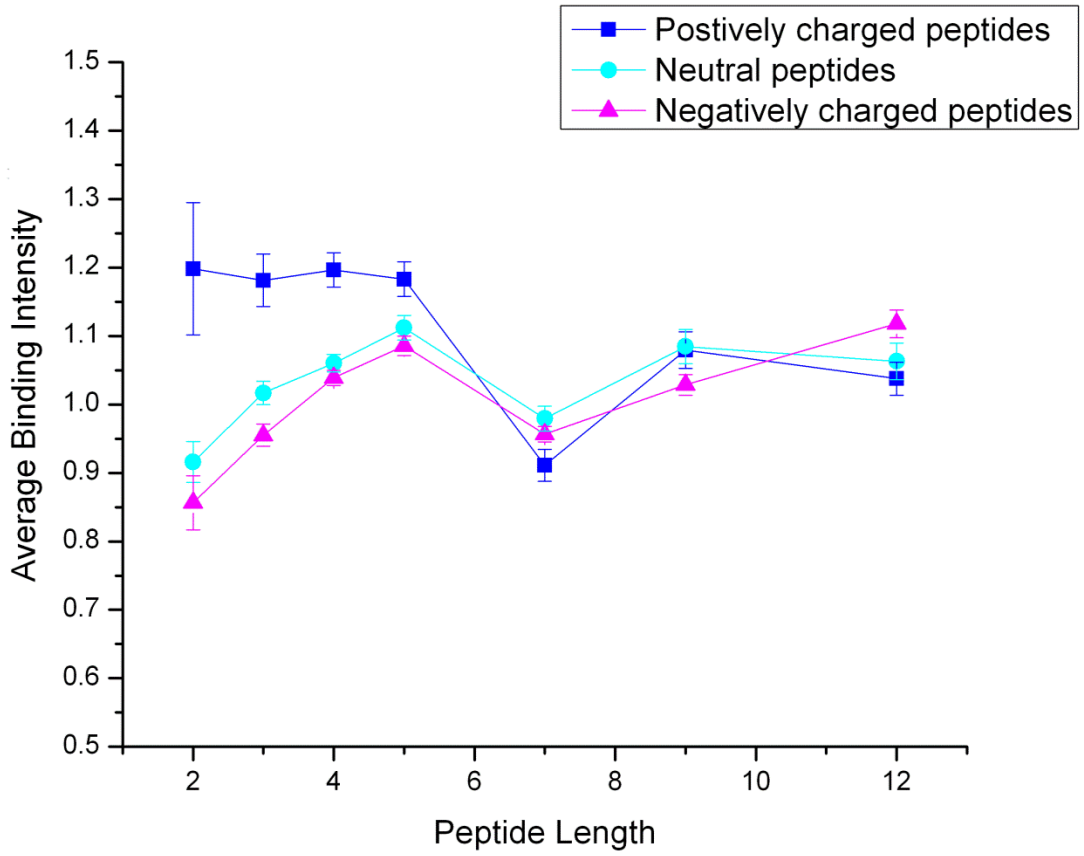

Figure S2.5 Protein mixture average binding intensity versus peptide length for each charge group. Details as in Figure 5, except that there were 1646 positively charged peptides (blue squares), 1539 neutral peptides (cyan circles) and 1690 negatively charged peptides (magenta triangles). The error bars shown represent the standard error in the mean for that peptide length.

The relationship between length and hydrophobicity

The average binding intensity for β-galactosidase versus peptide length for four hydrophobicity groups is shown in Figure S2.6. All four groups show lower binding intensities at long lengths.

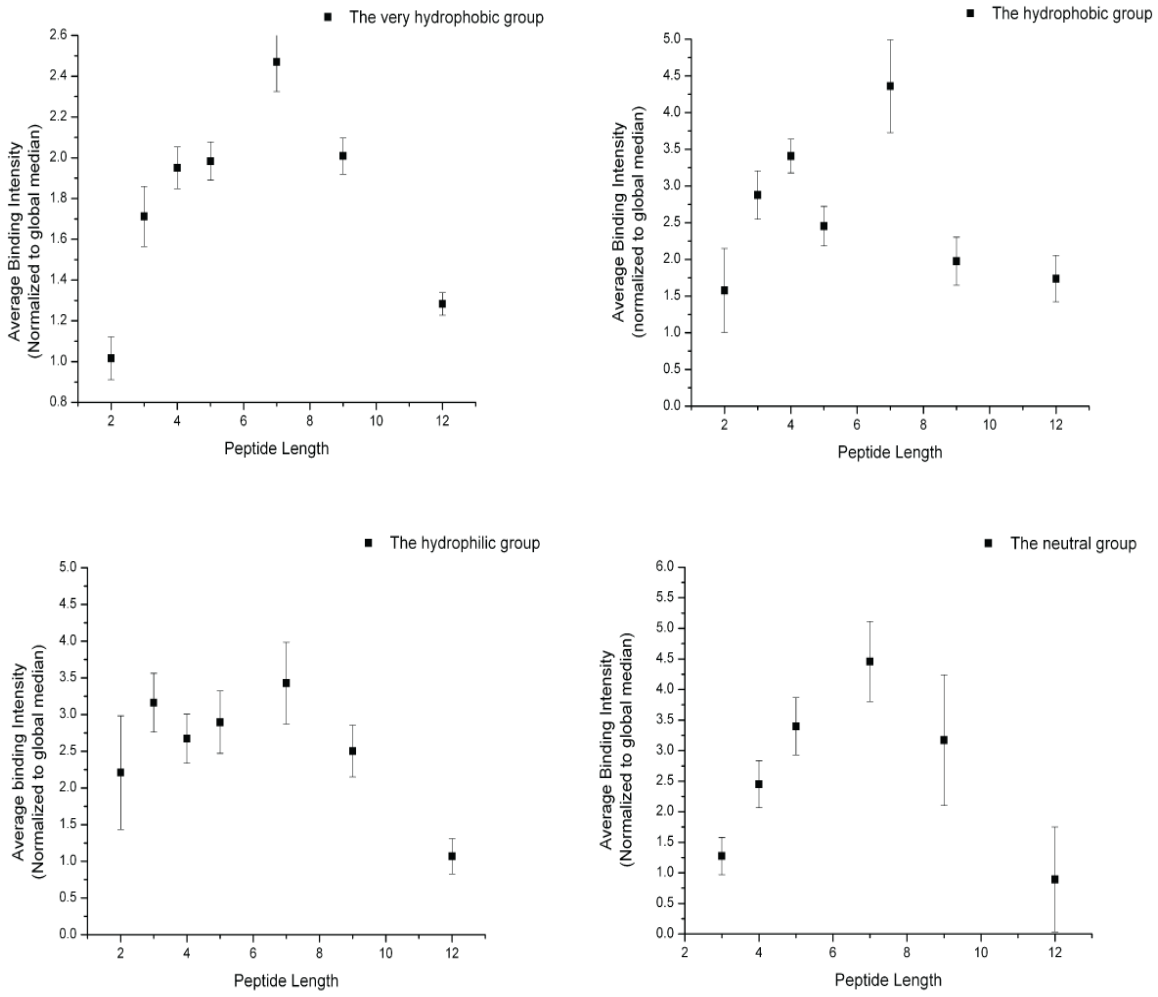

Figure S2.6 β-galactosidase average binding intensity versus peptide length for each hydrophobicity group. The binding intensities from peptides of each length were averaged together. There were 3014 peptides in the very hydrophobic group, 476 peptides in the hydrophobic group, 192 peptides in the neutral group and 296 peptides in the hydrophilic group. Intensity values shown are normalized to the global median for β-

galactosidase. The error bars shown represent the standard error in the mean for that peptide length.

Relative binding discrimination as a function of peptide hydrophobicity

Figure S2.7 shows the ratio between the binding of each of the two individual proteins (β-galactosidase and α1-antitrypsin) and the mixture as a function of hydrophobicity. The results suggest that β-galactosidase binds most selectively to hydrophillic peptides while α1-antitrypsin binds most selectively to hydrophobic peptides.

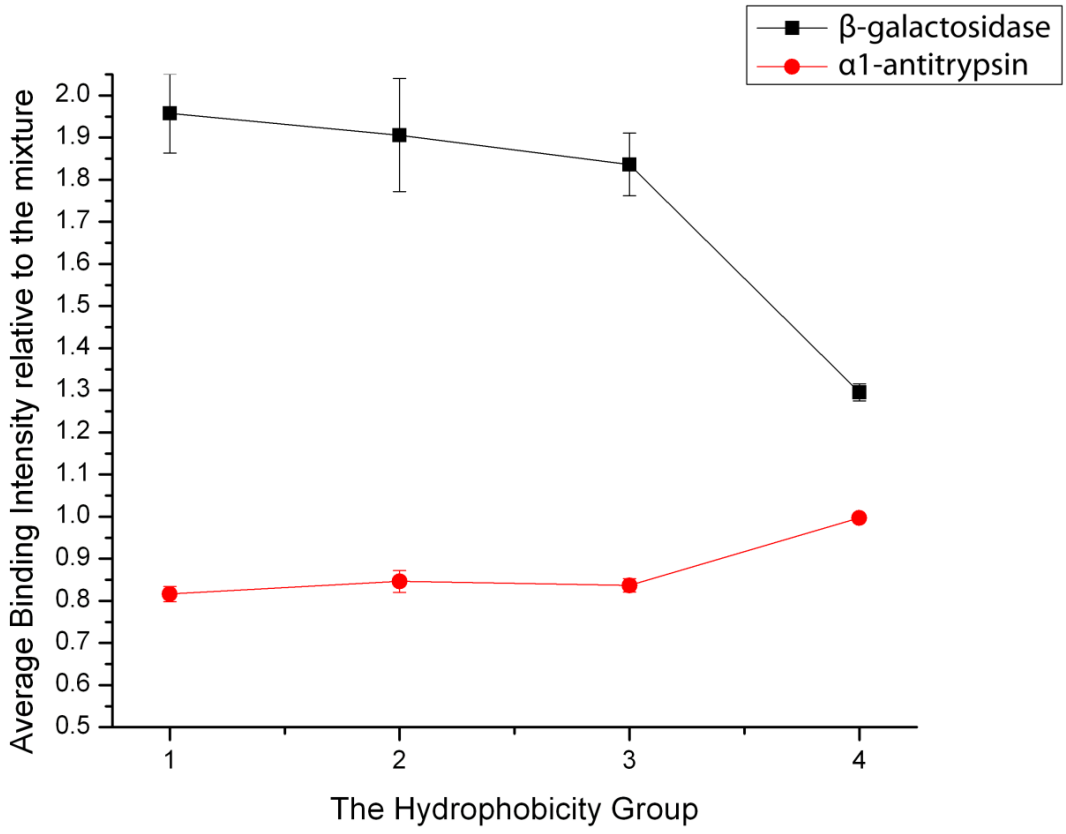

Figure S2.7. Relative binding discrimination as a function of peptide hydrophobicity. The relative average binding discrimination (see Fig. 2.7) between peptides of each hydrophobicity group (defined in Fig. 2.2) are shown for β-galactosidase and α1-antitrypsin. The error bars represent standard errors of the mean for each value.

Comparing β-galactosidase binding on two different peptide-conjugated surfaces

Prior to the protein binding experiment on the 30K LC Sciences array, similar binding experiments were performed on the same type array to both confirm that the sequences on the LC Science arrays were as advertised and that the nature of the binding surface itself does not dominate the results. To accomplish this, 14 peptides were selected

from an LC Sciences array, synthesized and purified. These were bound to microwell plates, protein was incubated with each peptide and the level of protein bound was measured on a SpectraMax M5 96 well plate reader (Molecular Device, Sunnyvale, CA) using a slightly modified version of a procedure previously published[33]. First, 30 µL of SMCC solution (10mM in 1× PBS buffer, pH 7.4) was added into each aminated microwell and incubated for one hour at room temperature to activate the surface. The microwell plate was briefly washed with pure water three times to get rid of extra SMCC. Second, 30 µL peptide solution (300 µM in 1× PBS buffer, 1 mM TCEP, pH 7.4) was added to the appropriate SMCC-activated microwells. The reaction was incubated for 4 hours at room temperature, in the dark to allow the peptide to conjugate to the surface. After the conjugation reaction was complete, the microwells were washed for 5 minutes in 1× TBST, three times, followed by three washes in water to remove free peptides. Third, 30 µL of biotin-labeled β-galactosidase (10nM in 10 mM phosphate buffer with 0.05% Tween 20 (v/v%), pH 7.4) was then added to corresponding wells and incubated for two hours at room temperature. The microwells were then washed for 5 minutes in 1× TBST, three times, followed by three washes in phosphate. Finally, β-galactosidase binding level was measured using an enzyme linked immunosorbant assay (ELISA). 30 µL of Alkaline phosphatase-conjugated strepavidin solution (0.4 mg/ml diluted at 1:1000 in 1× PBS, 0.05% (v/v) Tween 20) was added to the biotin-β-galactosidase-bound wells and incubated for one hour at room temperature. The streptavidin solution was then removed and the plate was washed three times with TBST buffer and three times with TBS buffer. Then, 200 µL of 1 mM PNPP was added to each well. The alkaline phosphatase activity was subsequently measured by reading the absorbance increase at

405 nm on the M5 plate reader. The β-galactosidase binding level was determined from the activity of alkaline phosphatase-conjugated strepavidin bound to the wells.

Table S2.1 shows the rank of protein binding level to each selected peptide sequence. Although the rank order varies somewhat, the stronger binders stay strong and the weaker binders stay weak on both binding platforms. This confirms that the measured binding signals reflect the binding levels between β-galactosidase and the surface-bound peptides, rather than the surface itself.

Table S2.1: The rank of β-galactosidase binding level to each selected peptide sequence

Sequence	Rank in LC Science platform	Rank in microwell platform
RYYSSRLRYGSG	1	1
SRYYGSG	2	4
YRSRYRQQQGSG	3	5
QRYYGSG	4	3
YRYYSGSG	5	2
WRYYQRSQYGSG	6	7
QRYQRSRSYGSG	7	8
RQYSGSG	8	6
YQWRRGSG	9	9
RLERSRQERGSG	10	11
RRQEGSG	11	10
EQRYREEREGSG	12	12
QLSEESELLGSG	13	14
RWESGSG	14	13

REFERENCES

[1] Ostuni E, Yan L, Whitesides GM. The interaction of proteins and cells with self-assembled monolayers of alkanethiolates on gold and silver. Colloid Surface B 1999;15:3-30.

[2] Johnson ME, Hummer G. Nonspecific binding limits the number of proteins in a cell and shapes their interaction networks. P Natl Acad Sci USA 2011;108:603-8.

[3] Andresen H, Grotzinger C. Deciphering the Antibodyome - Peptide Arrays for Serum Antibody Biomarker Diagnostics. Curr Proteomics 2009;6:1-12.

[4] Lee MH, Brass DA, Morris R, Composto RJ, Ducheyne P. The effect of non-specific interactions on cellular adhesion using model surfaces. Biomaterials 2005;26:1721-30.

[5] Vonhippe.Ph, Mcghee JD. DNA-Protein Interactions. Annu Rev Biochem 1972;41:231-&.

[6] Halford SE, Marko JF. How do site-specific DNA-binding proteins find their targets? Nucleic Acids Res 2004;32:3040-52.

[7] Oda M, Furukawa K, Ogata K, Sarai A, Nakamura H. Thermodynamics of specific and non-specific DNA binding by the c-Myb DNA-binding domain. J Mol Biol 1998;276:571-90.

[8] Woodson M, Liu J. Functional nanostructures from surface chemistry patterning. Phys Chem Chem Phys 2007;9:207-25.

[9] Gilles S, Winter S, Michael KE, Meffert SH, Li PG, Greben K, et al. Control of Cell Adhesion and Neurite Outgrowth by Patterned Gold Nanoparticles with Tunable Attractive or Repulsive Surface Properties. Small 2012;8:3357-67.

[10] Yu TY, Wang Q, Johnson DS, Wang MD, Ober CK. Functional hydrogel surfaces: Binding kinesin-based molecular motor proteins to selected patterned sites. Adv Funct Mater 2005;15:1303-9.

[11] Sykes KF, Legutki JB, Stafford P. Immunosignaturing: a critical review. Trends Biotechnol 2013;31:45-51.

[12] Hughes AK, Cichacz Z, Scheck A, Coons SW, Johnston SA, Stafford P. Immunosignaturing can detect products from molecular markers in brain cancer. Plos One 2012;7:e40201.

[13] http://www.lcsciences.com/applications/proteomics/binding-arrays/.

[14] Greving MP. Creating high-affinity ligands on surfaces and in solution: Arizona State University; 2009.

[15] Kyte J, Doolittle RF. A Simple Method for Displaying the Hydropathic Character of a Protein. J Mol Biol 1982;157:105-32.

[16] Pellois JP, Zhou XC, Srivannavit O, Zhou TC, Gulari E, Gao XL. Individually addressable parallel peptide synthesis on microchips. Nat Biotechnol 2002;20:922-6.

[17] Pilipenko OS, Atyaksheva LF, Poltorak OM, Chukhrai ES. Dissociation and catalytic activity of oligomer forms of beta-galactosidases. Russ J Phys Chem a+ 2007;81:990-4.

[18] Wallenfels K, Weil R. The Enzymes. P.D. Boyer ed1972. p. 617.

[19] Brown WT. Alpha-1-Antitrypsin - Apparent Molecular-Weight Heterogeneity Shown by Two-Dimensional Electrophoresis. Am J Hum Genet 1982;34:195-208.

[20] Spiro RG. Studies on Fetuin, a Glycoprotein of Fetal Serum .1. Isolation, Chemical Composition, and Physicochemical Properties. J Biol Chem 1960;235:2860-9.

[21] Welinder KG. Amino-Acid Sequence Studies of Horseradish-Peroxidase .4. Amino and Carboxyl Termini, Cyanogen-Bromide and Tryptic Fragments, the Complete Sequence, and Some Structural Characteristics of Horseradish Peroxidase-C. Eur J Biochem 1979;96:483-502.

[22] Shannon LM, Kay E, Lew JY. Peroxidase Isozymes from Horseradish Roots .I. Isolation and Physical Properties. J Biol Chem 1966;241:2166-&.

[23] Hirayama K, Akashi S, Furuya M, Fukuhara K. Rapid Confirmation and Revision of the Primary Structure of Bovine Serum-Albumin by Esims and Frit-Fab Lc Ms. Biochem Bioph Res Co 1990;173:639-46.

[24] Kaplan LJ, Foster JF. Isoelectric Focusing Behavior of Bovine Plasma Albumin, Mercaptalbumin, and Beta-Lactoglobulins-a and B. Biochemistry-Us 1971;10:630-&.

[25] Lindskog S. Purification and Properties of Bovine Erythrocyte Carbonic Anhydrase. Biochim Biophys Acta 1960;39:218-26.

[26] Betgovargez E, Knudson V, Simonian MH. Characterization of proteins in the human serum proteome. J Biomol Tech 2005;16:306-10.

[27] Najafi MF, Deobagkar D. Purification and characterization of an extracellular alpha-amylase from Bacillus subtilis AX20. Protein Expr Purif 2005;41:349-54.

[28] Tsuge H, Natsuaki O, Ohashi K. Purification, Properties, and Molecular Features of Glucose Oxidase from Aspergillus-Niger. J Biochem-Tokyo 1975;78:835-43.

[29] Pazur JH, Kleppe K. The Oxidation of Glucose and Related Compounds by Glucose Oxidase from Aspergillus Niger. Biochemistry-Us 1964;3:578-83.

[30] Diehnelt CW, Shah M, Gupta N, Belcher PE, Greving MP, Stafford P, et al. Discovery of High-Affinity Protein Binding Ligands - Backwards. Plos One 2010;5.

[31] Oliveros JC. VENNY. An interactive tool for comparing lists with Venn Diagrams.http://bioinfogp.cnb.csic.es/tools/venny/index.html. 2007.

[32] Monera OD, Sereda TJ, Zhou NE, Kay CM, Hodges RS. Relationship of Sidechain Hydrophobicity and alpha-Helical Propensity on the Stability of the Single-stranded Amphipathic alpha-Helix. J Pept Sci 1995;1:319-29.

[33] Fu JL, Reinhold J, Woodbury NW. Peptide-Modified Surfaces for Enzyme Immobilization. Plos One 2011;6.

CHAPTER 3: UNSTRUCTURED INTERACTIONS BETWEEN PEPTIDES AND PROTEINS: EXPLORING THE ROLE OF SEQUENCE MOTIFS IN AFFINITY AND SPECIFICITY

ABSTRACT

Unstructured interactions between proteins and other molecules or surfaces are often described as nonspecific, and have received relatively little attention in terms of their role in biology. However, despite their lack of a specific binding structure, these unstructured interactions can in fact be very selective. The lack of a specific structure for these interactions makes them more difficult to study in a chemically meaningful way, but one approach is statistical, simply looking at a large number of different ligands and using that to understand the chemistry of binding. Surface-bound peptide arrays are useful in this regard, and have been used as a model previously for this purpose (Wang et al., Acta Biomater. 2014;10:761-8). In that study, the binding of several proteins, including β-galactosidase, to all possible di-peptides, tri-peptides and tetra-peptides (using 7 selected amino acids) was performed and analyzed in terms of the charge characteristics, hydrophobicity, etc. of the binding interaction. The current work builds upon that study by starting with a representative subset of the tetrapeptides characterized previously and either extending them by adding all possible combinations of one, two and three amino acids, or by concatenating 57 of the previously characterized tetrapeptides to each other in all possible combinations (including order). The extended and concatenated libraries were analyzed by binding either labeled β-galactosidase to them or by binding a mixture of ten different labeled proteins of various sizes, hydrophobicities and charge characteristics to the peptide arrays. By comparing the binding signals from the

tetrapeptides or amino acid extensions alone to the binding signals from the complete extended or concatenated sequences, it was possible to evaluate the extent to which affinity and specificity of the whole sequence depends on the subsequences that make it up. The conclusion is that while joining two component sequences together can either greatly increase or decrease overall binding and specificity (relative to the component sequences alone), the contribution to the binding affinity and specificity of the individual binding components is strongly dependent on their position in the peptide; component sequences that bind strongly at the C-terminus of the peptide do not necessarily add substantially to binding and specificity when placed at the N-terminus. This work has been submitted to *Acta Biomaterialia* and is under review.

INTRODUCTION

When one thinks about the kinds of interactions proteins have in biological systems, generally what comes to mind are specific interactions: antibody-antigen binding, receptor-ligand binding, enzyme substrate binding, etc. However, given the extraordinarily crowded environment of the cell, the vast majority of the encounters that proteins have with each other or with the boundaries of the compartments that surround them are nonspecific. Here nonspecific interactions might also be referred to as unstructured interactions, interactions that do not involve the formation of a stable complex that has a homogeneous structure. Most biological processes comprise large, intricate interaction networks which include both specific and nonspecific interactions[1, 2] and some of these processes cannot function properly without the participation of nonspecific interactions. For example, nonspecific (unstructured) interactions between proteins (or enzymes) and the nucleic acids are important determinants of biological

function[3] . The initial, unstructured interactions of proteins with DNA or RNA can help to facilitate binding to specific sites by reducing the dimensionality of diffusion in DNA replication and DNA modification, by forming distorted binding geometries to activate transcription processes, or even by more indirect regulation of gene expression [4][5][6]. While most of these interactions do not involve the initial formation of specific structures, they can in fact be both quite strong and quite selective in terms of biological function.

Other examples of biological processes that emphasize the nature of unstructured interactions include proteins selectively NOT adsorbing to other macromolecules in crowded cellular environments[7-9] and Phosphatidylinositol transfer proteins (PITPs) localizing to the trans-Golgi network by both specific and nonspecific membrane-binding components[10]. Even the transfer of substrates between enzymes can utilize unstructured interactions to keep the local concentration of product/substrate high and facilitate transfer [11]. Indeed, protein interactions in biology almost certainly represent a continuum between structurally defined interactions with long residence times and transient unstructured interactions.

Of particular significance, both to our understanding of fundamental biology and to practical application, are nonspecific or unstructured interactions between proteins and complex surfaces. Though ubiquitous, these interactions are difficult to characterize for the very reason that they lack specific structures. Further, the interactions, often involve a protein and multiple molecules or parts of molecules on a surface. This can be quite different from a situation where one protein is interacting with one ligand; it is more one surface interacting with another. Though poorly defined structurally, these types of

interactions are clearly selective. Nature uses selective, unstructured interactions between protein surfaces and complex surfaces in their surroundings to its advantage and mimicking that kind of selectivity is a key aspect of developing biomaterials or biosensors with surfaces that avoid background binding and interference, but facilitate selected interactions[7, 8].

Biological surfaces are very complex and clearly have been tuned to encourage certain types of interactions, both structured and unstructured, in particular places. Thus far, the design and fabrication of artificial surfaces that interact with proteins has not taken full advantage of this kind of specific chemical tuning of interactions, particularly in a position dependent way. Previously, this laboratory has investigated the nature of nonspecific or unstructured binding between proteins and surfaces covered with particular peptide sequences by exposing *in situ* synthesized peptide arrays to protein solutions. Such arrays provide effective tools for characterizing the interaction of biological macromolecules with surfaces because they present a set of sequences with great chemical complexity that can be precisely controlled. In that study, two individual proteins, β-galactosidase and α1-antitrypsin, as well as a mixture of 9 different proteins, were bound to arrays of nearly 5000 different short peptides with a wide range of charge characteristics, hydrophobicities and peptide lengths[12]. Interestingly, even very short sequences (tripeptides, tetrapeptides) covering surfaces discriminate strongly between different proteins or between individual proteins and mixtures of proteins. While the general charge and hydrophobicity of the peptides are certainly important factors, more subtle aspects of the sequence are also important in driving the relative ability to bind one protein over another. From a practical point of view, the work demonstrated that it was

possible to simply and efficiently select short peptide sequences, directly on a surface, that dictated whether a particular protein would bind or not.

The fact that even short surface-bound peptides can have both strong and selective interactions with proteins is important from a practical point of view; short peptides are relatively easy and inexpensive to use for surface modification. Indeed, it is not unreasonable, given the size of peptide arrays now available, to perform essentially complete searches of tri-peptide and tetra-peptide and even penta-peptide sequence spaces to find the desired surface binding characteristics for the specific proteins, mixtures of proteins and/or conditions of interest. This brings up the question of to what extent the properties of short peptide sequences can be added together, creating longer sequences with composite binding function. For example, if one stacks two tetra-peptides that both bind a particular protein selectively, does the affinity and selectivity increase accordingly? Similarly, if one builds a new peptide library on top of a short peptide sequence with known binding characteristics, does one find that the same peptide sequences that work well at the surface in enhancing interaction also serves the same purpose at the N-terminus of a nascent peptide chain? Here, a series of tetra-peptide motifs, identified in the previous work cited above, and known to bind β-galactosidase, will be examined as individual motifs, in combination with each other, and after extension with a library of additional sequences.

It is important to note that the goal of the work described here is quite different from a study in which one is selecting a strong-binding ligand from a large library of possible peptides. The goal is instead to understand how modifying surfaces with a

relatively dense attachment of peptide sequences changes its affinity for proteins and protein mixtures, and how this depends on the details of the peptide sequences.

MATERIALS AND METHODS

Peptide microarray fabrication

The peptide microarrays used here were fabricated by LC Sciences (Houston, TX) using PepArray™ technology. The technology is developed based on proprietary μParaflo® microfluidics technology, which allows direct synthesis of peptides on a high density microfluidic chip (4000 features in a ~1.5cm^2 area) as features in specific locations on the chip using a photogenerated acid (PGA) to deprotect the amines of nascent peptides and conventional t-boc solid phase peptide synthesis [13, 14]. The chip is actually an enclosed microfluidic system, which contains fluid distribution channels and picoliter scale reaction chambers with physical isolation from each other[13]. Each chamber contains one specific peptide sequence. The in-situ peptide synthesis consists of the following steps: (1) derivatization of the surface with a protected NH2-linker, resulting in a surface density of less than 1pmol per 1mm^2 area; (2) applying a solution of the PGA precursor in dichloromethane; (3) deprotecting the protected amine in the desired reaction chamber using digital photolithography, which allows programmable light activation of chemical reactions; (4) coupling the amino acid, capping any unreacted linkers and deprotecting the side chain using standard peptide synthesis procedures [14-17].

Peptide libraries synthesized on peptide array surfaces

Two different peptide libraries were designed and synthesized on the arrays to study the relationship between sequence position within a peptide and binding affinity for β-galactosidase (see Results). The same set of 7 amino acids {E, L, S, R, Q, W, Y} as in the previous studies was used to synthesis all peptides [12, 18, 19]. Each peptide library contains 3918 custom-made peptides (including test sequences and control sequences), up to 8 amino acids in length, and 82 additional control spots (including blank and LC Science internal controls) provided by the company. Two identical arrays were made for each peptide library and each was bound under the same binding condition, but with samples labeled with different dyes (see below). A GSG tri-peptide linker was added to the C terminus of each sequence to maintain a uniform linkage between the peptides and the array surfaces in addition to the proprietary surface linker (20-30nm in length) used by LC Sciences.

The first set of peptide arrays ("extension arrays") included peptide sequences with the following motifs: "(N') X abcd GSG (C')"; "(N') XX abcd GSG(C')"; "(N') XXX abcd GSG(C')". The C-terminal "abcd GSG" sequence was one of 8 sequences selected from a peptide library that included all possible variations of the tetrapeptide sequence "abcd" created in a previous study using the same sub-set of 7 amino acids specified above[12]. The sequences used are given in Table 3.1. For both the previous array and the arrays described here, the peptide sequences are attached to the surface via the C-terminal glycine. These selected tetrapeptides (actually a selected tetrapeptide plus the GSG linker) were chosen so that between them they cover a range of affinities to β-galactosidase. The N-terminal "X, XX and XXX" segments constitute all possible

sequence combinations with one amino acid (7 combinations), two amino acids (49 combinations) and three amino acids (343 combinations), using the 7 different amino acids specified above. All possible combinations of the eight different "abcd" tetrapeptide sequences and the "X, XX, XXX" N-terminal sequences were synthesized in the same arrays. In addition, all of the one amino acid, two amino acid and three amino acid N terminal extensions themselves were synthesized directly on a GSG linker (e.g., "(N') XXX GSG (C')") as were the eight base tetrapeptide sequences without further extention ("(N') abcd GSG (C')"). The base tetrapeptide sequences were synthesized with 10 replicates. The library also contains 319 peptides randomly selected from previous arrays to monitor the chip-to-chip variation.

The second set of peptide arrays ("concatenation arrays") included peptide sequences with the motif: "(N') xxxx yyyy GSG (C')". Both of the "xxxx" and "yyyy" were sequences chosen from 57 tetrapeptides again selected from the previous study of all tetrapeptide sequences. The 57 sequences used in this study were chosen by performing 'K-mean' clustering of the binding data from all tetrapeptide sequences into 6 groups depending on the level of apparent affinity and relative specificity to β-galactosidase, and then selecting sequences that covered all 6 groups. The sequences of the peptides that make up the arrays in the second set consist all possible combinations of two of the 57 tetrapeptide peptide sequences, spliced together in the motif described above. The individual 57 tetrapeptides used in this study were also included on the array as controls in the form "(N') xxxx GSG (C')". In addition, the library contained 669 peptides randomly selected from previous arrays to monitor the chip-to-chip variation and normalization.

Protein binding

Alexa Fluor®-555 and Alexa Fluor®-647 (AF555 and AF647, Invitrogen/Life Technologies) labeled proteins were used for the protein binding assays. The labeling process followed the standard protocol provided by the manufacturer for the dye (Invitrogen/Life Technologies). Proteins used in the experiments here were the same as the proteins used in previous studies with the exception of α1-antitrypsin, which was added to the protein mixture instead of being tested as an individual protein [12]. (Note that while α1-antitrypsin is known to bind its reactive center loop (FLGAIG) peptide [20-23], and one might be concerned this specific binding would bias the peptide binding of the protein mixture, previous work from this lab [12] has not shown such a peptide bias in these arrays and in fact the dependence of binding on charge and hydrophobicity of this protein is very similar to the overall protein mixture). Two different binding concentrations, 10nM and 100nM, were tested for β-galactosidase (from *E.coli*). The concentration for each protein in the mixture was then determined using the same approach as in previous experiments [12]. The dye/protein labeling ratio for each protein was also maintained as previously. The fluorescent signal from the bound labeled proteins was measured and used as an indicator for peptide protein binding activity (referred to as the binding signal below). The concept of dye-swapping (labeling with two dyes and testing to see if binding is driven by the specific dye or by the protein) was employed as previously [12]. Protein samples separately labeled with AF555 and AF647 were applied on identical arrays respectively under the same experimental conditions. Very similar results were obtained regardless of which dye was used (e.g., Figures S3, S4, S5 and S6). For the extension study, binding signals from the array with AF555

labeled β-galactosidase were used, due to a problem with one of the AF647 arrays, while for the concatenation studies, binding signals from proteins with the two dye labels were averaged.

The actual protein binding assays were performed by LC sciences using the following procedures and conditions (procedures and conditions are from LC Sciences final customer reports). First, the chips were blocked in the blocking buffer (super block, a proprietary blocking solution, with 0.05% Tween-20 and 0.05% Triton X-100, pH7.0), overnight at 4ºC, to minimize non-specific binding and then washed in washing buffer (1×PBS, pH 7.0). Second, two different protein samples (AF555-β-galactosidase and AF647-β-galactosidase,) were bound to two identical arrays respectively for 1 hour in binding buffer (1×PBS, pH7.4) at 25°C, then washed with washing buffer. The array was then imaged using either 635 nm excitation and a 655-695 nm emission filter or 532 nm excitation and a 550-600 nm emission filter, depending on the dye used. The arrays were adapted to fit into standard *Molecular Devices* GenePix scanner and the PMT voltage level was adjusted based on a pre-scan image. After recording the first image, the chip surface was stripped with stripping buffer cocktail (pH7.0) at 25°C for 2 hours to remove surface-bound proteins, and then washed in washing buffer. The chips were imaged again at the same wavelength to confirm the stripping. Fourth, the AF555-protein mixture was bound to the stripped array that had AF647-β-galactosidase and the AF647-protein mixture was bound to the stripped arrays that had AF555-β-galactosidase for 1 hour in binding buffer at 25°C, then washed with washing buffer. The array was then re-imaged for AF647 fluorescence. Note that an advantage of using β-galactosidase for these studies is its large size. There should be very little interaction between dyes on different

proteins. As a result, binding signals should be linear with the concentration of β-galactosidase.

Data analysis

The binding between protein and peptides is measured as the fluorescence intensities of dye-labeled proteins captured by surface bound peptides. Higher fluorescence intensities suggest higher protein binding to a particular peptide feature. The term "binding signal" is used in all subsequent descriptions to refer to the measured fluorescence intensity. Peptide binding signal values were provided for each feature by LC Sciences as the median fluorescence intensity for each peptide feature. The background level, which is generally ~5% -10% of the lowest binding signal, depending on the sample used, was also provided by the company in the data report. For all of the analysis reported here, the binding signal for each peptide was normalized to the median intensity of each array [24]. Binding signals from arrays bound with same sample but using different dye labels were averaged and the mean of the binding signals was used in further analyses. Locally written scripts in Matlab were used to sort the peptides and to perform statistics.

RESULTS

Tetrapeptide extension studies

A set of arrays ("extension arrays") was designed to explore how the affinity of a tetrapeptide (the "base tetrapeptide) for β-galactosidase changes when it is extended by all possible one, two or three amino acid sequences (the "extension sequence). Specifically the question is how the binding properties of the base tetrapeptide alone

combine with the binding properties of the extension sequence to generate the binding properties of the extended peptide sequence (the "extended peptide sequence" is the base tetrapeptide plus the extension sequence). The binding signal distribution after incubating β-galactosidase with these arrays is shown Figure 3.1. As can be seen, in most cases, there is a rather wide distribution of binding signals associated with each family of extended peptide sequences (a family being defined by its base tetrapeptide sequence).

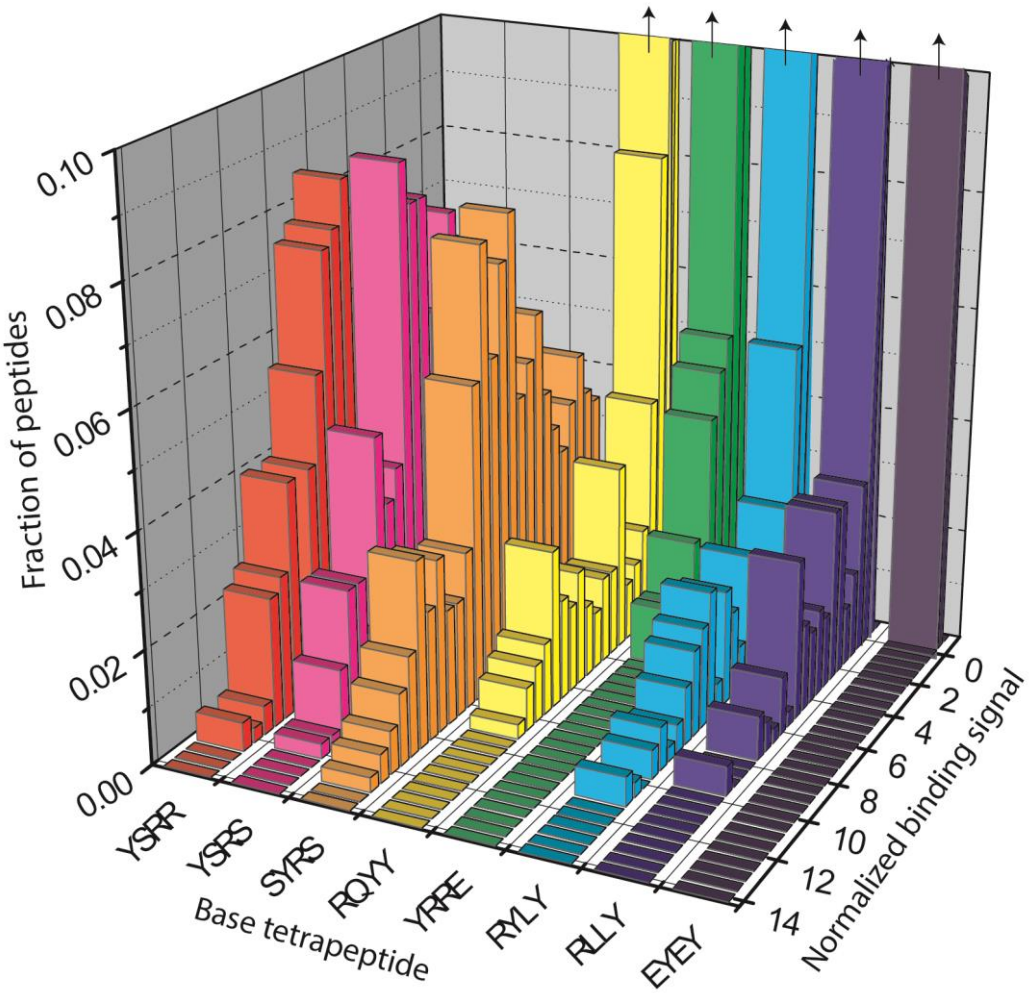

Figure 3.1. Three dimensional histogram of β-galactosidase binding signal distribution resulting from the extension of selected 8 tetrapeptides with 1, 2 or 3 amino acids. For the base tetrapeptdies EYEY, RLLY, RYLY, YRRE and RQYY, the

peptides with low binding signal (essentially background) dominate and have fractional values exceeding the height of the y-axis (indicated by arrows)**.** For example, for the base tetrapeptide EYEY, all of the extension peptides are in the "0" bin because their binding is at background levels (this bin has a vertical value of 1). The tetrapeptide groups were ordered by their inherent binding affinities to β-galactosidase without extension (highest on left). The binding signals for the base tetrapeptide sequences alone are shown in Table 3.1, but in each case they are near the median of the distributions shown. All binding signal values were normalized to the global median.

Table 3.1 shows that, roughly speaking, the average binding signal of all extended peptide sequences derived from a particular base tetrapeptide has a value similar to the binding signal of the base tetrapeptide sequence itself. Another way to visualize these results is shown in Figure 3.2. Here, the average β-galactosidase binding signal for all extended peptide sequences with the same base tetrapeptide sequence is plotted versus the binding signal of the base tetrapeptide sequence alone. Apparently, on average, addition of an extension sequence to the base tetrapeptide is as likely to reduce as to increase the affinity of the base tetrapeptide sequence for β-galactosidase.

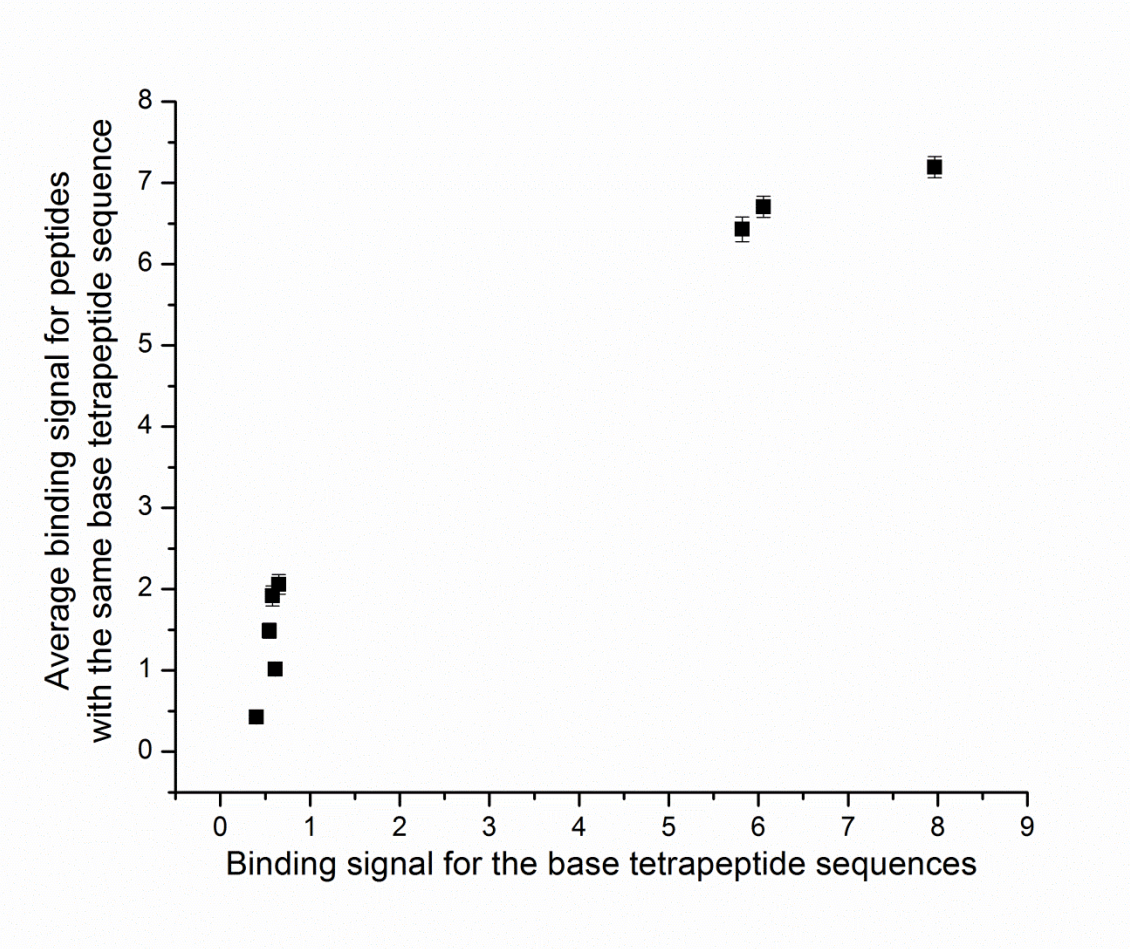

Figure 3.2. β-galactosidase average binding signal for the extended peptide sequences vs the inherent binding signal of the base tetrapeptide sequence alone. The fluorescence intensities (AF555- β-galactosidase) of the 399 extended peptide sequences for each base tetrapeptide sequence were averaged. All intensities were normalized to the global median. The error bars shown represent the standard error of the mean for that base tetrapeptide group.

Even though the base tetrapeptide appears to dictate the <u>average</u> binding to β-galactosidase of all the extended tetrapeptides, it is clear from Figure 3.1 that some extended tetrapeptides bound much more strongly or weaklly than the corresponding non-extended tetrapeptide. Table 3.1 shows the average binding signal for both the 10

elongated peptides with the highest affinity derived from each base tetrapeptide sequence and for the 10 elongated peptides with the lowest affinity for each base tetrapeptide sequence. In some of the cases in which binding to the base tetrapeptide sequence alone was weak, the highest affinity extended tetrapeptide sequences increased binding signal by as much as ~18-fold. Weak-binding base tetrapeptide sequences containing negatively charged residues (EYEY and YRRE) were not improved as much by extension as neutral or positive sequences. Base tetrapeptides that were strong binders alone also showed less apparent improvement. However, this appears to be an issue of dynamic range of the experiment. For peptides based on these high affinity base tetrapeptide sequences, there was relatively little change in binding signal when 100 nM β-galactosidase was used instead of 10 nM (i.e., binding was nearly saturated). Technically, it was difficult to find a concentration range in which the top end of the binding signal was not close to saturation, but the bottom end was still measurable in the same experiment. Trying to compare values between arrays that used different concentrations of protein added enough to the noise in the data so that it was not reliable. Note that for the tetrapeptide sequence EYEY, the binding is always essentially at the level of the background signal. Thus while the measured improvement due to extension is small, it is not clear whether this is because of lack of improvement or because the binding is so weak to begin with that the improvement is not sufficient to raise the signal above the background.

Table 3.1.Binding information for base tetrapeptides and extended sequences

Base Tetrapeptide	Net charge for base tetrapeptide	Base tetrapeptide binding signal to ß-galactosidase	Average binding of the elongated peptides with the same base tetrapeptide	Average binding of the top 10 elongated peptides with the same base tetrapeptide	Average binding of the bottom 10 elongated peptides with the same base tetrapeptide
EYEY	-2	0.40±0.01*	0.4	0.64±0.02	0.32±0.01
RLLY	1	0.55±0.03	1.5±0.1	8.3±0.4	0.34±0.02
RYLY	1	0.58±0.03	1.9±0.1	10.5±0.3	0.34±0.01
YRRE	1	0.61±0.04	1.01±0.04	4.2±0.1	0.34±0.02
RQYY	1	0.65±0.04	2.1±0.1	9.2±0.1	0.36±0.03
SYRS	1	5.8±0.3	6.4±0.2	12.9±0.2	0.94±0.36
YSRS	1	6.1±0.5	6.7±0.1	12.1±0.1	1.6±0.5
YSRR	2	7.9±0.3	7.2±0.1	12.5±0.2	1.4±0.5

*Errors are represented in text as ± standard error of the mean

The effect of the length of the extension was also considered. There is very little difference in the average binding signal of all base tetrapeptides extended by one, two or three amino acids (Figure S3.1). However, the 10 highest binding signals for the three amino acid extensions is ~2 fold higher than for one amino acid extensions and ~1.2 fold higher than two amino acid extensions. Thus while the average binding signal is not very sensitive to the length of extension, the absolute value of the extent of change is. This could be an effect of length per se, or it could be simply that there are more three amino acid extensions to choose from than two amino acid or one amino acid extensions (increased sequence complexity).

As described above, extension can either reduce or increase the affinity of a particular tetrapeptide sequence for β-galactosidase. One question is whether there is a discernible sequence dependence of extension on binding. One might expect that

extending a base tetrapeptide sequence with an extension sequence of one to three amino acids that by itself is a strong binder would lead to an extended peptide sequence with increased binding affinity. Conversely, adding an extension sequence that binds β-galactosidase only weakly by itself might be expected to lower the binding affinity of the extended peptide sequence. Figure 3.3 shows the result of plotting the average binding signal of the extended peptide sequences that used a particular one, two and three amino acid extension sequence vs. the inherent binding of the extension sequence alone. Note that each point in the plot is the average of the eight possible extended sequences made from a particular extension sequence and one of the eight base tetrapeptide sequences. There is a correlation between the binding signal of the extension sequence alone and the binding of the average extended peptide sequence, but it is less pronounced than that between the binding signal of the base tetrapeptide alone vs. the average signal from the extended peptide sequence in Figure 3.2. This may be due in part simply to the size of some of the extensions: when added to a tetrapeptide, a single amino acid is unlikely to have as dramatic an effect as a larger sequence, particularly on the low end of the binding scale.

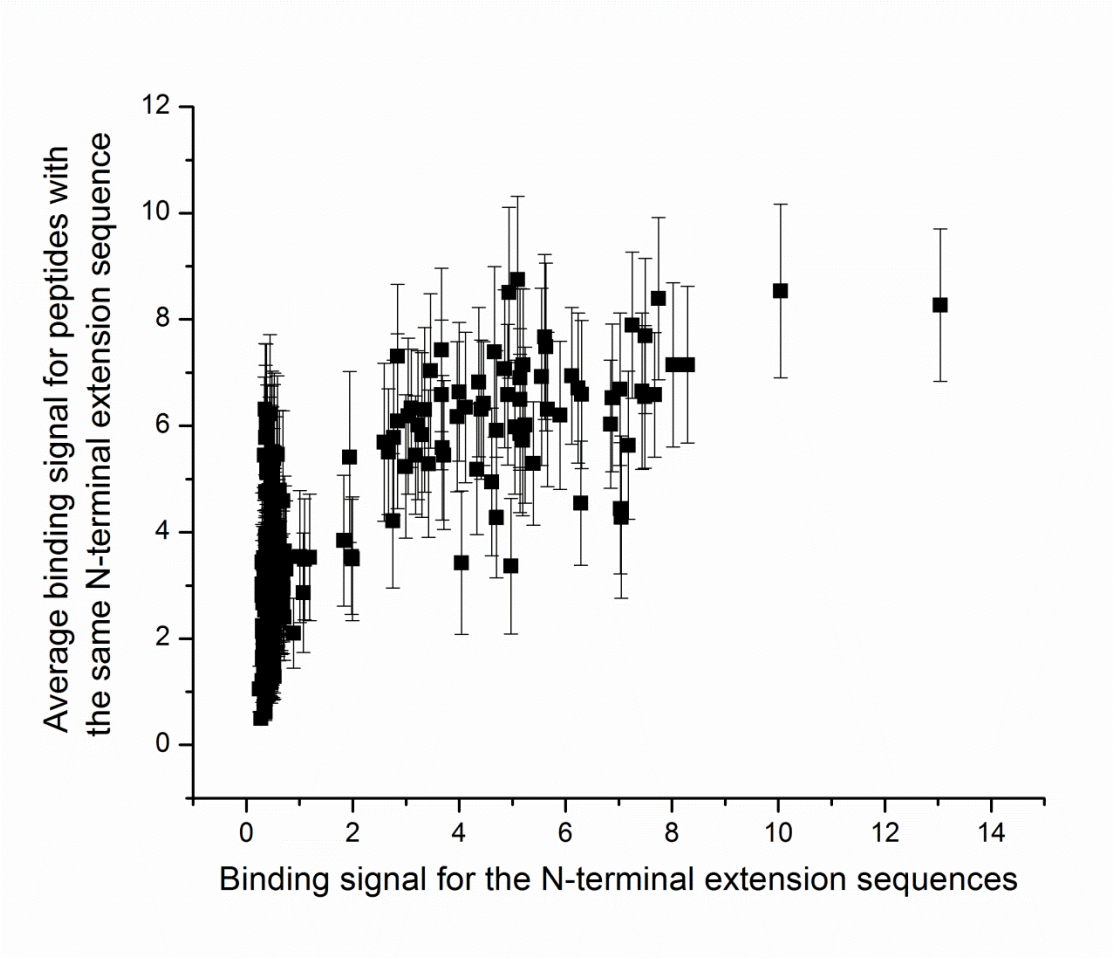

Figure 3.3. β-galactosidase average binding signal for the extended peptide sequences vs the inherent binding signal of the extension sequence alone.. The binding signals (AF555- β-galactosidase) from peptides with the same N-terminal extension sequence were averaged together. For each N-terminal extension sequence, there were 8 peptides. Binding signals shown were normalized to the global median. The error bars shown represent the error in the mean for each N-terminal extension sequence group.

To explore this issue further, extended sequences with the highest and lowest binding signals for β-galactosidase were compared to the highest and lowest signals seen

for the extension sequences by themselves. Of the extended peptide sequences with the top 10% of binding signals, about half of these included extension sequences that ranked in the top 10% of the binding signals from extension sequences themselves. Similarly, of the extended peptide sequences with the bottom 10% of binding signals, about a quarter of these included extension sequences that ranked in the bottom 10% of the binding signals from extension sequences themselves. Thus there is a statistical bias for the extension sequences with the greatest inherent affinity for β-galactosidase to be present in the extended peptide sequence with the greatest β-galactosidase affinity.

Figure 3.4 considers in more detail the amino acid composition of the sequences of the extended segments themselves. The charge characteristics of the amino acids used in the N-terminal extension sequences is the most important factor in determining binding to β-galactosidase. This is also true of the extension sequences by themselves. Figure 3.4 shows a heatmap of the amino acid composition of the extension sequence added to different base tetrapeptides that gave rise to extended peptide sequences with binding signals in the top 10%. The amino acid composition values were determined from the ratio of the amino acid's frequency of occurrence in top 10% of binding sequences divided by its frequency of occurrence in all extended sequences. A similar heatmap for the bottom 10% of the binding sequences is not shown but is essentially the reverse of Figure 3.4. As one might expect given the low pI of β-galactosidase (4.61), the positively charged amino acid, arginine, occurs in almost all strong binding sequences, while the negatively charged amino acid, glutamic acid, occurs in almost all weakly binding sequences. Perhaps of more interest, tyrosine has a somewhat elevated frequency of occurrence in the top 10% of binding sequences while glutamine and serine have a higher

likelihood of being in the bottom 10% of the binding sequences. The composition of amino acids that gave rise to a strong enhancement of binding when added to the N-terminus of the tetrapeptide "EYEY" was quite different from that of extensions from the other base tetrapeptides, however the signal is so low for this base tetrapeptide alone that calculation of enhancement factors is subject to substantial noise.

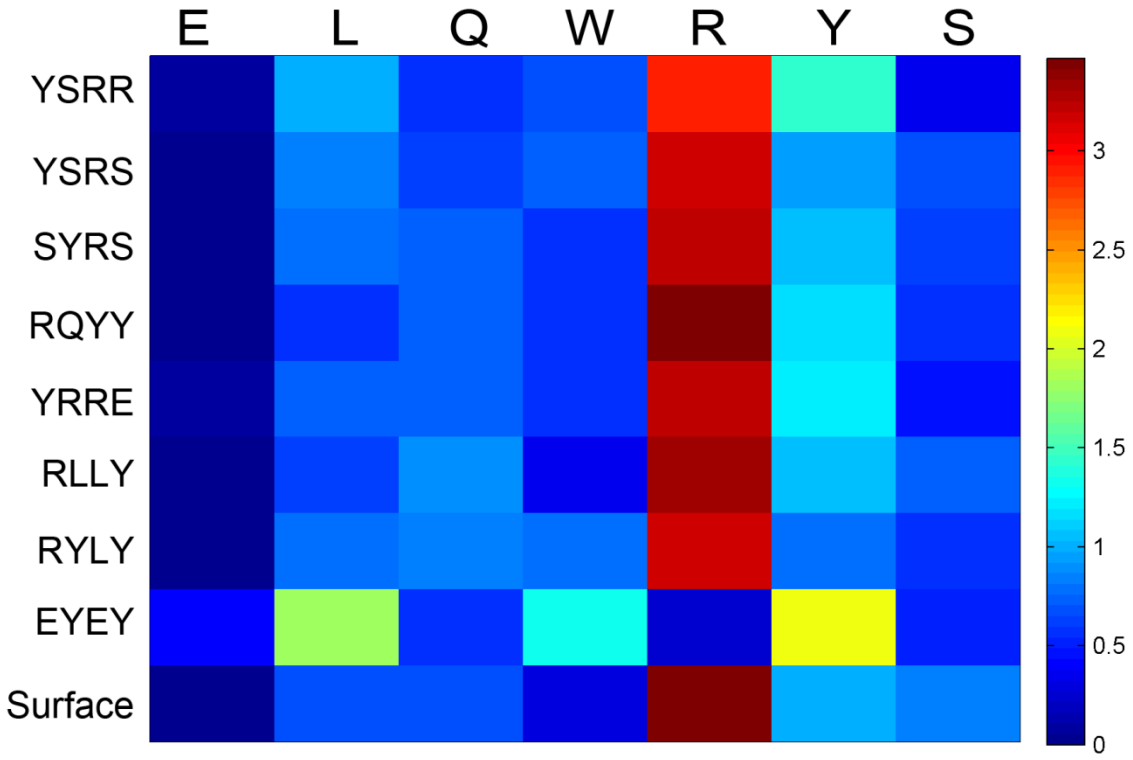

Figure 3.4. A heatmap of the relative frequency of specific amino acids in the extension sequences. The y-axis shows the different base tetrapeptides used ("Surface" means there was no tetrapeptide; amino acids were added directly to the linker on the surface). Each column represents a different amino acid used in the extension sequence. Each row represents a different base tetrapeptide. The color represents the relative frequency of occurrence of a particular amino acid in the highest binding 10% of the

extension sequences built on the base tetramers. A color value of "1.0" represents a frequency of occurrence of a particular amino acid equivalent to that expected from a randomly defined sequence (1/7 since there are seven different amino acids used).

Relative specificity, the tendency to bind one protein vs. another, was explored by comparing the binding of β-galactosidase protein to the binding of a mixture of 10 different proteins (these were the same proteins used in a previous study of protein binding to ordered peptide libraries plus α1 antitrypsin [12]). When the relative specificity is plotted against the binding signal, a strong correlation between affinity and relative specificity for these peptides is observed (data not shown), as was found previously in studies of β-galactosidase to peptide libraries[12]. Normalizing the data of Figures 3.2 and 3.3 to the intensity of the mixture of 10 proteins did not appreciably change the trends, again suggesting that binding and relative specificity track. Table 3.2 shows the relative specificity values for the base tetrapeptides and the elongated peptide sequences. As one might expect, the most specific base tetrapeptides give rise to most specific extended peptide sequences.

Table 3.2. Relative specificity for base tetrapeptides and extended sequences

Tetrapeptide	Net charge	Relative specificity to ß-galactosidase	Average relative specificity of all elongated peptides	Average relative specificity of the top 10 elongated peptides
EYEY	-2	0.56±0.02*	0.65	0.95±0.03
RYLY	1	0.67±0.03	0.90±0.03	0.7±0.1
RLLY	1	0.67±0.04	1.2±0.1	3.9±0.2
YRRE	1	0.75±0.04	1.24±0.03	3.7±0.1
RQYY	1	0.76±0.03	1.01±0.03	3.1±0.1
SYRS	1	2.0±0.2	3.8±0.1	7.3±0.2
YSRS	1	2.4±0.3	3.9±0.1	6.6±0.1
YSRR	2	2.8±0.2	4.7±0.1	9.1±0.3

*Errors are represented in text as ± standard error of the mean

Finally, the effect of amino acid order in the extension sequences on binding was considered in the case of 3 amino acid extensions. Extended peptide sequences in which a particular amino acid was in the first, second or third position were averaged to try and understand if sequence or composition was more important in the effect of the extension sequence. What was found was that the effect of a particular amino acid, on average, was largely independent of where it was placed (data not shown). Thus, composition of the extension sequence appears to be more important than the actual order of amino acids.

Concatenation Arrays

A set of 57 tetrapeptides ("component tetrapeptides") were selected as described in Materials and Methods and all possible pairs were concatenated to form 3249 different sequences ("concatenated sequences") of the form "(N') xxxx yyyy GSG (C')". The objective of this study was to explore how the affinity of each concatenated sequence relates to the known affinities of its component tetrapeptides and to what extent the order of those component tetrapeptides matter. Figure 3.5 shows the average binding signal

from the concatenated peptides that have a particular C-terminal component tetrapeptide sequence ("yyyy" above) versus the inherent binding signal of that individual C-terminal component tetrapeptide alone. The correlation between the average binding signal for the concatenated sequence and the average binding signal for its C-terminal component tetrapeptide is very strong (almost 1 to 1).

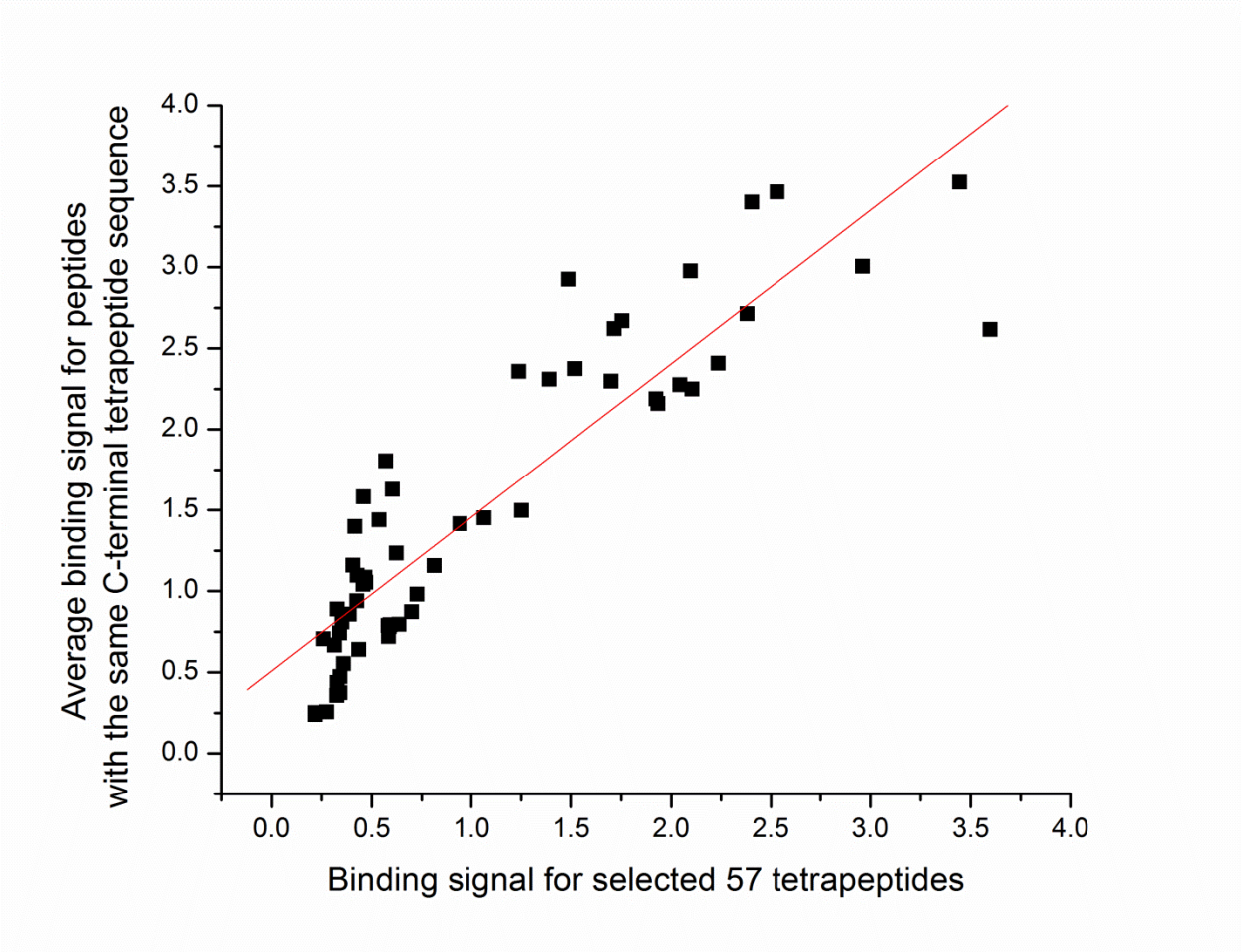

Figure 3.5. β-galactosidase average binding signal of the concatenated sequences vs. the inherent binding signal of the C-terminal component tetrapeptide sequence alone. The average of the binding signals from all 57 peptides containing a particular C-terminal component tetrapeptide sequence are shown as a function of the binding signal from the C-terminal component tetrapeptide by itself. Intensity values

shown are normalized to the global median and signals from both AF555 and AF647 labeled protein binding measurements were averaged. The correlation is very similar when either dye labeled protein measurement is considered separately, implying that binding is not very dependent on the dye (Figs S3 and S4).

Figure 3.6 shows the average binding signal from the 57 concatenated peptides with a particular N-terminal component tetrapeptide sequence ("xxxx" in the motif defined above) plotted as a function of the inherent binding strength of that N-terminal component tetrapeptide sequence alone. In contrast to Figure 3.5, the dependence is weak, suggesting that a sequence that was originally selected to bind well a particular distance from the C-terminus of the peptide (which is where the peptide is attached to the surface) does not necessarily contribute strongly to binding when displaced from the surface. In addition to the weaker correlation shown in Figure 3.6, the set of concatenated peptides that have the same component tetrapeptide sequence at N-terminus exhibit a different binding dynamic range than the set of concatenated peptides with that same component tetrapeptide sequence at the C-terminus. More specifically, the standard deviation of the binding signals of concatenation peptides that have a particular C-terminal tetrapeptide sequence is 0.63 while the standard deviation in the binding of families of sequences with the same N-terminal component tetrapeptide is 1.07; grouping sequences by their N-terminus is much more variable that grouping sequences by their C-terminus. This again implies that the position of the binding sequence relative to the surface affects the nature of its contribution to the binding signal.

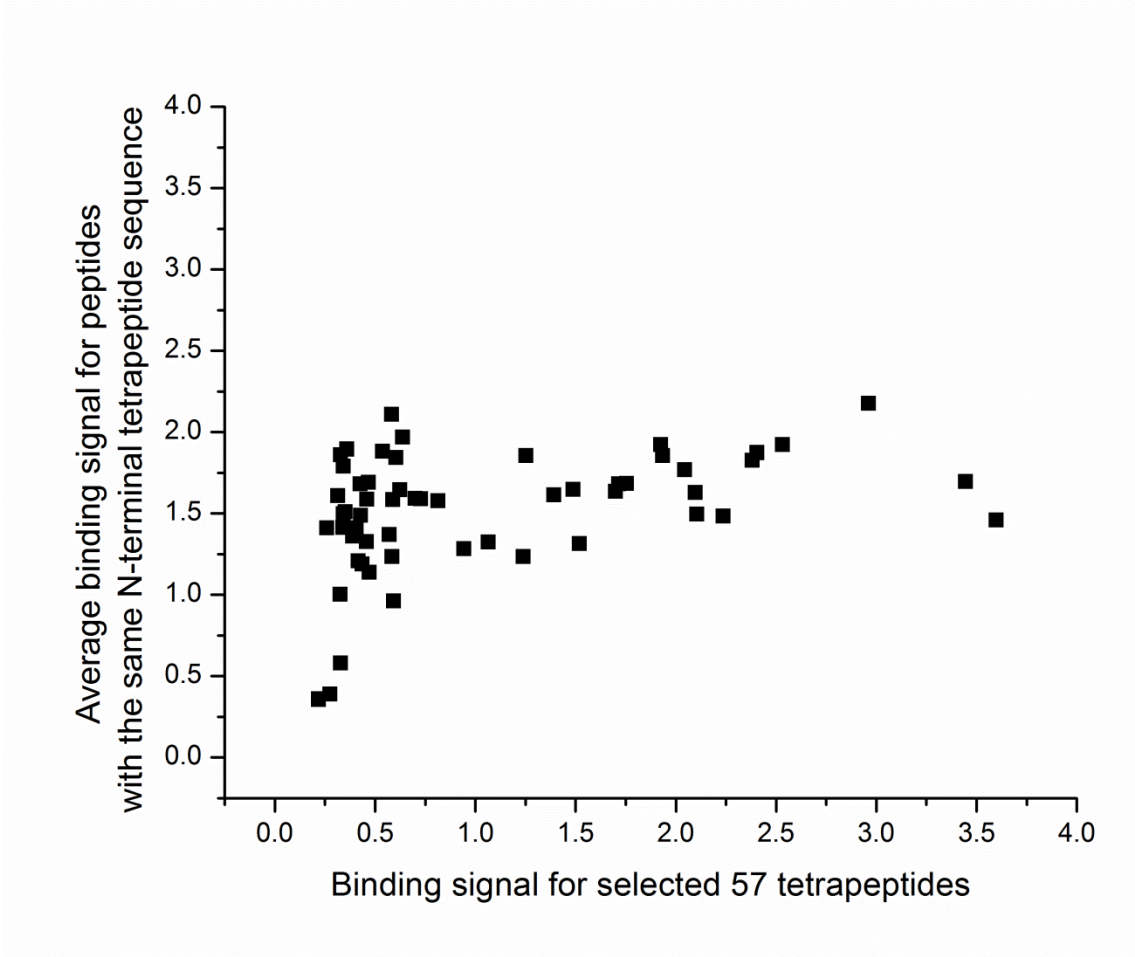

Figure 3.6. β-galactosidase average binding signal of the concatenated sequences vs. the inherent binding signal of the N-terminal component tetrapeptide sequence alone. The average binding signals from all peptides that contain a particular N-terminal component sequence were plotted as a function of the average binding signal of the N-terminal component sequence alone. Intensity values shown are normalized to the global median and the binding signals of AF555 and AF647 labeled proteins were averaged. Either dye-bound protein gives a very similar binding trend (Figs S5 and S6).

The effect of charge on peptide binding was also explored in the concatenation studies. β-galactosidase has a pI of 4.61, and thus is negatively charged at neutral pH.

Therefore, one would expect that peptides with positive charges have higher binding signal in general. Consistent with this expectation, the concatenated peptides with a negatively charged component tetrapeptide on either the N-terminus or C-terminus show relatively weak binding to β-galactosidase. This trend is a little more pronounced when the negatively charged component tetrapeptide sequence is placed near surface at C-terminal. As was seen for the extension study, concatenation of an N-terminal component tetrapeptide to a negatively charged C-terminal component tetrapeptide hardly increases binding to β-galactosidase at all, regardless of how strongly the N-terminal component tetrapeptide sequence binds by itself.

The relative specificity of β-galactosidase binding was also studied for the concatenation library. Similar to the extension studies, when the relative specificity is plotted against the binding signal, a strong correlation between affinity and relative specificity is observed (data not shown). There is no significant change in correlations when the data of Figures 3.5 and 3.6 are normalized to the intensity of the mixture of 10 proteins (data not shown).

DISCUSSION

Traditionally, the focus of molecular recognition studies in biology has been on so-called specific interactions. In general, nonspecific interactions are largely ignored or considered uninteresting. However, by definition, any macromolecule in the cell undergoes many more nonspecific than specific interactions with molecules and particularly molecular surfaces around it. Indeed, as our understanding of the crowded environment of the cell increases, it has become more and more apparent how important controlling such interactions in both time and space are. This is thought to have been a

significant constraint in the evolution of protein structure [25, 26]. In this regard, "specific" and "nonspecific" are really misnomers. A more precise nomenclature is "structured" and "unstructured" interactions; the difference between them being whether a certain metastable structure is formed (a narrow potential well) or whether the affinity is driven by longer range forces or by a summation of many interactions that allow for multiple relative positions of the binding pair (a broad potential well).

Classical approaches to understanding what forces and types of chemical interactions are involved in structured interactions often do not work well for unstructured interactions due to the heterogeneity of the system. Instead, statistical approaches can be used, and the systematic analysis of binding to diverse arrays of binding partners is one such approach. The studies described above were built on earlier work in which a comprehensive analysis of tetrapeptides binding to a common protein, β-galactosidase, was performed, resulting in a catalogue of peptide sequences and relative affinities[12]. This made it possible to ask several simple, fundamental questions about the nature of unstructured interactions:

- Is the binding of the component motifs additive?
- Does the order or context of the binding motifs matter?
- What kinds of intermolecular forces dominate binding to β-galactosidase?
- Does additive binding result in additive specificity?

β-galactosidase is well suited for this study as it is a protein with a large, diverse surface area that does not have known natural peptide binding partners. This increases the likelihood of finding unstructured binding interactions with the peptide covered surfaces.

Additivity. In the first set of experiments described above, extending a tetrapeptide motif by one, two or three amino acids resulted in both increases and decreases in overall binding of β-galactosidase, as shown in Figure 3.1. In general, the ability of the extension sequences to increase or decrease the binding of the base tetrapeptide correlated weakly with the binding of the extension sequence alone (Figure 3.3). The shallow correlation is particularly evident if one excludes the very weakest binding peptides (<1 on the scale given in Figure 3.3). The weak additivity is even more apparent when two component tetrapeptide sequences that have been tested independently for relative affinity are concatenated together. As shown in Figure 3.6, simply adding a high affinity sequence to the N-terminus of another sequence of known relative binding affinity does not increase binding affinity substantially on average. Thus, while adding sequences to the N-terminus of a known binder can indeed either enhance or reduce binding by a large factor (see Tables 3.1 &3.2 and associated text), that change does not correlate well with the affinity of that N-terminal addition when it is attached directly via the GSG linker to the surface.

Order or Context. Figures 3.2, 3.3, 3.5 and 3.6 address the question of position or order of binding sequences. In Figures 3.2 and 3.5, there is a strong, and nearly one to one, correlation between the relative affinity of the C-terminal sequence and the average relative affinity of the extended or concatenated sequences. In Figures 3.3 and 3.6, the correlation with the N-terminal sequence is considerably weaker. In other words, the binding contribution of the C-terminal tetrapeptide remains more or less constant, regardless of what is added to it, and the relative affinity of the sequence added is not

very predictive of overall affinity. Thus the order or context of the binding motif is critical in determining its contribution to the binding in these surface attached peptides.

Forces. Forces that drive unstructured interactions can be any of a number of non-covalent forces including charge-charge, hydrophobic or hydrogen bonding interactions. Here, charge-charge interactions appear to play the dominant role in determining the binding activity for β-galactosidase (a very negatively charged protein). This is reflected in Figure 3.4 which shows that β-galactosidase prefers Arg, a positively charged residue, to any other amino acid for extending a tetrapeptide motif by one, two or three amino acids. The importance of charge-charge interactions has been confirmed in previous studies in which β-galactosidase binds strongly and selectively to some of the most positively charged peptides[12]. Because of its large size, there is more opportunity for long-range interactions, like columbic interactions to take place. In addition, because of its low pI, it is strongly negatively charged at neutral pH; there are approximately 97 aspartic acid residues and 135 glutamic acid residues exposed to the solvent. This makes it very easy to form charge-charge interactions with positively charged peptides. The fact that extension of peptides with tyrosine-containing sequences appear to enhance binding could suggest a role for hydrogen bonding, but additional comparison (e.g. to phenylalanine that was not included in this study) would be necessary.

Specificity. In both experiments performed here, the relative specificity values for β-galactosidase are strongly correlated to its binding affinity values. In other words, the ratio between β-galactosidase binding and the binding of a diverse set of ten other proteins increases and decreases in essentially the same way as does the affinity. Thus, extension sequences that result in more affinity usually result in more specificity. Clearly,

unstructured interactions between proteins and surface bound peptides can be quiet selective in some cases as shown by β-galactosidase, making it possible to create complex patterns of relatively short peptides on surfaces that are capable of quite selective interactions with different proteins.

ACKNOWLEDGEMENTS

The authors would like to thank Drs. Zhan-gong Zhao, Matthew Greving, Phillip Stafford and Stephen Johnston for helpful discussions and input. This material is based in part upon work supported by the National Science Foundation under Grant No. (MCB-1243082).

SUPPORTING INFORMATION

The effect of the length of the N terminal extension

The effect of the length of the N terminal extension was studied by plotting the average β-galactosidase binding signal for one, two and three amino acid extended peptide sequences (each point is an average over all extended peptide sequences with a particular base tetrapeptide sequence) versus the binding signal for the base tetrapeptide alone (Figure S3.1). The average binding signals are not substantially different as a function of extension length, though as described in the main text, the extended peptide sequences that show the highest binding signals have three amino acid extensions. In other words, the effect of the extension at any length is more or less equal in either direction (increasing or decreasing the binding signal), but the absolute value of the effect is greater with longer extensions.

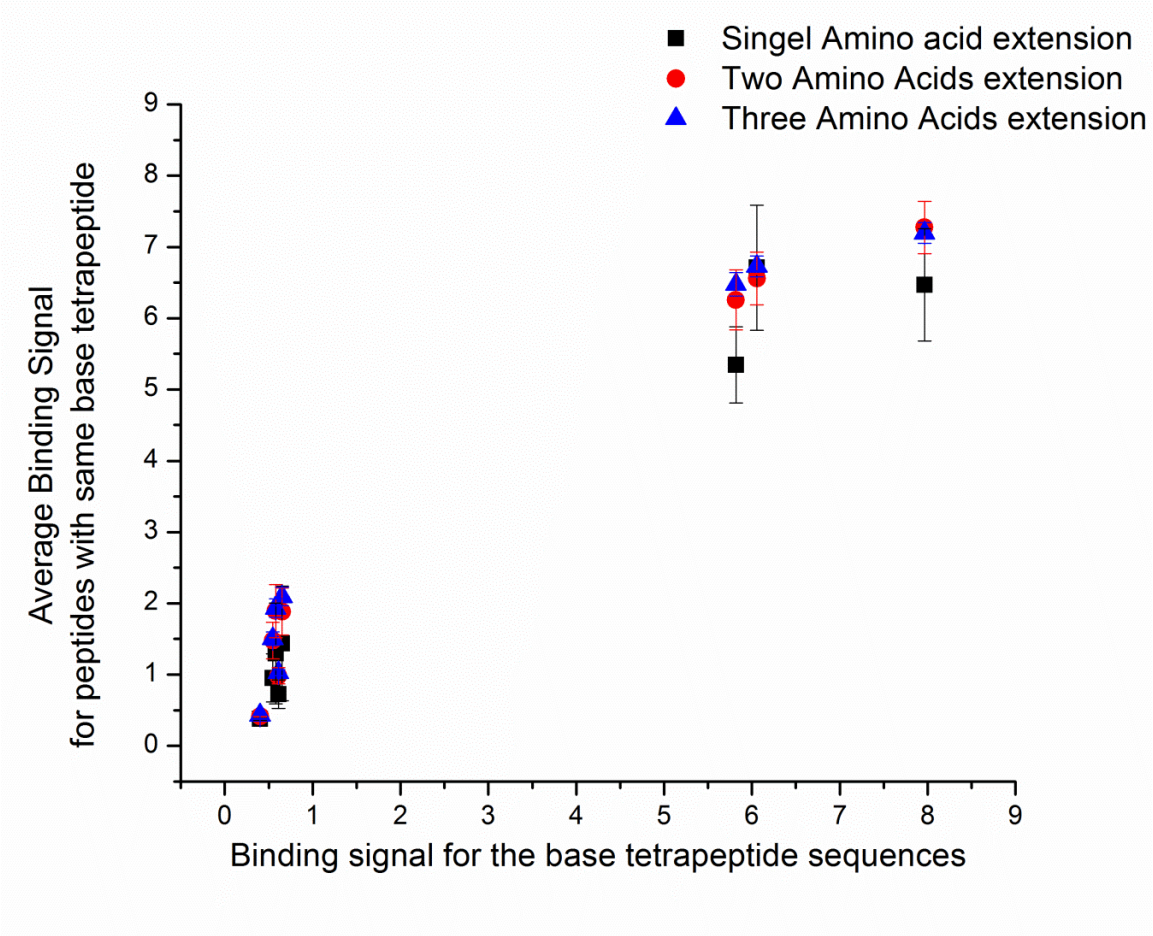

Figure S3.1 β-galactosidase average binding signal for the one, two and three amino acid extended peptide sequences vs. the binding signal of the base tetrapeptide sequence upon which the extended sequences were built. Details as in figure 3.2.

The effect of dye-driven binding

To moniter the effect of dye-bias on protein binding, a dye-swaping expereiment was performed using AF555 labeled β-galactosidase on one array and AF647 labeled β-galactosidase on another identical array under the same binding conditions at the same time. Figure S3.2 shows the results of plotting the binding signal from each feature in the array using AF555 labeled protein vs. the same feature in the array using AF647 labeled

protein. In general, the distribution of intensities is what one might expect for a protein dominated binding interactions; almost all of the peptides that give high binding with one dye label also give high binding with the other. To specifically address the question of whether dye binding is affecting the results of this study, Figures S3.3, S3.4, S3.5 and S3.6 show the average binding signal from the concatenation peptides that have a particular C (or N)-terminal base tetrapeptide sequence versus the inherent binding signal of that individual C (or N)-terminal tetrapeptide alone using either binding data from AF555 (Figures S3.3 and S3.4) or AF647 (Figures S3.5 and S3.6) labeled β-galactosidase. The binding trends and conclusions one would draw are very simialr to what was observed with the average binding data of AF555 and AF647 labeled β-galactosidase (Figure 3.5 and Figure 3.6). This indicates that the dye-bias effect is not the driving force in the binding trends observed here.

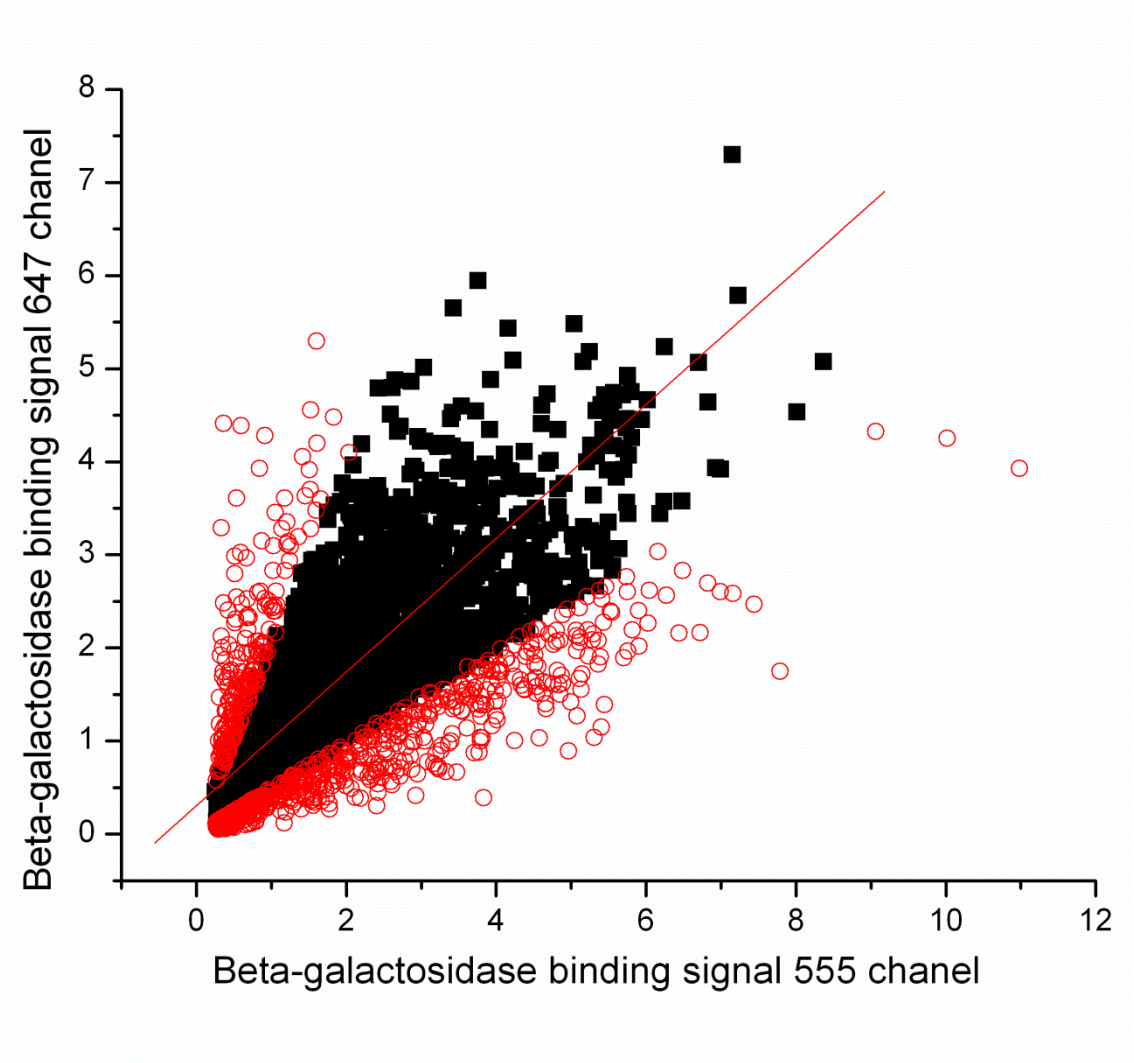

Figure S3.2 AF555 labeled beta-galactosidase binding intensity versus AF647 labeled beta-galactosidase binding intensity. The fluorescent signals from the bound labeled proteins were measured and normalized to the global median on each array. A linear regression was performed between two groups and the binding signal ratio between two labeled proteins was calculated. Points that have a ratio smaller than 0.5 or larger than 2.0 are represented by red circles. While there is likely some dye-driven binding, the vast majority of the signal appears to correlate well between the two arrays, given a Pearson correlation coefficient of 0.74.

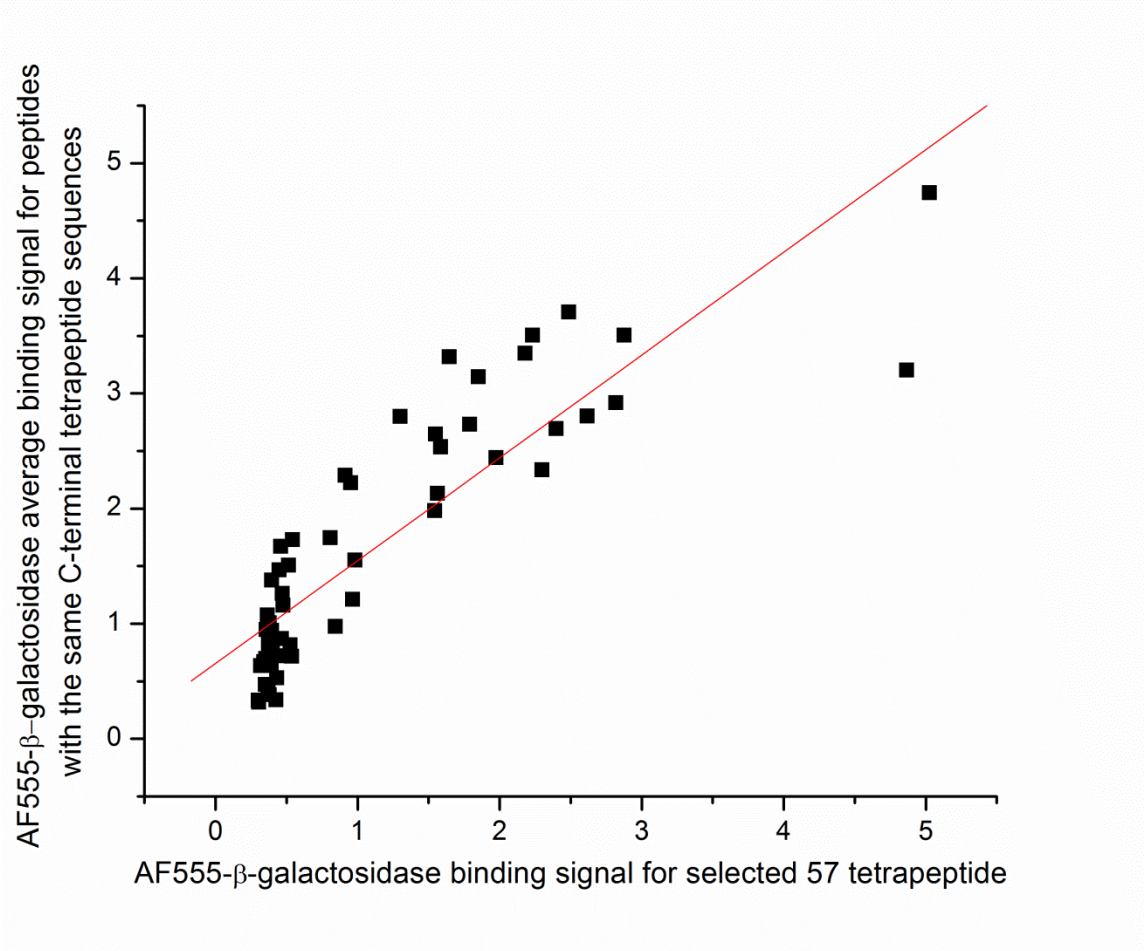

Figure S3.3 The average binding signal from AF555 labeled β-galactosidase incubated with concatenated peptides sharing the same C-terminal base tetrapeptide sequence versus the inherent binding signal of the C-terminal base tetrapeptide sequence alone. This plot is identical to Figure 3.5 except that only AF555 labeled β-galactosidase was used.

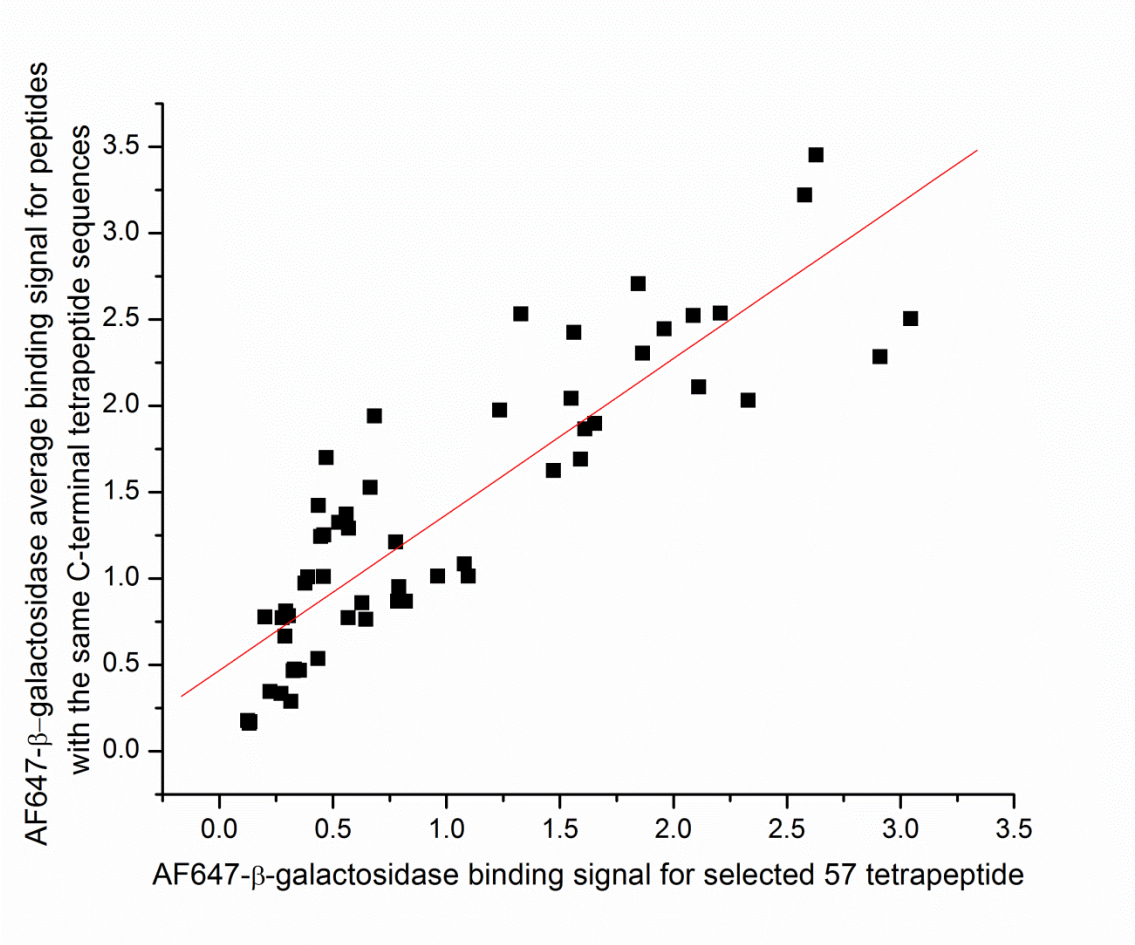

Figure S3.4 The average binding signal from AF647 labeled β-galactosidase incubated with concatenated peptides sharing the same C-terminal base tetrapeptide sequence versus the inherent binding signal of the C-terminal base tetrapeptide sequence alone. This plot is identical to Figure 3.5 except that only AF647 labeled β-galactosidase was used.

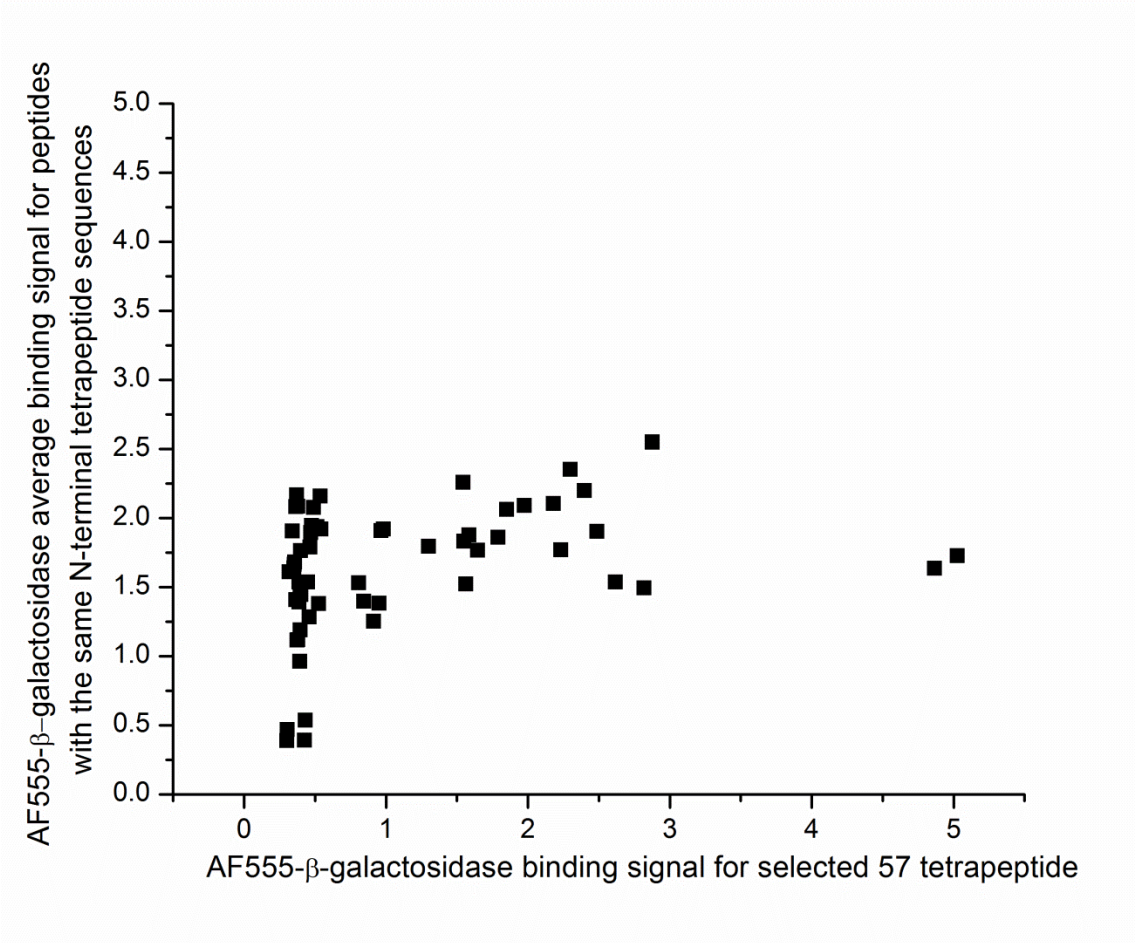

Figure S3.5 The average binding signal from AF555 labeled β-galactosidase incubated with concatenated peptides sharing the same N-terminal base tetrapeptide sequence versus the inherent binding signal from the N-terminal base tetrapeptide sequence alone. This plot is identical to Figure 3.6 except that only AF555 labeled β-galactosidase was used.

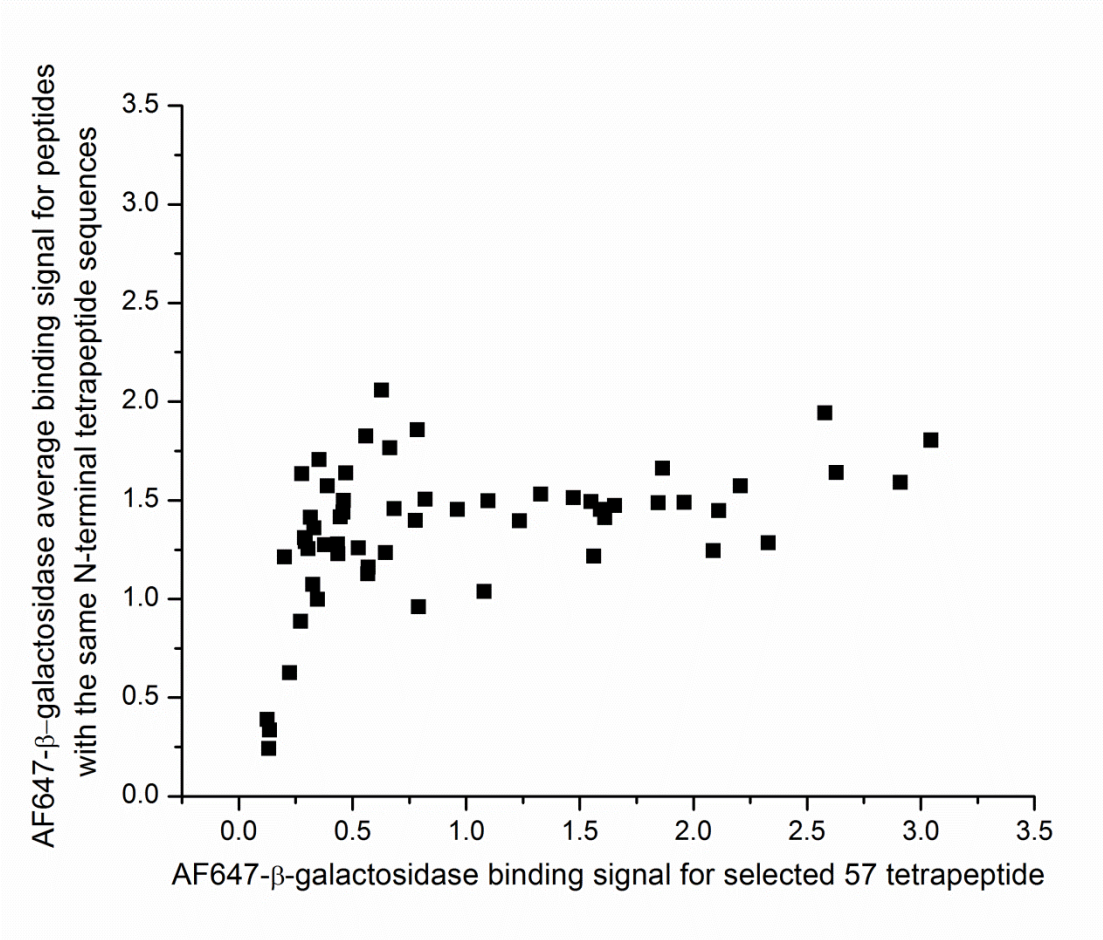

Figure S3.6 The average binding signal from AF647 labeled β-galactosidase incubated with concatenated peptides sharing the same N-terminal base tetrapeptide sequence versus the inherent binding signal from the N-terminal base tetrapeptide sequence alone. This plot is identical to Figure 3.6 except that only AF647 labeled β-galactosidase was used.

Graphical Abstract

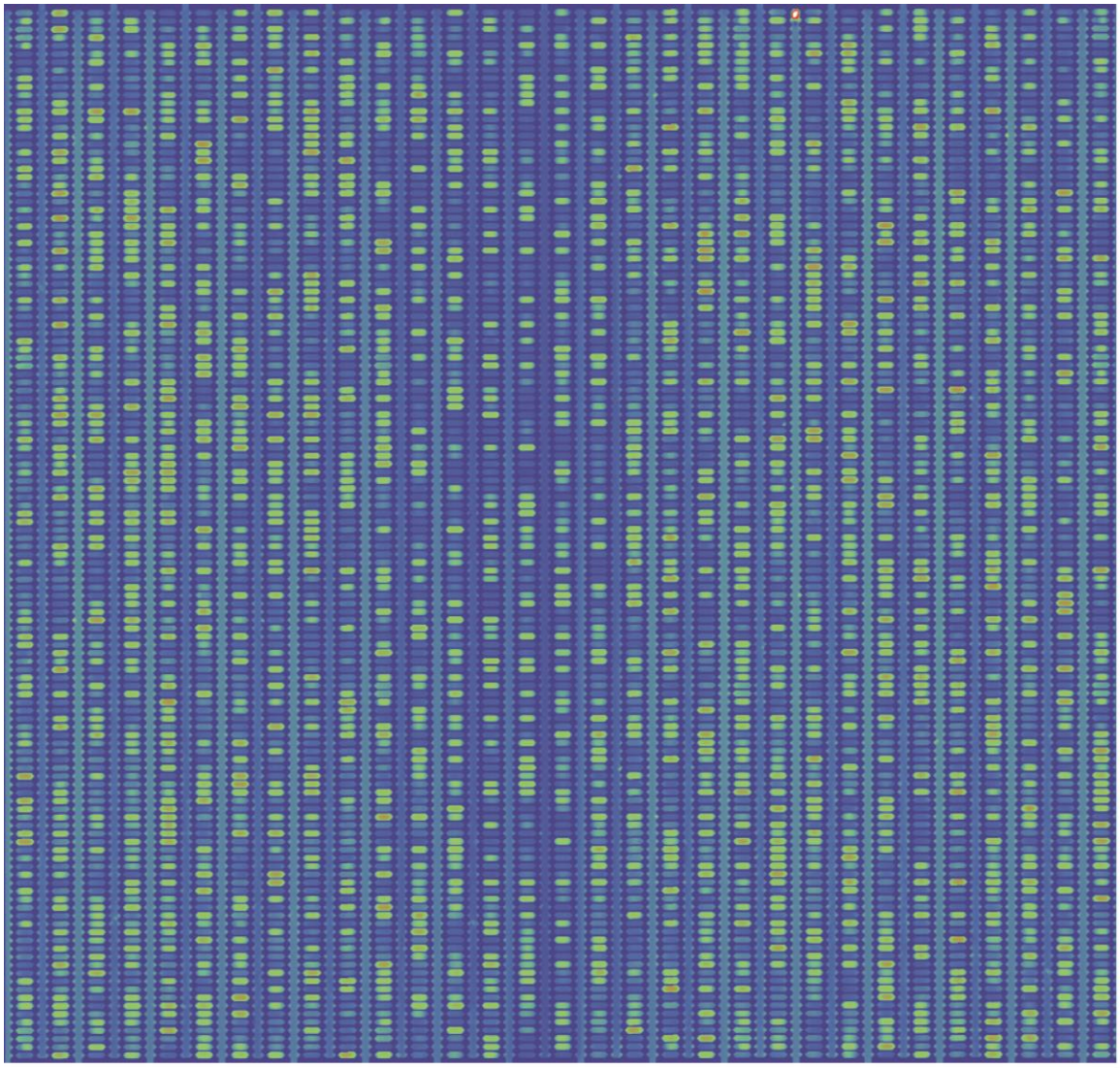

Figure S3.7. Labeled β-galactosidase binding to an “extension array”. This is an example of the array images received from LC Sciences.

REFERENCES

[1] Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets - The roles of intrinsic disorder in protein interaction networks. Febs Journal 2005;272:5129-48.

[2] Ratner BD, Bryant SJ. Biomaterials: Where we have been and where we are going. Annu Rev Biomed Eng 2004;6:41-75.

[3] Magee J, Warwicker J. Simulation of non-specific protein-mRNA interactions. Nucleic Acids Res 2005;33:6694-9.

[4] Halford SE, Marko JF. How do site-specific DNA-binding proteins find their targets? Nucleic Acids Res 2004;32:3040-52.

[5] Schultz SC, Shields GC, Steitz TA. Crystal-Structure of a Cap-DNA Complex - the DNA Is Bent by 90-Degrees. Science 1991;253:1001-7.

[6] Ganguly A, Rajdev P, Williams SM, Chatterji D. Nonspecific Interaction between DNA and Protein allows for Cooperativity: A Case Study with Mycobacterium DNA Binding Protein. J Phys Chem B 2012;116:621-32.

[7] White AD, Nowinski AK, Huang WJ, Keefe AJ, Sun F, Jiang SY. Decoding nonspecific interactions from nature. Chem Sci 2012;3:3488-94.

[8] White AD, Huang WJ, Jiang SY. Role of Nonspecific Interactions in Molecular Chaperones through Model-Based Bioinformatics. Biophys J 2012;103:2484-91.

[9] Johnson ME, Hummer G. Nonspecific binding limits the number of proteins in a cell and shapes their interaction networks. P Natl Acad Sci USA 2011;108:603-8.

[10] Phillips SE, Ile KE, Boukhelifa M, Huijbregts RPH, Bankaitis VA. Specific and nonspecific membrane-binding determinants cooperate in targeting phosphatidylinositol transfer protein beta-isoform to the mammalian trans-golgi network. Mol Biol Cell 2006;17:2498-512.

[11] Fu JL, Cai K, Johnston SA, Woodbury NW. Exploring Peptide Space for Enzyme Modulators. J Am Chem Soc 2010;132:6419-24.

[12] Wang W, Woodbury NW. Selective protein-peptide interactions at surfaces. Acta Biomater 2014;10:761-8.

[13] http://www.lcsciences.com/applications/proteomics/.

[14] Gao XL, Zhou XC, Gulari E. Light directed massively parallel on-chip synthesis of peptide arrays with t-Boc chemistry. Proteomics 2003;3:2135-41.

[15] Pellois JP, Zhou X, Srivannavit O, Zhou T, Gulari E, Gao X. Individually addressable parallel peptide synthesis on microchips. Nat Biotechnol 2002;20:922-6.

[16] Pellois JP, Wang W, Gao X. Peptide synthesis based on t-Boc chemistry and solution photogenerated acids. Journal of combinatorial chemistry 2000;2:355-60.

[17] Stewart JM, Young,J.D. Solid Phase Peptide Synthesis. 2nd edition ed: Pierce Chemical Co; 1984.

[18] Kyte J, Doolittle RF. A Simple Method for Displaying the Hydropathic Character of a Protein. J Mol Biol 1982;157:105-32.

[19] Greving MP. Creating high-affinity ligands on surfaces and in solution: Arizona State University; 2009.

[20] Chang YP, Mahadeva R, Chang WSW, Lin SC, Chu YH. Small-molecule peptides inhibit Z alpha(1)-antitrypsin polymerization. J Cell Mol Med 2009;13:2304-16.

[21] Chang YP, Mahadeva R, Chang WSW, Shukla A, Dafforn TR, Chu YH. Identification of a 4-mer peptide inhibitor that effectively blocks the polymerization of pathogenic Z alpha(1)-antitrypsin. Am J Resp Cell Mol 2006;35:540-8.

[22] Crystal RG. The alpha 1-antitrypsin gene and its deficiency states. Trends in genetics : TIG 1989;5:411-7.

[23] Elliott PR, Pei XY, Dafforn TR, Lomas DA. Topography of a 2.0 angstrom structure of alpha(1)-antitrypsin reveals targets for rational drug design to prevent conformational disease. Protein Sci 2000;9:1274-81.

[24] Diehnelt CW, Shah M, Gupta N, Belcher PE, Greving MP, Stafford P, et al. Discovery of High-Affinity Protein Binding Ligands - Backwards. Plos One 2010;5.

[25] Levy ED, De S, Teichmann SA. Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. Proc Natl Acad Sci U S A 2012;109:20461-6.

[26] Yang JR, Liao BY, Zhuang SM, Zhang J. Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. Proc Natl Acad Sci U S A 2012;109:E831-40.

CHAPTER 4: EXPLORING THE NATURE OF PEPTIDE-ANTIBODY INTERACTIONS USING DIVERSE-SEQUENCE PEPTIDE ARRAYS

CONTRIBUTION

The following chapter describes a study that exploring the general interactions between antibodies and surfaces covered with peptides. The binding of monoclonal antibodies and serum to the HealthTell array surface covered with particular peptides is systematically studied and the binding information is used to understand the chemistry of binding. The array synthesis and binding assays were performed by HealthTell and by the CIM peptide array core. For each peptide feature, the raw binding signal from the binding assay is the median intensity of each spot on the array. Wei Wang took the raw binding data and performed the analysis of the binding data as described below.

INTRODUCTION

Antibody-antigen reactions are widely used both clinically and in the laboratory to detect the level of biomarkers or pathogens in human serum associated with a variety of diseases. In a number of cases the antibody raised in response to the disease itself serves as a key biomarker [1, 2]. The Center for Innovation in Medicine in the Biodesign Institute at Arizona State University has developed an approach for profiling the repertoire of circulating antibodies in the blood that utilizes high density peptide arrays. Specifically, human serum containing large amount of antibodies is applied to the surface of an array of thousands of diverse-sequence peptides and the binding of blood-borne antibodies to these peptides can be measured in terms of the fluorescence intensity of a labeled secondary antibody. The resulting so-called immunosignature, the pattern of binding of antibodies to the diverse-sequence peptide features in the array, can both

detect and identify a wide variety of specific diseases[2]. There are several published studies using immunosignatures to characterize vaccines[3] and it has been successfully applied to detection of such diseases as Alzheimer's disease[4], various brain cancers[5], valley fever[6] and breast cancer[7]. The approach has also been used to simultaneously detect and discriminate multiple cancers as well as multiple infectious diseases[8, 9]. Results from these studies strongly suggest that the unique immunosignatures derived from the binding between antibodies and diverse-sequence peptides can reflect disease-driven changes in the antibody profile and the approach is currently being commercialized as a diagnostic tool with HealthTell (www.healthtell.com).

Although linear antibodies (antibodies that bind a contiguous sequence of amino acids) are known bind to specific epitopes, most antibodies also have substantial cross-reactivity with non-cognate targets [10]. Moreover, using the immunosignature peptide arrays mentioned above, Halperin et al. have found that monoclonal antibodies can bind non-cognate peptide sequences with high affinity[11]. This suggests that the interactions between antibodies and these surface-bound peptides are not entirely driven by interactions with one specific sequence. In fact, when a paratope binds to its corresponding epitope, the initial interactions are thought to be driven by general long-range forces, such as ionic and hydrophobic interactions. These attractive forces locally overcome the hydration energies and allow the epitope and the paratope to approach each other more closely. Once the surfaces of the paratope and epitope are close enough, a strong interfacial bond resulting from Van der Waals forces will arise and the strength of this bond is usually depends on how closely the two surfaces fit each other and the total contact area involved [12, 13]. While the specific interfacial interaction between a

paratope and a known epitope can be characterized in detail with crystallography, exploring the range and chemical nature of possible mimotope sequences is more difficult.

Here, the interactions between antibodies and dense arrays of peptide features will be studied using peptide arrays that have been intentionally designed for measuring immunosignatures. The peptide arrays were manufactured using a photolithography-based, in situ synthesis approach and each array consists of ~330,000 peptide sequences with an average length of ~12 amino acids (plus a three amino acid linker, GSG) and a length range from 3 to 17 amino acids[8]. Sixteen of the twenty natural amino acids were used in the synthesis (cysteine, methionine, isoleucine and threonine were excluded). The ~330,000 feature arrays each cover an area of ~0.5 cm^2 and each feature in an array is 8 microns in diameter and contains a uniquely synthesized peptide sequence. The peptide sequences were generated using a pseudo-random algorithm designed to minimize the number of synthesis steps but sample combinatorial sequence space fairly evenly[8].

As described above, these arrays are being developed for health monitoring by profiling the pattern of molecular recognition of circulating antibodies in the blood. In the context of that application, there are a number of fundamental questions about general interactions between antibodies and peptide affixed to surfaces that need to be answered. The antibody samples used here include both single, commercially available monoclonal antibodies as well as serum samples from either healthy individuals or individuals known to have a specific infectious disease (Malaria). In particular, two well characterized monoclonal antibodies (p53Ab1, one of the p53 epitopes and DM1A, an antibody against

tubulin), four normal serum samples and four disease serum samples were tested on immunosignature peptide arrays from two different manufacturing lots (the arrays are made on the surface of silicon wafers and the two sets of arrays used came from two different wafers made at different times). While the monoclonal antibody samples each contain a single type of immunoglobulin, the serum samples are heterogeneous mixtures of many biomolecules including a large diversity of different immunoglobulins. Antibody binding to different peptide features on the surface is visualized in each case through the use of a labeled secondary antibody to the constant region of the primary antibody (for the studies described below, a labeled secondary antibody to IgG was used). The dynamic range of this binding is very high for both monoclonal antibodies and serum. The highest binding signals are roughly 150-fold higher than the lowest signals. Because the lengths of peptides used in the array are relatively short, the peptides are likely to be largely unstructured so that the majority interactions between antibodies and these surface bound peptides are not dependent on specific tertiary structures, a concept supported by past work comparing the binding of many different monoclonal antibodies to arrays of printed peptides[11]. Indeed, the binding of monoclonal antibodies to these dense arrays is quite diverse, with some binding observed to cognate and near-cognate sequences (in the case of monoclonal antibodies that recognize linear epitopes) as well as binding to mimitopes that bear no resemblance to the original cognate sequence. In order to explore the nature of the binding of antibodies to peptide covered surfaces in more detail, the binding of antibody samples was monitored as a function of charge, hydrophobicity, length, specific amino acid contribution and position within sequence of the peptides in the array.

METHODS AND MATERIALS

HealthTell Peptide Microarrays

The immunosignaturing peptide microarray used in this work has been described previously[8]. ~330,000 peptides of near-random sequence were synthesized by light-directed photolithography. The common linker sequence, GSG, exists on the C-terminus of every peptide. 152 control peptides containing the epitope"NH3-RHSVV-COOH" are present at random locations throughout the array. This sequence is the complete epitope for p53 antibody Ab1, and is a positive control. Other partial epitopes for Ab1, Ab8 and DM1A are embedded into other control peptides.

Binding Assays

Microarrays are pre-incubated with blocking buffer (BB = 10nM Phosphate Buffered Saline, pH 7.3 and 05% BSA, 0.5% Tween) for 1 hour prior to addition of a 1:20,000 dilution of serum into sample buffer (SB = BB less 0.5% Tween) for one hour at 25ºC. The primary antibody is washed off with BB and then the peptide-bound antibodies are detected by addition of 5nM of AlexaFluor 555-conjugated anti-human IgG1 secondary (Rockland Antibodies, Gilbertsville, PA) for 1 hour in SB at 25ºC, then washed 3x in SB, then 5x in 18MOhm water followed by centrifugation at 1800g for 5 seconds to dry. Arrays are scanned in an Innopsys Innoscan 900 scanner at 647nm and 555nm using high laser power and 70% PMT at 1µm resolution. TIFF images are aligned with the corresponding gal file that defines which measured intensity is associated with which peptide feature[6].

Samples

Serum samples were received at ASU through IRB#0912004625, "Profiling Biological Sera for Unique Antibody Signatures", renewed March, 2013 by the Western Institutional Review Board (Olympia, WA). Monoclonals were obtained from the manufacturer. All disease states were assessed by SeraCare using the FDA approved ELISA-based diagnostic. Monoclonals were quality checked by the manufacturer, and at ASU using dot-blot against peptides containing the correct epitope sequence.

Data Analysis

Binding signals were normalized to the mean intensity for each respective array. All subsequent analysis utilized the mean-normalized data. Locally written scripts in Matlab were used to sort the peptide sequences and to perform most of the statistical studies.

RESULTS AND DISCUSSION

Peptide binding intensity dependence on position or neighboring features

The nature of interactions between immunoglobulins and surface bound peptides was explored using high density peptide arrays intentionally designed for immunnosignature purposes. While the peptide arrays used for this analysis have a high feature density, the size of each feature (8 microns diameter) is still much larger than the size of immunoglobulin molecules. Thus one would expect that binding at each peptide feature would be independent of binding at neighboring features, unless there were issues associated uneven binding caused by the assay itself. To verify binding independence of each peptide feature in the array, two different methods were used. The first method focused on the intensity of the binding signal to 'RHSVV' which is the epitope sequence

for the monoclonal antibody p53Ab1. There were 152 'RHSVV' peptide features synthesized on the array surface at different locations. The binding signals for both the P53 Ab1 monoclonal antibody and for binding of serum IgG at these peptide features was constant (within the noise of the measurement), regardless of the location of the peptide feature (data not shown). Furthermore, the binding signals for the RHSVV sequences are not correlated to the binding intensities observed for the neighboring peptides or to their physical properties (length, charge, hydrophobicity, etc.). The second method of testing binding equivalence across the array was more general and was applied to all peptides in the array. The binding intensity for each peptide sequence was plotted against the average binding intensity of the 6 nearest neighbor peptides surrounding it (data not shown). This plot was then compared to the binding intensity for the same sequences versus the average binding intensity of 6 randomly selected peptides in the array. There is essentially no difference between two plots and the correlation between the binding for any particular peptide and the average binding of its neighboring peptides is very weak.

Effect of Total Peptide Charge on IgG Binding

Charge-charge interactions are one of the most important driving forces for general binding. Figure 4.1 shows the average binding signal for each sample tested as a function of net peptide charge at neutral pH. Typical monoclonal antibodies are slightly negatively charged at neutral pH. Thus, it is not surprising that both monoclonal antibodies (p53Ab1 and DM1A) show higher binding affinity to surface features with more positive charges. The effect is more pronounced for monoclonal antibody p53Ab1, where the average binding intensity for positively charged peptides is roughly 3.2-fold higher than that of negatively charged peptides. Monoclonal antibody DM1A shows a

much less pronounced increase at 1.2-fold. In the case of monoclonal antibodies, the nature of the cognate sequence and the tendency of any particular monoclonal antibody to bind to off-target sequences will also play a major role in dictating the charge dependence of binding. Interestingly, the cognate sequence of p53Ab1 (RHSVV) is itself net positive, consistent of what one might expect from the charge dependence of binding, while that of DM1A (AALEKDY) is net negative. It is interesting that DM1A on average binds more strongly to positively charged peptides. However, the net overall negative charge of the IgG molecule could be an issue. In addition, past work has shown that p53Ab1 tends to be more specific for its cognate sequence and near-cognate sequences than does DM1A[11]. Thus, it is perhaps not surprising that the less specific monoclonal would have a less distinct preference with regard to the overall characteristics of the peptide, though it is interesting that the recognition sequence per se is not dominating the charge interaction.

Two types of serum samples were tested, including 8 uninfected individuals and 8 samples from individuals infected with Malaria. Most of the serum samples tested show a similar dependence on total peptide charge, such that the average binding intensity for positively charged peptides is approximated 1.5 to 2-fold higher than that of negatively charged or neutral peptides. However, 6 serum samples from the normal group show in addition to the overall trend a strong preference for very highly negatively charged peptides (peptides with total charge -7 or -6). This preference appears to be individual rather than related to infection with Malaria.

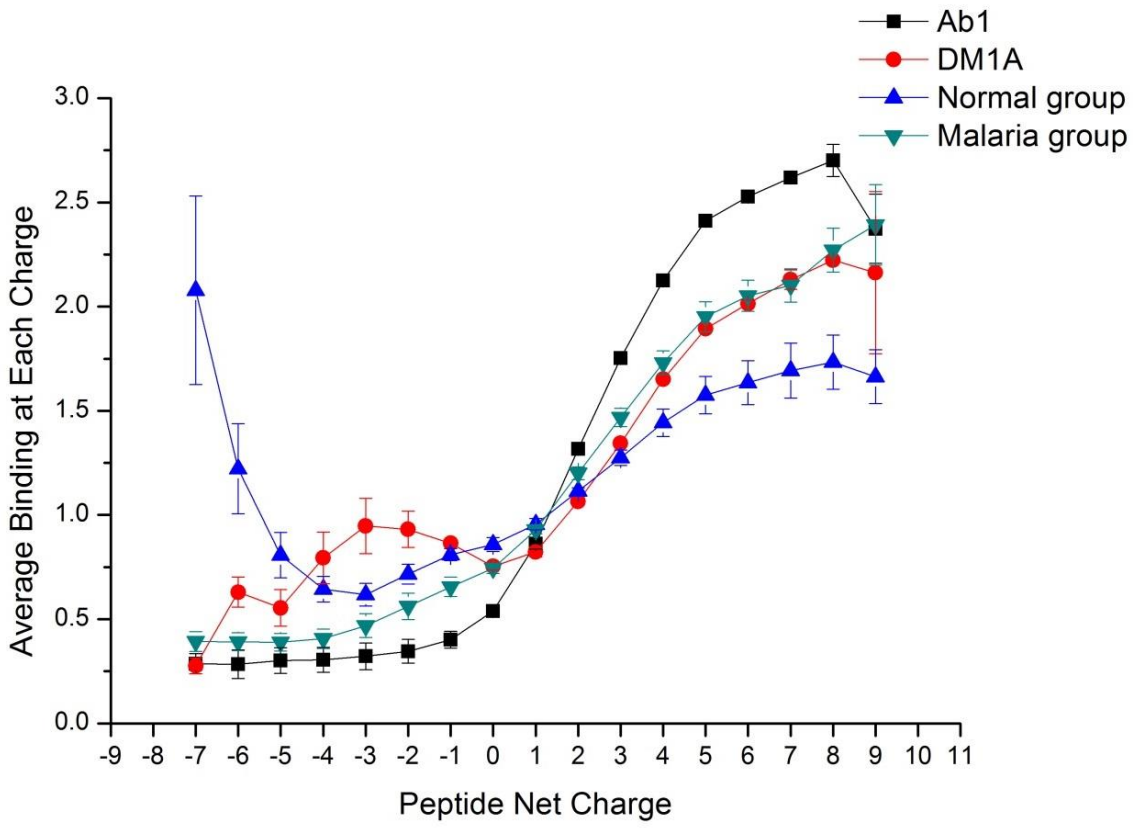

Figure 4.1. Average Binding Intensity as a Function of Net Charge of Peptides. The net charge of the peptides at neutral pH varies from -7 to +9. Intensity values shown are mean normalized fluorescence intensities of dye-labeled secondary antibodies captured by antibodies that are bound to surface bound peptides. The data shown is the average of all samples tested for each sample group (four sample groups: p53Ab1, DM1A, normal serum and malaria serum). The error bars shown represents the standard error of the mean for samples in the same sample group.

Effect of peptide hydrophobicity on IgG binding

Hydrophobic interactions are also important in driving binding interactions. Figure 4.2 shows the average binding signal for the monoclonal antibody and serum samples as a function of peptide hydrophobicity. The hydrophobicity value of each

peptide was estimated from the average of the hydrophobicity index values[14] for each amino acid in the peptide sequence. The peptides were then ranked and evenly sorted into 12 groups, from the least hydrophobic (hydrophilic) to the most hydrophobic, according to their average hydrophobicity values. The effects of hydrophobicity are very similar for all samples. The average binding signal peaks at an average hydrophobicity value between -10 to 0. At the highest and lowest hydrophobicity, the binding for all samples decreases by more than 50% relative to the peak.

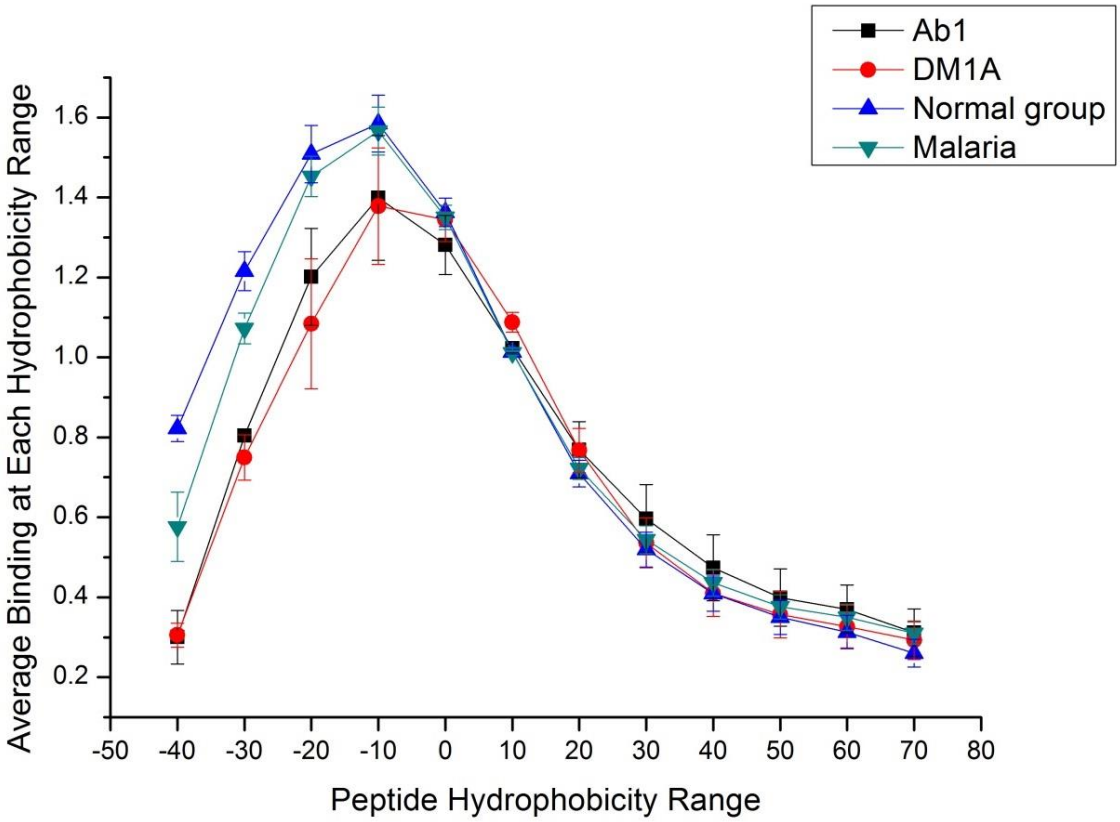

Figure 4.2. Average Binding Intensity as a Function of Peptide Hydrophobicity. The binding intensities from peptides of each hydrophobicity group were averaged together. There were twelve hydrophobicity groups as describe in the text. Intensity values shown

are normalized to the mean for each sample. The data shown is the average of all samples tested for each sample group. The error bars shown represents the standard error of the mean for samples in the same sample group.

Peptide Length Dependence

Normally, increasing the length of a peptide would increase its chemical complexity. Longer peptides have a larger potential number of amino acid combinations. However, the way these sequences were generated, the complexity is, by one measure, less dependent on length. The reason for this is that there is simply a fixed probability that an amino acid will be added at each particular step in the synthesis. Thus short sequences end up being relatively random while longer and longer sequences have more in common with the sequence in which the amino acids were added. Thus if one asks in a 5-mer how often an 'A' follows a 'G' or the other way around, it will be closer to 50/50% than if one measures this for a 15-mer. Though imperfect, this is more a measure of the dependence on length per se than on the order complexity of the sequences that can be made. In previous studies on the binding of β-galactosidase using unbiased sequences of different lengths, the intensity of binding decreased as the length increasing beyond approximately 7 amino acids[15]. This suggested a more complex dependence of binding on peptide length than simply more sites to bind to, and it was suggested that this might have involved entropy restriction effects. Moreover, peptide length is not entirely independent of the net charge. For the chosen peptide sequences on the peptide arrays described here, longer peptides tend to be more positively charged on average. Since average binding intensity for positively charged peptides is higher than that of negatively charged or neutral peptides, one might expect that longer peptides would have higher

binding intensities if charge-charge interactions are dominant. However, the situation is much more complex than that. Although antibodies may prefer charged motifs, the way these peptides were synthesized, it becomes more difficult to have any particular charged motif or fragment within a longer peptide sequence (again because in this approach, longer peptides are more biased towards the sequence in which amino acids were added). Thus, even though the longer peptides are more positively charged than shorter peptides, they may be less likely to contain certain binding sites. As seen in Figure 4.3, for serum samples (uninfected and malaria), the average binding signal peaked at 7 amino acids length and slowly dropped so that by a length of 17 amino acids the average binding is similar to the average binding for peptides of length 3. For monoclonal antibodies, length dependence is more complicated as the specific interactions with the epitope sequence are involved and the length of the corresponding epitope would then have large effect on the average binding. For p53Ab1, the length of the epitope is 5 and the average binding signal for p53Ab1 peaked at 5 and 6 amino acids in length. For DM1A, the small binding peak at length of 7 can be explained by the length of its epitope, which is 7, but the increase in binding at longer lengths may be because DM1A is able to take advantage of a wider variety of interaction sites or that binding is greater when the site is farther from the surface.

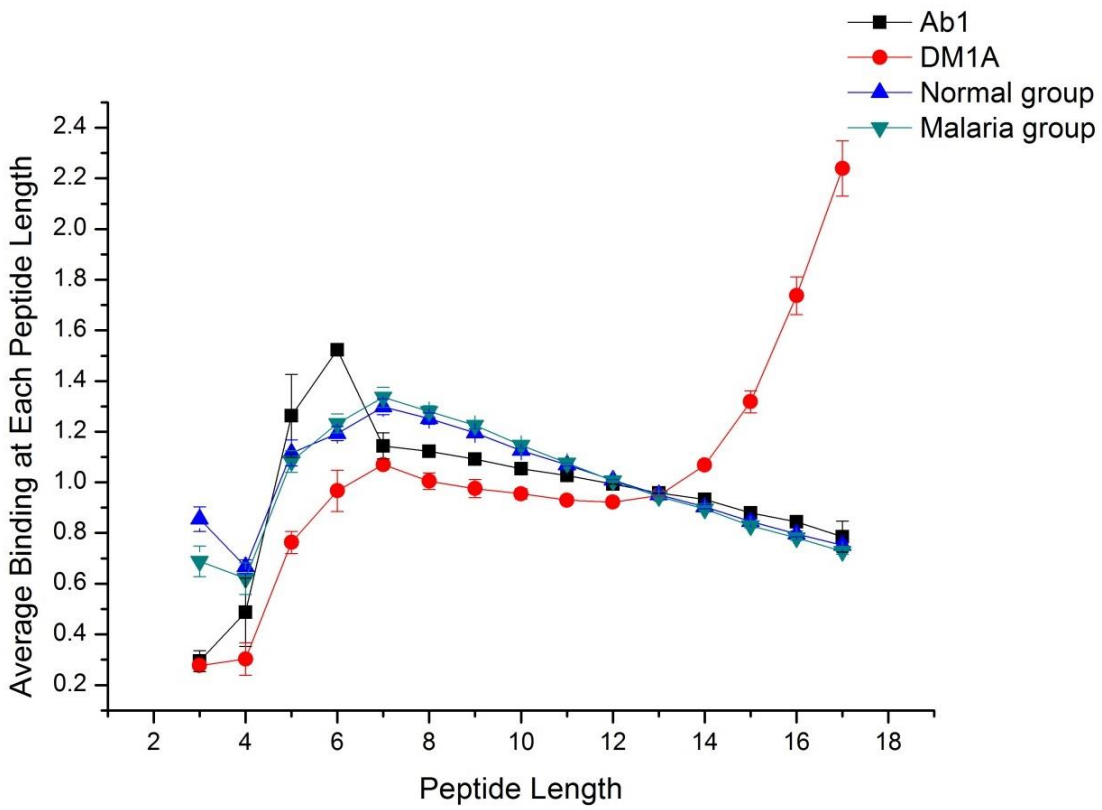

Figure 4.3. Average Binding Intensity as a Function of Length for Each Sample. The binding signals from peptide features of each length were averaged together. Intensity values shown are normalized to the mean for each sample. The data shown is the average of all samples tested for each sample group. The error bars shown represents the standard error of the mean for samples in the same sample group.

Contribution of Specific Amino Acid Residues to Serum Binding

Figure 4.4 shows the contribution of each amino acid to serum binding. The average binding intensities were calculated for peptide sequences containing the same amino acid. More specifically, the binding intensities from all peptides with each particular amino acid residue were averaged together (e.g. All peptide sequences possessing Alanine were averaged together to give the average binding intensity value for

Alanine in Figure 4.4). As seen in Figure 4.4, the average binding intensities for peptide sequences containing Arginine, Lysine, Proline or Glycine are ~20% higher than the global average binding intensity. This suggests that these amino acids may contribute to high binding for serum samples from both the normal and malaria groups. In contrast, the average binding intensities for peptide sequences containing Phenylalanine or Leucine are noticeably lower (more than 20% lower) than the global mean.

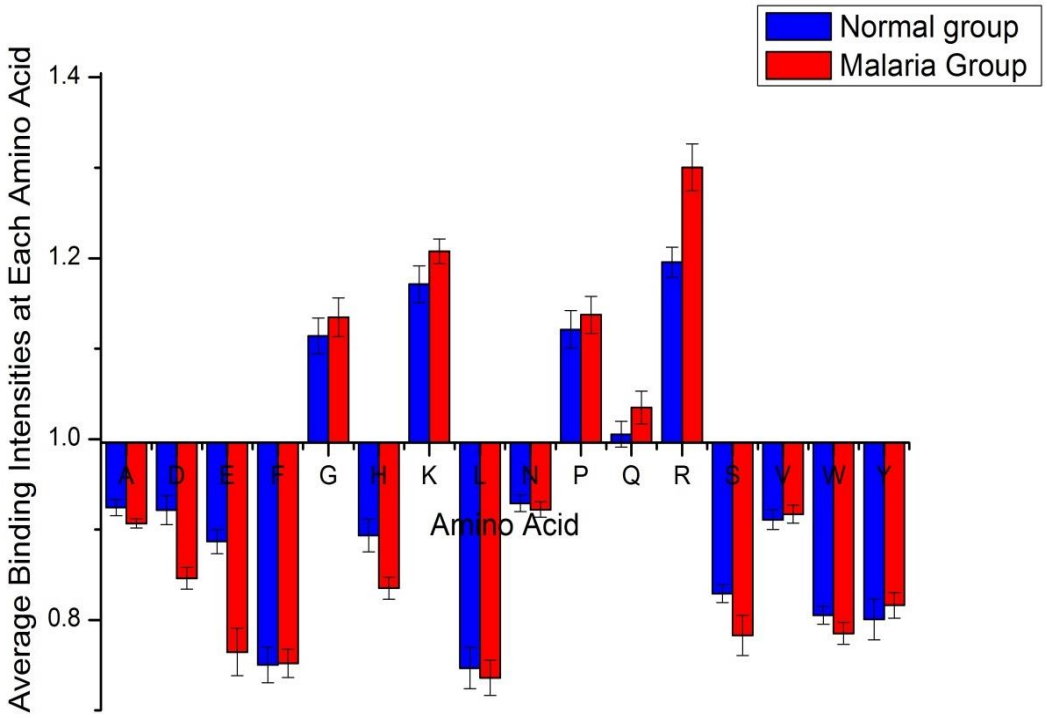

Figure 4.4. Average Binding Intensities as a Function of Amino Acid per Peptide. The binding intensities from peptides with each specific amino acid residue were averaged together. Intensity values shown are normalized to the mean for each sample. The data shown is the average of all samples tested for each sample group. The error bars shown represents the standard error of the mean for samples in the same sample group.

Another way to look at the contribution of specific amino acids to binding would be by calculating the amino acid's frequency of occurrence in the top percentage and bottom percentage of the binding sequences. The results show that Arginine, Lysine, Proline and Glycine are the most frequently occurring amino acids in the top 1% of binding sequences (Figure 4.5) while Phenylalanine and Leucine are the most frequently occurring amino acids in the bottom 1% binding sequences(Figure 4.6), consistent with what was shown in Figure 4.4.

The fact that serum samples prefer amino acids with positively charged side chains such as Arginine and Lysine suggests the importance of charge-charge interactions in serum-peptide binding on the surface. Although serum contains a large number of different biomolecules, the measured binding signals are specifically from IgG molecules and peptides, in part because at the high dilution used only antibodies have small enough dissociation constants to remain bound, and also because the binding is detected with fluorescently labeled anti-human-IgG. Because the pI of IgG molecules is slightly lower than neutral pH, this may promote electrostatic forces and long-range interactions with positively charge amino acid side chains. In addition, the amino acids contributing to high binding intensities here are significantly different from amino acids preferred by the general antibody-epitope interfaces[16, 17]. This suggests that the interactions between serum IgG and surface-bound peptides here may not be driven entirely by specific antibody-antigen interactions; instead, interactions driven by general forces may also play a key role.

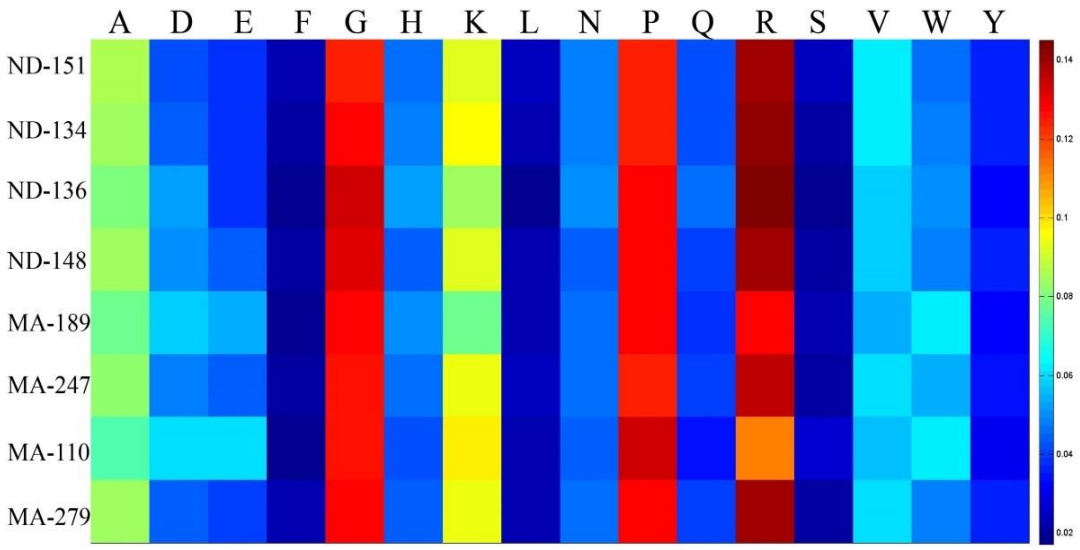

Figure 4.5. A heatmap of the frequency of occurrence of specific amino acids in the top 1% serum binding sequences. Each column represents a different amino acid used. Each row represents a different sample tested (ND represents normal sample and MA represents malaria sample). The color represents the frequency of occurrence of a particular amino acid in the highest 1% binding sequences. A color value of 0.06 represents a frequency of occurrence of a particular amino acid equivalent of that expected from a randomly defined sequence (1/16 since there are 16 different amino acids used).

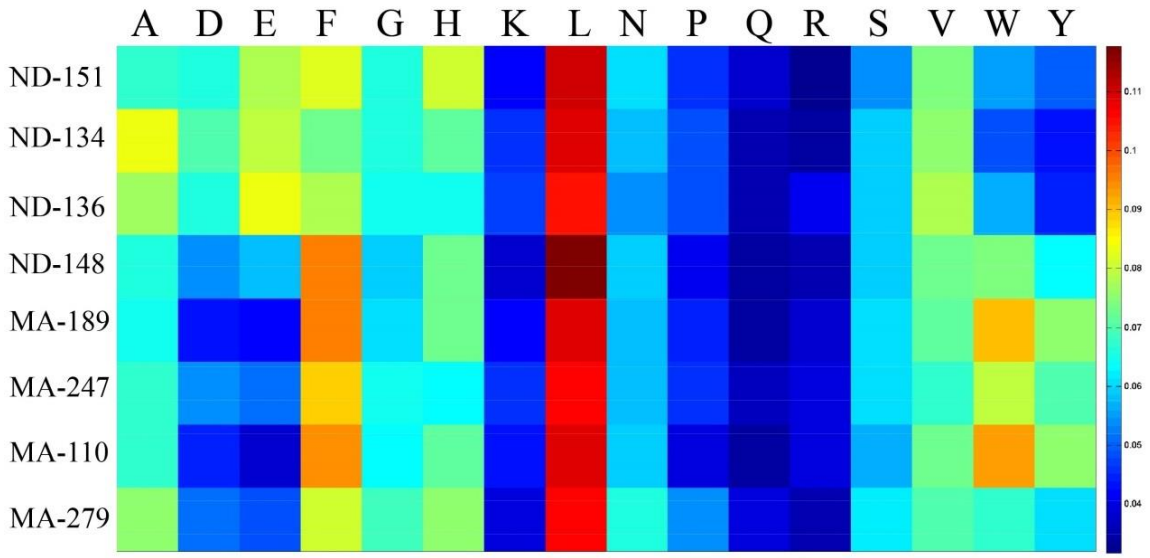

Figure 4.6. A heatmap of the frequency of occurrence of specific amino acids in the bottom 1% serum binding sequences. Each column represents a different amino acid used. Each row represents a different sample tested (ND represents normal sample and MA represents malaria sample). The color represents the frequency of occurrence of a particular amino acid in the lowest 1% binding sequences. A color value of 0.06 represents a frequency of occurrence of a particular amino acid equivalent of that expected from a randomly defined sequence (1/16 since there are 16 different amino acids used).

Effect of Amino Acid Position within a Peptide Sequence

Amino acid position within a peptide sequence was also considered. A 16 (amino acid name) by 17 (amino acid position number, position 1 is N-terminus and position 17 is C-terminus for a peptide with 17 amino acid residues) matrix was created to examine the effect of amino acid position on binding. For each amino acid, the binding intensities for all peptides with the amino acid at the same position were averaged together and

plotted against the amino acid position relative to the N-terminus (Figure 4.7). Generally, the effect on binding of the specific amino acid used is not very dependent on where the amino acid is placed within a peptide. But the position of some amino acids has bigger effect on binding than that of others. For example, the contribution of glycine and proline to the strong binding signals was ~50% higher when they are near the N-terminus compared to the C-terminus, whereas the contribution of lysine remains the roughly same. In general, the effect of an amino acid identity is less significant toward C-terminus. Note however that the position number is defined as the position of an amino acid relative to N-terminus. Because the length of peptides are range from 3 to 17 amino acids, there are fewer and fewer peptides at any given position as one progresses from the N to C terminus.

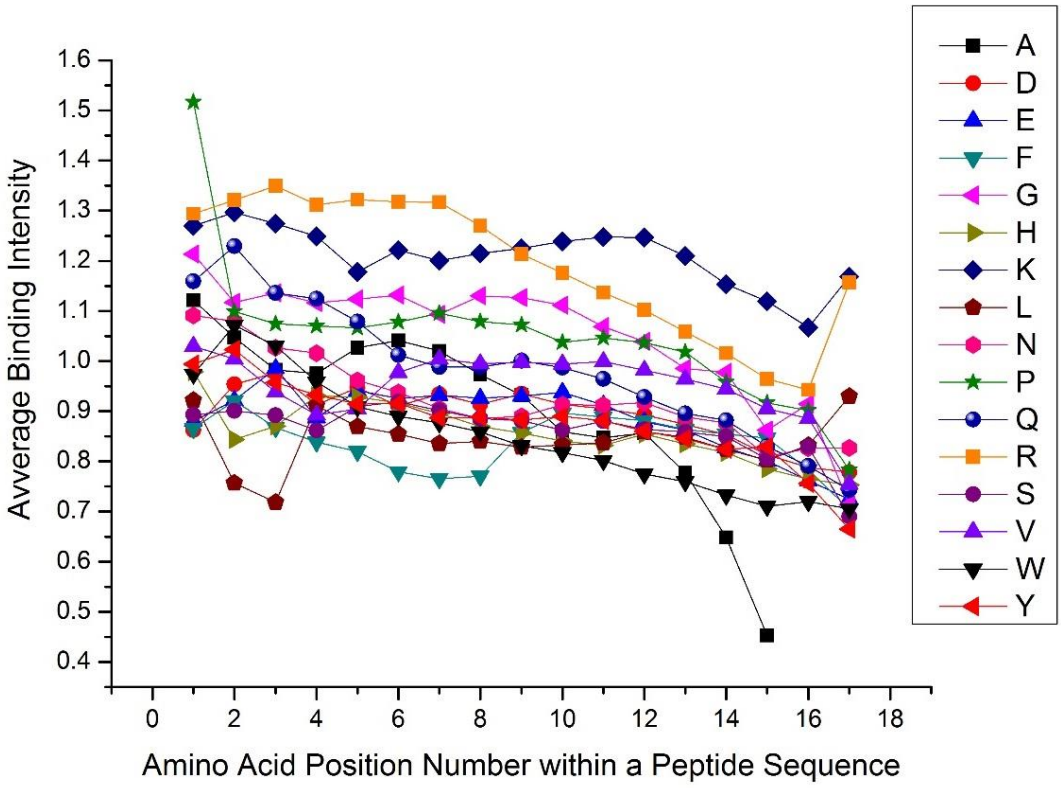

Figure 4.7. Average Binding Intensity as a function of amino acid position within a peptide sequence for each amino acid. The sample used for creating this figure is ND-151, other samples give very similar binding trends (Figures not shown). For each amino acid, the binding intensities for peptides with such amino acid at the same position were averaged together. Intensity values shown are normalized to the mean value.

REFERENCES

[1] Andresen H, Grotzinger C. Deciphering the Antibodyome - Peptide Arrays for Serum Antibody Biomarker Diagnostics. Curr Proteomics 2009;6:1-12.

[2] Sykes KF, Legutki JB, Stafford P. Immunosignaturing: a critical review. Trends Biotechnol 2013;31:45-51.

[3] Legutki JB, Johnston SA. Immunosignatures can predict vaccine efficacy. P Natl Acad Sci USA 2013;110:18614-9.

[4] Restrepo L, Stafford P, Magee DM, Johnston SA. Application of immunosignatures to the assessment of Alzheimer's disease. Annals of neurology 2011;70:286-95.

[5] Hughes AK, Cichacz Z, Scheck A, Coons SW, Johnston SA, Stafford P. Immunosignaturing can detect products from molecular markers in brain cancer. Plos One 2012;7:e40201.

[6] Stafford P, Halperin R, Legutki JB, Magee DM, Galgiani J, Johnston SA. Physical characterization of the "immunosignaturing effect". Mol Cell Proteomics 2012;11:M111 011593.

[7] Brown JR, Stafford P, Johnston SA, Dinu V. Statistical methods for analyzing immunosignatures. Bmc Bioinformatics 2011;12:349.

[8] Legutki JB, Zhao ZG, Greving JP, Woodbury NW, Johnston SA, Stafford P. Scalable High-Density Peptide Arrays for Comprehensive Health Monitoring Nature Communications 2014.(In Press)

[9] Stafford P, Cichacz Z, Woodbury NW, Johnston SA. An Immunosignature system for diagnosis of cancer. Proc Natl Acad Sci U S A 2014. (Under Review)

[10] James LC, Tawfik DS. The specificity of cross-reactivity: promiscuous antibody binding involves specific hydrogen bonds rather than nonspecific hydrophobic stickiness. Protein Sci 2003;12:2183-93.

[11] Halperin RF, Stafford P, Johnston SA. Exploring antibody recognition of sequence space through random-sequence peptide microarrays. Mol Cell Proteomics 2011;10:M110 000786.

[12] Reverberi R, Reverberi L. Factors affecting the antigen-antibody reaction. Blood transfusion = Trasfusione del sangue 2007;5:227-40.

[13] Van Oss CJ. Hydrophobic, hydrophilic and other interactions in epitope-paratope binding. Molecular immunology 1995;32:199-211.

[14] Monera OD, Sereda TJ, Zhou NE, Kay CM, Hodges RS. Relationship of Sidechain Hydrophobicity and alpha-Helical Propensity on the Stability of the Single-stranded Amphipathic alpha-Helix. J Pept Sci 1995;1:319-29.

[15] Wang W, Woodbury NW. Selective protein-peptide interactions at surfaces. Acta Biomater 2014;10:761-8.

[16] Soga S, Kuroda D, Shirai H, Kobori M, Hirayama N. Use of amino acid composition to predict epitope residues of individual antibodies. Protein Engineering Design & Selection 2010;23:441-8.

[17] Soga S, Shirai H, Kobori M, Hirayama N. Use of amino acid composition to predict ligand-binding sites. Journal of chemical information and modeling 2007;47:400-6.

CHAPTER 5: CONCLUSIONS AND FUTURE CONSIDERATIONS

The interaction of proteins with chemically complex surfaces is of critical significance in both biological and man-made systems. A complete understanding of the protein-surface interactions is useful because it will potentially tell us how proteins function in many biological processes involving cellular interfaces and contribute to developing surfaces with selective binding or nonbinding properties for practical use. Up to this time, the majority of the studies on molecular interactions have mainly focused on specific interactions with well-defined final complexes. In contrast, interactions that do not result in a unique structure for the final complex are usually considered nonspecific and treated as a nuisance to experiments. However, as described in chapter 1, unstructured interactions are more ubiquitous in the cellular environments and participate in many biological processes. Therefore, there is a need to characterize and further study these interactions. In fact, the strength of either type of interactions largely comes from long-range forces such as general charge-charge interactions (examples include DNA binding proteins, antigen-antibody interactions). While for a specific interaction, the final structure is defined by interactions like specific hydrogen bonds or short-range van der Waals forces and the affinity depends on the goodness of fit between the interacting surfaces; for an unstructured interaction or so called nonspecific interaction, the affinity is driven by longer range forces only or by a summation of many interactions that allow a set of heterogeneous binding structures instead of stabilizing one particular binding structure.

It is difficult to detect and characterize unstructured interactions with conventional techniques. As mentioned in Chapter 1, most existing approaches to study

molecular interactions are trying to search and characterize the most favored single structure among all the complexes formed. This makes these approaches not suitable for studying the unstructured interactions because an unstructured interaction does not have a predominant binding structure, instead, a variety of low affinity structures that are equally important are often formed in one interaction. The research projects in this dissertation demonstrated a statistical approach to study molecular interactions, including both structured and unstructured interactions, between proteins and heterogonous surfaces covered with peptides. Unlike most classical approaches for studying molecular interactions, this approach does not force the system to converge on a single binding structure; instead, it allows systematic analysis of binding to diverse arrays with a large number of potential binding partners and uses the binding information to understand the chemistry of molecular interactions at surfaces.

Chapter 2 described the binding of β-galactosidase, α1-antitrypsin and a protein mixture with a set of 9 different soluble proteins to the LC Sciences array surface-attached peptides from a library, covering lengths from 2 to 12 amino acids and sequence permutations of 7 amino acids {E, L, S, R, Q, W, Y}, representing a broad span of overall charge and hydrophobicity. In spite of apparently unstructured interactions of the immobilized peptides with the proteins of interest, the interactions can be quiet strong and diverse. In addition, considerable selectivity of protein binding to the peptide-coated surface can be obtained. The work presented in Chapter 2 supports the concept that specific and nonspecific binding are better thought of as two ends of amore continuous spectrum and suggests the possibility of creating surfaces covered with short peptide sequences with relative specific protein interaction profiles. However, due to the

limitations associated with LC Sciences PepArray platform (e.g. cost, lack of controls on binding experiments, etc), although the work was able to demonstrate the possibility of using peptide arrays in exploring the interaction space of proteins with complex surfaces, additional experiments are required to fully understand the concepts. For example, a larger number of additional individual proteins with different physicochemical properties and of different origin could be tested with the same surface and compare to the binding profiles of β-galactosidase and α1-antitrypsin. This would allow a more extensive study on the correlation between the binding profile and the protein physicochemical properties. For instance, the correlation between the binding profile and the protein isoelectric point could be examined. Similarly, the hydrophobicity of the proteins can be taken into account by using the accessible area of hydrophobic amino acid residues. This can further verify the driving forces for unstructured surface interactions and how they contribute to discriminating proteins. In addition, the binding experiments could be performed at variable pH and ionic strength, to estimate relative contributions of electrostatic forces and hydrophobic interactions into the surface peptide-protein binding capacity. Another interesting question is whether using a larger subset of amino acids as building blocks would change the correlation between surface binding and peptide sequence properties. Creating a peptide library using an increased amino acid alphabet size would increase the chemical complexity represented by the library, though a smaller percentage of the full sequence space would be covered and used for the statistical study.

Since short surface-bound peptides can have both strong and selective interactions with proteins, one interesting fundamental question to ask is how addition or combination of short peptide sequences change the properties of a surface covered with a particular

short peptide. The answers to these questions are very important for developing the principles needed to engineer new surfaces with specific properties using relatively short peptides. The strategy employed in Chapter 3 for preparing peptide arrays, which involved the use of selected tetrapeptides with known binding characteristics to β-galactosidase as C-terminal bases tethered to the surface, and either an N-terminal extension by one, two or three amino acids, or a concatenation with another tetrapeptide, were designed to answer these questions. One of the most important observations is that while both additions and combinations of short peptide sequences can either greatly increase or decrease the binding and relative specificity of a particular tetrapeptide on the surface, the contribution to the binding affinity and relative specificity of the individual binding components is strongly dependent on their position within the peptide sequence. Moreover, component sequences that bind strongly at the C-terminus of the peptide do not necessarily add substantially to binding and specificity when placed at the N-terminus. It is important to note that although a surface covered with random-sequence peptides with longer lengths may possess similar or even higher chemical complexity and degree of differential binding, short peptides are easier and cheaper to use for surface modification.

Due to the limitations associated with LC Sciences arrays, it would be useful to perform similar studies with a different array system. As described in Chapter 1, the HealthTell array system is another array platform that is completely different from the LC Sciences array system. HealthTell Arrays are in a different format and are built on a different substrate. These arrays are developed for the purpose of performing immunosignature diagnostics, which attempts to identify a unique binding pattern of

circulating antibodies in serum that differentiates healthy individuals from those with a particular disease. In chapter 4, fundamental questions about general interactions between antibodies and peptide attached to surfaces are answered by systematically studying the binding of monoclonal antibodies and serum to the HealthTell array surface covered with particular peptides, from a library consisting of ~330,000 peptide sequences with a length range from 3 to 17 amino acids and amino acid alphabet size of 16 (cysteine, methionine, isoleucine and threonine were excluded). The peptide sequences include certain epitope sequences and random sequences that are not derived from any natural binding sequence space. While it is confirmed that the linear epitopes can be recognized and specifically targeted upon binding to the antibodies, the random sequences may or may not behave the same. The interactions driven by general forces may play an important role in peptide-surface binding with HealthTell array system. Consistent with what was seen before, the interactions can be strong and diverse. Effects of charge, hydrophobicity, length and amino acid position on IgG binding on HealthTell array surface are quiet similar to what was observed with β-galactosidase on LC Science array surface. This suggests that general interactions play a role in addition to any cognate-type binding. In addition, positively charged amino acids are preferred by serum samples further confirming the importance of general electrostatic interactions in serum binding on surfaces.

To further study the general interactions between antibodies and the HealthTell array surface, a number of other sequence properties of surface-bound peptides can be examined and correlated to the binding of antibodies to the surfaces. For example, sequence complexity, flexibility and accessibility can be considered. These studies may

provide information on the entropic cost upon binding. Also, the binding assays could be performed at a variable pH and instead of looking at the total charge for each peptide at neutral pH, the isoelectric point could be considered. It may also worthwhile to look at the correlation between binding and the position of certain tetrapeptides within a sequence. Although it has been shown that the binding is not strongly dependent on the position of any single amino acid within the sequence, position of tetrapeptides or longer binding units within a sequence might still be important (As in Chapter 3, single amino acid position is not important but the tetrapeptide position is important in term of the contribution to binding).

In summary, unstructured affinity driven by multiple weak interactions plays an important role in many biological processes. As our understanding of the crowded environment of the cell increases, it has become more and more apparent how important controlling such interactions is. Although still in early stages, this dissertation work demonstrates the possibility of using peptide arrays in exploring the interaction space of proteins with surfaces. With the technology advancing and expanding rapidly, there is tremendous space to explore many more studies with this approach, and we could expect a brand new way to create complex surfaces for our use.

COMPLETE LIST OF REFERENCES

Chapter 1

[1] Ostuni E, Yan L, Whitesides GM. The interaction of proteins and cells with self-assembled monolayers of alkanethiolates on gold and silver. Colloid Surface B 1999;15:3-30.

[2] Volkmer R, Tapia V, Landgraf C. Synthetic peptide arrays for investigating protein interaction domains. FEBS Lett 2012;586:2780-6.

[3] Zhu X, Gerstein M, Snyder M. Getting connected: analysis and principles of biological networks. Genes & development 2007;21:1010-24.

[4] Mager MD, LaPointe V, Stevens MM. Exploring and exploiting chemistry at the cell surface. Nature chemistry 2011;3:582-9.

[5] Zaia J. At last, functional glycomics. Nature methods 2011;8:55-7.

[6] Taniguchi CM, Emanuelli B, Kahn CR. Critical nodes in signalling pathways: insights into insulin action. Nature reviews Molecular cell biology 2006;7:85-96.

[7] Kasemo B. Biological surface science. Surface science 2002;500:656-77.

[8] Ratner BD. The engineering of biomaterials exhibiting recognition and specificity. J Mol Recognit 1996;9:617-25.

[9] Gopel W. Bioelectronics and nanotechnologies. Biosens Bioelectron 1998;13:723-8.

[10] Andresen H, Grotzinger C. Deciphering the Antibodyome - Peptide Arrays for Serum Antibody Biomarker Diagnostics. Curr Proteomics 2009;6:1-12.

[11] Graves DJ. Powerful tools for genetic analysis come of age. Trends in Biotechnology 1999;17:127-34.

[12] Ball V, Maechling C. Isothermal microcalorimetry to investigate non specific interactions in biophysical chemistry. Int J Mol Sci 2009;10:3283-315.

[13] Johnson ME, Hummer G. Nonspecific binding limits the number of proteins in a cell and shapes their interaction networks. Proc Natl Acad Sci U S A 2011;108:603-8.

[14] Kalodimos CG, Biris N, Bonvin AM, Levandoski MM, Guennuegues M, Boelens R, et al. Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. Science 2004;305:386-9.

[15] Halford SE, Marko JF. How do site-specific DNA-binding proteins find their targets? Nucleic Acids Res 2004;32:3040-52.

[16] Tan JM, Wong ES, Lim KL. Protein misfolding and aggregation in Parkinson's disease. Antioxidants & redox signaling 2009;11:2119-34.

[17] Bucciantini M, Giannoni E, Chiti F, Baroni F, Formigli L, Zurdo J, et al. Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. Nature 2002;416:507-11.

[18] White AD, Nowinski AK, Huang WJ, Keefe AJ, Sun F, Jiang SY. Decoding nonspecific interactions from nature. Chem Sci 2012;3:3488-94.

[19] White AD, Huang WJ, Jiang SY. Role of Nonspecific Interactions in Molecular Chaperones through Model-Based Bioinformatics. Biophys J 2012;103:2484-91.

[20] Stafford P, Halperin R, Legutki JB, Magee DM, Galgiani J, Johnston SA. Physical characterization of the "immunosignaturing effect". Mol Cell Proteomics 2012;11:M111 011593.

[21] Dey B, Thukral S, Krishnan S, Chakrobarty M, Gupta S, Manghani C, et al. DNA-protein interactions: methods for detection and analysis. Mol Cell Biochem 2012;365:279-99.

[22] Berggard T, Linse S, James P. Methods for the detection and analysis of protein-protein interactions. Proteomics 2007;7:2833-42.

[23] Phizicky EM, Fields S. Protein-protein interactions: methods for detection and analysis. Microbiological reviews 1995;59:94-123.

[24] Miernyk JA, Thelen JJ. Biochemical approaches for discovering protein-protein interactions. The Plant journal : for cell and molecular biology 2008;53:597-609.

[25] Shoemaker BA, Panchenko AR. Deciphering protein-protein interactions. Part I. Experimental techniques and databases. Plos Comput Biol 2007;3:e42.

[26] Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol 1999;17:1030-2.

[27] Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, et al. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods 2001;24:218-29.

[28] Bauch A, Superti-Furga G. Charting protein complexes, signaling pathways, and networks in the immune system. Immunological reviews 2006;210:187-207.

[29] Wilson DS, Szostak JW. In vitro selection of functional nucleic acids. Annu Rev Biochem 1999;68:611-47.

[30] Smith GP. Filamentous Fusion Phage - Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. Science 1985;228:1315-7.

[31] Keefe AD, Szostak JW. Functional proteins from a random-sequence library. Nature 2001;410:715-8.

[32] Hanes J, Pluckthun A. In vitro selection and evolution of functional proteins by using ribosome display. P Natl Acad Sci USA 1997;94:4937-42.

[33] Clackson T, Wells JA. In vitro selection from protein and peptide libraries. Trends Biotechnol 1994;12:173-84.

[34] Zahnd C, Amstutz P, Pluckthun A. Ribosome display: selecting and evolving proteins in vitro that specifically bind to a target. Nature methods 2007;4:269-79.

[35] Sutherland BW, Toews J, Kast J. Utility of formaldehyde cross-linking and mass spectrometry in the study of protein-protein interactions. J Mass Spectrom 2008;43:699-715.

[36] Fu JL, Reinhold J, Woodbury NW. Peptide-Modified Surfaces for Enzyme Immobilization. Plos One 2011;6.

[37] Woodson M, Liu J. Functional nanostructures from surface chemistry patterning. Phys Chem Chem Phys 2007;9:207-25.

[38] Gao X, Pellois JP, Na Y, Kim Y, Gulari E, Zhou X. High density peptide microarrays. In situ synthesis and applications. Molecular diversity 2004;8:177-87.

[39] Katz C, Levy-Beladev L, Rotem-Bamberger S, Rito T, Rudiger SG, Friedler A. Studying protein-protein interactions using peptide arrays. Chemical Society reviews 2011;40:2131-45.

[40] Morales Betanzos C, Gonzalez-Moa MJ, Boltz KW, Vander Werf BD, Johnston SA, Svarovsky SA. Bacterial glycoprofiling by using random sequence peptide microarrays. Chembiochem : a European journal of chemical biology 2009;10:877-88.

[41] Legutki JB, Magee DM, Stafford P, Johnston SA. A general method for characterization of humoral immunity induced by a vaccine or infection. Vaccine 2010;28:4529-37.

[42] Breitling F, Nesterov A, Stadler V, Felgenhauer T, Bischoff FR. High-density peptide arrays. Mol Biosyst 2009;5:224-34.

[43] Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. Light-directed, spatially addressable parallel chemical synthesis. Science 1991;251:767-73.

[44] http://www.lcsciences.com/applications/proteomics/.

[45] Gao XL, Zhou XC, Gulari E. Light directed massively parallel on-chip synthesis of peptide arrays with t-Boc chemistry. Proteomics 2003;3:2135-41.

[46] Pellois JP, Zhou X, Srivannavit O, Zhou T, Gulari E, Gao X. Individually addressable parallel peptide synthesis on microchips. Nat Biotechnol 2002;20:922-6.

[47] Pellois JP, Wang W, Gao X. Peptide synthesis based on t-Boc chemistry and solution photogenerated acids. Journal of combinatorial chemistry 2000;2:355-60.

[48] Stewart JM, Young,J.D. Solid Phase Peptide Synthesis. 2nd edition ed: Pierce Chemical Co; 1984.

[49] Gao X, Zhou X, Gulari E. Light directed massively parallel on-chip synthesis of peptide arrays with t-Boc chemistry. Proteomics 2003;3:2135-41.

[50] Matthews BW. The structure of E. coli beta-galactosidase. Comptes rendus biologies 2005;328:549-56.

[51] Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 1961;3:318-56.

[52] Kalnins A, Otto K, Ruther U, Muller-Hill B. Sequence of the lacZ gene of Escherichia coli. The EMBO journal 1983;2:593-7.

[53] Jacobson RH, Zhang XJ, DuBose RF, Matthews BW. Three-dimensional structure of beta-galactosidase from E. coli. Nature 1994;369:761-6.

[54] Juers DH, Jacobson RH, Wigley D, Zhang XJ, Huber RE, Tronrud DE, et al. High resolution refinement of beta-galactosidase in a new crystal form reveals multiple metal-binding sites and provides a structural basis for alpha-complementation. Protein Sci 2000;9:1685-99.

[55] Welply JK, Fowler AV, Zabin I. Beta-Galactosidase Alpha-Complementation - Effect of Single Amino-Acid Substitutions. J Biol Chem 1981;256:6811-6.

[56] Welply JK, Fowler AV, Zabin I. Beta-Galactosidase Alpha-Complementation - Overlapping Sequences. J Biol Chem 1981;256:6804-10.

[57] Juers DH, Matthews BW, Huber RE. LacZ beta-galactosidase: structure and function of an enzyme of historical and molecular biological importance. Protein Sci 2012;21:1792-807.

[58] Accolla RS, Cina R, Montesoro E, Celada F. Antibody-mediated activation of genetically defective Escherichia coli beta-galactosidases by monoclonal antibodies produced by somatic cell hybrids. Proc Natl Acad Sci U S A 1981;78:2478-82.

[59] Roth RA, Rotman B. Inactivation of normal beta-D-galactosidase by antibodies to defective forms of the enzyme. J Biol Chem 1975;250:7759-65.

[60] Fu J, Cai K, Johnston SA, Woodbury NW. Exploring peptide space for enzyme modulators. J Am Chem Soc 2010;132:6419-24.

[61] Fu J, Reinhold J, Woodbury NW. Peptide-modified surfaces for enzyme immobilization. Plos One 2011;6:e18692.

[62] Kalsheker N. Alpha 1-antitrypsin: structure, function and molecular biology of the gene. Biosci Rep 1989;9:129-38.

[63] Crystal RG. The alpha 1-antitrypsin gene and its deficiency states. Trends in genetics : TIG 1989;5:411-7.

[64] Carrell RW, Jeppsson JO, Vaughan L, Brennan SO, Owen MC, Boswell DR. Human alpha 1-antitrypsin: carbohydrate attachment and sequence homology. FEBS Lett 1981;135:301-3.

[65] Huntington JA, Read RJ, Carrell RW. Structure of a serpin-protease complex shows inhibition by deformation. Nature 2000;407:923-6.

[66] Patschull AOM, Segu L, Nyon MP, Lomas DA, Nobeli I, Barrett TE, et al. Therapeutic target-site variability in a1-antitrypsin characterized at high resolution. Acta Crystallogr F 2011;67:1492-7.

[67] Elliott PR, Pei XY, Dafforn TR, Lomas DA. Topography of a 2.0 angstrom structure of alpha(1)-antitrypsin reveals targets for rational drug design to prevent conformational disease. Protein Sci 2000;9:1274-81.

[68] Parfrey H, Mahadeva R, Ravenhill NA, Zhou AW, Dafforn TR, Foreman RC, et al. Targeting a surface cavity of alpha(1)-antitrypsin to prevent conformational disease. J Biol Chem 2003;278:33060-6.

[69] Gooptu B, Lomas DA. Conformational Pathology of the Serpins: Themes, Variations, and Therapeutic Strategies. Annu Rev Biochem 2009;78:147-76.

[70] Lomas DA, Evans DL, Finch JT, Carrell RW. The Mechanism of Z-Alpha-1-Antitrypsin Accumulation in the Liver. Nature 1992;357:605-7.

[71] Chang YP, Mahadeva R, Chang WSW, Lin SC, Chu YH. Small-molecule peptides inhibit Z alpha(1)-antitrypsin polymerization. J Cell Mol Med 2009;13:2304-16.

[72] Mahadeva R, Dafforn TR, Carrelll RW, Lomas DA. 6-mer peptide selectively anneals to a pathogenic serpin conformation and blocks polymerization - Implications for the prevention of Z alpha(1)-antitrypsin-related cirrhosis. J Biol Chem 2002;277:6771-4.

[73] Chang YP, Mahadeva R, Chang WSW, Shukla A, Dafforn TR, Chu YH. Identification of a 4-mer peptide inhibitor that effectively blocks the polymerization of pathogenic Z alpha(1)-antitrypsin. Am J Resp Cell Mol 2006;35:540-8.

[74] Delves PJ, Martin SJ, Burton DR. Essentials : Roitt's Essential Immunology (12th Edition). Hoboken, NJ, USA: Wiley-Blackwell; 2011.

[75] Woof JM, Burton DR. Human antibody-Fc receptor interactions illuminated by crystal structures. Nature reviews Immunology 2004;4:89-99.

[76] Halperin RF, Stafford P, Johnston SA. Exploring antibody recognition of sequence space through random-sequence peptide microarrays. Mol Cell Proteomics 2011;10:M110 000786.

[77] Reineke U, Ivascu C, Schlief M, Landgraf C, Gericke S, Zahn G, et al. Identification of distinct antibody epitopes and mimotopes from a peptide array of 5520 randomly generated sequences. Journal of immunological methods 2002;267:37-51.

Chapter 2

[1] Ostuni E, Yan L, Whitesides GM. The interaction of proteins and cells with self-assembled monolayers of alkanethiolates on gold and silver. Colloid Surface B 1999;15:3-30.

[2] Johnson ME, Hummer G. Nonspecific binding limits the number of proteins in a cell and shapes their interaction networks. P Natl Acad Sci USA 2011;108:603-8.

[3] Andresen H, Grotzinger C. Deciphering the Antibodyome - Peptide Arrays for Serum Antibody Biomarker Diagnostics. Curr Proteomics 2009;6:1-12.

[4] Lee MH, Brass DA, Morris R, Composto RJ, Ducheyne P. The effect of non-specific interactions on cellular adhesion using model surfaces. Biomaterials 2005;26:1721-30.

[5] Vonhippe.Ph, Mcghee JD. DNA-Protein Interactions. Annu Rev Biochem 1972;41:231-&.

[6] Halford SE, Marko JF. How do site-specific DNA-binding proteins find their targets? Nucleic Acids Res 2004;32:3040-52.

[7] Oda M, Furukawa K, Ogata K, Sarai A, Nakamura H. Thermodynamics of specific and non-specific DNA binding by the c-Myb DNA-binding domain. J Mol Biol 1998;276:571-90.

[8] Woodson M, Liu J. Functional nanostructures from surface chemistry patterning. Phys Chem Chem Phys 2007;9:207-25.

[9] Gilles S, Winter S, Michael KE, Meffert SH, Li PG, Greben K, et al. Control of Cell Adhesion and Neurite Outgrowth by Patterned Gold Nanoparticles with Tunable Attractive or Repulsive Surface Properties. Small 2012;8:3357-67.

[10] Yu TY, Wang Q, Johnson DS, Wang MD, Ober CK. Functional hydrogel surfaces: Binding kinesin-based molecular motor proteins to selected patterned sites. Adv Funct Mater 2005;15:1303-9.

[11] Sykes KF, Legutki JB, Stafford P. Immunosignaturing: a critical review. Trends Biotechnol 2013;31:45-51.

[12] Hughes AK, Cichacz Z, Scheck A, Coons SW, Johnston SA, Stafford P. Immunosignaturing can detect products from molecular markers in brain cancer. Plos One 2012;7:e40201.

[13] http://www.lcsciences.com/applications/proteomics/binding-arrays/.

[14] Greving MP. Creating high-affinity ligands on surfaces and in solution: Arizona State University; 2009.

[15] Kyte J, Doolittle RF. A Simple Method for Displaying the Hydropathic Character of a Protein. J Mol Biol 1982;157:105-32.

[16] Pellois JP, Zhou XC, Srivannavit O, Zhou TC, Gulari E, Gao XL. Individually addressable parallel peptide synthesis on microchips. Nat Biotechnol 2002;20:922-6.

[17] Pilipenko OS, Atyaksheva LF, Poltorak OM, Chukhrai ES. Dissociation and catalytic activity of oligomer forms of beta-galactosidases. Russ J Phys Chem a+ 2007;81:990-4.

[18] Wallenfels K, Weil R. The Enzymes. P.D. Boyer ed1972. p. 617.

[19] Brown WT. Alpha-1-Antitrypsin - Apparent Molecular-Weight Heterogeneity Shown by Two-Dimensional Electrophoresis. Am J Hum Genet 1982;34:195-208.

[20] Spiro RG. Studies on Fetuin, a Glycoprotein of Fetal Serum .1. Isolation, Chemical Composition, and Physicochemical Properties. J Biol Chem 1960;235:2860-9.

[21] Welinder KG. Amino-Acid Sequence Studies of Horseradish-Peroxidase .4. Amino and Carboxyl Termini, Cyanogen-Bromide and Tryptic Fragments, the Complete Sequence, and Some Structural Characteristics of Horseradish Peroxidase-C. Eur J Biochem 1979;96:483-502.

[22] Shannon LM, Kay E, Lew JY. Peroxidase Isozymes from Horseradish Roots .I. Isolation and Physical Properties. J Biol Chem 1966;241:2166-&.

[23] Hirayama K, Akashi S, Furuya M, Fukuhara K. Rapid Confirmation and Revision of the Primary Structure of Bovine Serum-Albumin by Esims and Frit-Fab Lc Ms. Biochem Bioph Res Co 1990;173:639-46.

[24] Kaplan LJ, Foster JF. Isoelectric Focusing Behavior of Bovine Plasma Albumin, Mercaptalbumin, and Beta-Lactoglobulins-a and B. Biochemistry-Us 1971;10:630-&.

[25] Lindskog S. Purification and Properties of Bovine Erythrocyte Carbonic Anhydrase. Biochim Biophys Acta 1960;39:218-26.

[26] Betgovargez E, Knudson V, Simonian MH. Characterization of proteins in the human serum proteome. J Biomol Tech 2005;16:306-10.

[27] Najafi MF, Deobagkar D. Purification and characterization of an extracellular alpha-amylase from Bacillus subtilis AX20. Protein Expr Purif 2005;41:349-54.

[28] Tsuge H, Natsuaki O, Ohashi K. Purification, Properties, and Molecular Features of Glucose Oxidase from Aspergillus-Niger. J Biochem-Tokyo 1975;78:835-43.

[29] Pazur JH, Kleppe K. The Oxidation of Glucose and Related Compounds by Glucose Oxidase from Aspergillus Niger. Biochemistry-Us 1964;3:578-83.

[30] Diehnelt CW, Shah M, Gupta N, Belcher PE, Greving MP, Stafford P, et al. Discovery of High-Affinity Protein Binding Ligands - Backwards. Plos One 2010;5.

[31] Oliveros JC. VENNY. An interactive tool for comparing lists with Venn Diagrams.http://bioinfogp.cnb.csic.es/tools/venny/index.html. 2007.

[32] Monera OD, Sereda TJ, Zhou NE, Kay CM, Hodges RS. Relationship of Sidechain Hydrophobicity and alpha-Helical Propensity on the Stability of the Single-stranded Amphipathic alpha-Helix. J Pept Sci 1995;1:319-29.

[33] Fu JL, Reinhold J, Woodbury NW. Peptide-Modified Surfaces for Enzyme Immobilization. Plos One 2011;6.

Chapter 3

[1] Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets - The roles of intrinsic disorder in protein interaction networks. Febs Journal 2005;272:5129-48.

[2] Ratner BD, Bryant SJ. Biomaterials: Where we have been and where we are going. Annu Rev Biomed Eng 2004;6:41-75.

[3] Magee J, Warwicker J. Simulation of non-specific protein-mRNA interactions. Nucleic Acids Res 2005;33:6694-9.

[4] Halford SE, Marko JF. How do site-specific DNA-binding proteins find their targets? Nucleic Acids Res 2004;32:3040-52.

[5] Schultz SC, Shields GC, Steitz TA. Crystal-Structure of a Cap-DNA Complex - the DNA Is Bent by 90-Degrees. Science 1991;253:1001-7.

[6] Ganguly A, Rajdev P, Williams SM, Chatterji D. Nonspecific Interaction between DNA and Protein allows for Cooperativity: A Case Study with Mycobacterium DNA Binding Protein. J Phys Chem B 2012;116:621-32.

[7] White AD, Nowinski AK, Huang WJ, Keefe AJ, Sun F, Jiang SY. Decoding nonspecific interactions from nature. Chem Sci 2012;3:3488-94.

[8] White AD, Huang WJ, Jiang SY. Role of Nonspecific Interactions in Molecular Chaperones through Model-Based Bioinformatics. Biophys J 2012;103:2484-91.

[9] Johnson ME, Hummer G. Nonspecific binding limits the number of proteins in a cell and shapes their interaction networks. P Natl Acad Sci USA 2011;108:603-8.

[10] Phillips SE, Ile KE, Boukhelifa M, Huijbregts RPH, Bankaitis VA. Specific and nonspecific membrane-binding determinants cooperate in targeting phosphatidylinositol transfer protein beta-isoform to the mammalian trans-golgi network. Mol Biol Cell 2006;17:2498-512.

[11] Fu JL, Cai K, Johnston SA, Woodbury NW. Exploring Peptide Space for Enzyme Modulators. J Am Chem Soc 2010;132:6419-24.

[12] Wang W, Woodbury NW. Selective protein-peptide interactions at surfaces. Acta Biomater 2014;10:761-8.

[13] http://www.lcsciences.com/applications/proteomics/.

[14] Gao XL, Zhou XC, Gulari E. Light directed massively parallel on-chip synthesis of peptide arrays with t-Boc chemistry. Proteomics 2003;3:2135-41.

[15] Pellois JP, Zhou X, Srivannavit O, Zhou T, Gulari E, Gao X. Individually addressable parallel peptide synthesis on microchips. Nat Biotechnol 2002;20:922-6.

[16] Pellois JP, Wang W, Gao X. Peptide synthesis based on t-Boc chemistry and solution photogenerated acids. Journal of combinatorial chemistry 2000;2:355-60.

[17] Stewart JM, Young,J.D. Solid Phase Peptide Synthesis. 2nd edition ed: Pierce Chemical Co; 1984.

[18] Kyte J, Doolittle RF. A Simple Method for Displaying the Hydropathic Character of a Protein. J Mol Biol 1982;157:105-32.

[19] Greving MP. Creating high-affinity ligands on surfaces and in solution: Arizona State University; 2009.

[20] Chang YP, Mahadeva R, Chang WSW, Lin SC, Chu YH. Small-molecule peptides inhibit Z alpha(1)-antitrypsin polymerization. J Cell Mol Med 2009;13:2304-16.

[21] Chang YP, Mahadeva R, Chang WSW, Shukla A, Dafforn TR, Chu YH. Identification of a 4-mer peptide inhibitor that effectively blocks the polymerization of pathogenic Z alpha(1)-antitrypsin. Am J Resp Cell Mol 2006;35:540-8.

[22] Crystal RG. The alpha 1-antitrypsin gene and its deficiency states. Trends in genetics : TIG 1989;5:411-7.

[23] Elliott PR, Pei XY, Dafforn TR, Lomas DA. Topography of a 2.0 angstrom structure of alpha(1)-antitrypsin reveals targets for rational drug design to prevent conformational disease. Protein Sci 2000;9:1274-81.

[24] Diehnelt CW, Shah M, Gupta N, Belcher PE, Greving MP, Stafford P, et al. Discovery of High-Affinity Protein Binding Ligands - Backwards. Plos One 2010;5.

[25] Levy ED, De S, Teichmann SA. Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. Proc Natl Acad Sci U S A 2012;109:20461-6.

[26] Yang JR, Liao BY, Zhuang SM, Zhang J. Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. Proc Natl Acad Sci U S A 2012;109:E831-40.

Chapter 4

[1] Andresen H, Grotzinger C. Deciphering the Antibodyome - Peptide Arrays for Serum Antibody Biomarker Diagnostics. Curr Proteomics 2009;6:1-12.

[2] Sykes KF, Legutki JB, Stafford P. Immunosignaturing: a critical review. Trends Biotechnol 2013;31:45-51.

[3] Legutki JB, Johnston SA. Immunosignatures can predict vaccine efficacy. P Natl Acad Sci USA 2013;110:18614-9.

[4] Restrepo L, Stafford P, Magee DM, Johnston SA. Application of immunosignatures to the assessment of Alzheimer's disease. Annals of neurology 2011;70:286-95.

[5] Hughes AK, Cichacz Z, Scheck A, Coons SW, Johnston SA, Stafford P. Immunosignaturing can detect products from molecular markers in brain cancer. Plos One 2012;7:e40201.

[6] Stafford P, Halperin R, Legutki JB, Magee DM, Galgiani J, Johnston SA. Physical characterization of the "immunosignaturing effect". Mol Cell Proteomics 2012;11:M111 011593.

[7] Brown JR, Stafford P, Johnston SA, Dinu V. Statistical methods for analyzing immunosignatures. Bmc Bioinformatics 2011;12:349.

[8] Legutki JB, Zhao ZG, Greving JP, Woodbury NW, Johnston SA, Stafford P. Scalable High-Density Peptide Arrays for Comprehensive Health Monitoring Nature Communications 2014. (In perss)

[9] Stafford P, Cichacz Z, Woodbury NW, Johnston SA. An Immunosignature system for diagnosis of cancer. Proc Natl Acad Sci U S A 2014. (Under Review)

[10] James LC, Tawfik DS. The specificity of cross-reactivity: promiscuous antibody binding involves specific hydrogen bonds rather than nonspecific hydrophobic stickiness. Protein Sci 2003;12:2183-93.

[11] Halperin RF, Stafford P, Johnston SA. Exploring antibody recognition of sequence space through random-sequence peptide microarrays. Mol Cell Proteomics 2011;10:M110 000786.

[12] Reverberi R, Reverberi L. Factors affecting the antigen-antibody reaction. Blood transfusion = Trasfusione del sangue 2007;5:227-40.

[13] Van Oss CJ. Hydrophobic, hydrophilic and other interactions in epitope-paratope binding. Molecular immunology 1995;32:199-211.

[14] Monera OD, Sereda TJ, Zhou NE, Kay CM, Hodges RS. Relationship of Sidechain Hydrophobicity and alpha-Helical Propensity on the Stability of the Single-stranded Amphipathic alpha-Helix. J Pept Sci 1995;1:319-29.

[15] Wang W, Woodbury NW. Selective protein-peptide interactions at surfaces. Acta Biomater 2014;10:761-8.

[16] Soga S, Kuroda D, Shirai H, Kobori M, Hirayama N. Use of amino acid composition to predict epitope residues of individual antibodies. Protein Engineering Design & Selection 2010;23:441-8.

[17] Soga S, Shirai H, Kobori M, Hirayama N. Use of amino acid composition to predict ligand-binding sites. Journal of chemical information and modeling 2007;47:400-6.