A Molecular Simulation Study of Antibody-Antigen Interactions on Surfaces for the Rational Design of Next-Generation Antibody Microarrays

Derek B. Bush
Brigham Young University
A Molecular Simulation Study of Antibody-Antigen Interactions on Surfaces
for the Rational Design of Next-Generation Antibody Microarrays

Derek B. Bush

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Thomas A. Knotts IV, Chair
Bradley C. Bundy
Dean R. Wheeler
John D. Hedengren
Morris D. Argyle

Department of Chemical Engineering
Brigham Young University

Copyright © 2017 Derek B. Bush
All Rights Reserved
ABSTRACT

A Molecular Simulation Study of Antibody-Antigen Interactions on Surfaces for the Rational Design of Next-Generation Antibody Microarrays

Derek B. Bush
Department of Chemical Engineering, BYU
Doctor of Philosophy

Antibody microarrays constitute a next-generation sensing platform that has the potential to revolutionize the way that molecular detection is conducted in many scientific fields. Unfortunately, current technologies have not found mainstream use because of reliability problems that undermine trust in their results. Although several factors are involved, it is believed that undesirable protein interactions with the array surface are a fundamental source of problems where little detail about the molecular-level biophysics are known. A better understanding of antibody stability and antibody-antigen binding on the array surface is needed to improve microarray technology. Despite the availability of many laboratory methods for studying protein stability and binding, these methods either do not work when the protein is attached to a surface or they do not provide the atomistic structural information that is needed to better understand protein behavior on the surface. As a result, molecular simulation has emerged as the primary method for studying proteins on surfaces because it can provide metrics and views of atomistic structures and molecular motion. Using an advanced, coarse-grain, protein-surface model this study investigated how antibodies react to and function on different types of surfaces. Three topics were addressed: (1) the stability of individual antibodies on surfaces, (2) antibody binding to small antigens while on a surface, and (3) antibody binding to large antigens while on a surface. The results indicate that immobilizing antibodies or antibody fragments in an upright orientation on a hydrophilic surface can provide the molecules with thermal stability similar to their native aqueous stability, enhance antigen binding strength, and minimize the entropic cost of binding. Furthermore, the results indicate that it is more difficult for large antigens to approach the surface than small antigens, that multiple binding sites can aid antigen binding, and that antigen flexibility simultaneously helps and hinders the binding process as it approaches the surface. The results provide hope that next-generation microarrays and other devices decorated with proteins can be improved through rational design.

Keywords: antibody, antigen, coarse-grain, ligand binding, molecular simulation, microarray, protein stability, umbrella sampling, replica exchange, lysozyme, hemagglutinin
I foremost wish to thank my graduate advisor, Dr. Tommy Knotts. He has been instrumental in educating me on the fundamentals of molecular simulation, analyzing simulation results, and writing papers for publication. He has always been kind and patient, and his advice has helped me to understand how to do scientific research in a proper and clear way.

I also wish to thank my undergraduate research advisor, Dr. Brad Bundy. He provided me the first opportunity to work in scientific research shortly after becoming a faculty member at BYU. The time I spent working in his lab taught me many of the things that helped me be successful later on during my graduate program. His excitement for scientific discovery was contagious and instrumental in directing my attention to graduate school.

I wish to note the friendship and help from various undergraduate and graduate students I have worked with. These include Ryan Welling, Anthony Gillespie, Tyler Madsen, Jacob Marshall, Josh Wilkerson, Shuai Wei, Rebecca Grawe, Addison Smith, and Daniel Carlson. I wish to particularly thank Jonathan Gallacher, Jacob Williams, and Richard Messerly for their friendship.

I appreciate the love and support I have received from my parents during both my undergraduate and graduate years. I am grateful for the supportive Church leaders that I have interacted with during my time at BYU. I am also grateful for BYU itself, for the ideals that it stands for, the education it has provided me, and for opportunities to conduct scientific research. I acknowledge the many financial contributions that Church members make to sustain BYU. I also wish to thank my Father in Heaven for the opportunity to study at BYU, and for His guidance that brought me to BYU.

I acknowledge the financial support of the National Science Foundation (No. CBET-1054867). I am also grateful for the numerous resources provided and maintained by the Fulton Supercomputing Facility at BYU. The work in this dissertation ultimately involved over 904,000 jobs, consuming more than 11.8 million CPU hours.
# TABLE OF CONTENTS

**LIST OF TABLES** .............................................. viii

**LIST OF FIGURES** ........................................... ix

Chapter 1  Introduction ........................................ 1

Chapter 2  Literature Review ................................. 4
  2.1  Basic antibody concepts .............................. 4
    2.1.1  Value of proteins ............................ 4
    2.1.2  Structure-function relationship ............. 4
    2.1.3  Protein folding ................................ 5
    2.1.4  Physical characteristics of antibodies .... 7
    2.1.5  Antibody crystal structures ................. 7
    2.1.6  Summary of protein biochemistry .......... 9
  2.2  Current antibody microarrays ....................... 9
    2.2.1  Designs ....................................... 9
    2.2.2  Substrate materials ......................... 12
    2.2.3  Immobilization techniques ................... 13
    2.2.4  Summary of microarray designs .............. 15
  2.3  Microarray performance issues ...................... 15
    2.3.1  Excessive variability ........................ 15
    2.3.2  Underperformance of immobilized antibodies 16
    2.3.3  Unbalanced antibody performance ............. 17
    2.3.4  Cross-reactivity ................................ 18
    2.3.5  Summary of performance issues .............. 19
  2.4  Laboratory methods for studying proteins on surfaces 19
    2.4.1  Methods ....................................... 19
    2.4.2  Resolution limitations ....................... 20
    2.4.3  Summary of laboratory methods .............. 21
  2.5  Molecular simulation of proteins .................... 21
    2.5.1  Methods ....................................... 21
    2.5.2  All-atom models ................................ 22
    2.5.3  Coarse-grain models .......................... 23
    2.5.4  Summary of molecular simulation .......... 24
  2.6  Antibody behavior on surfaces ....................... 25
    2.6.1  Antibody stability ........................... 25
    2.6.2  In silico antibody-surface studies .......... 26
    2.6.3  In silico studies of small proteins on surfaces 27
    2.6.4  Summary of antibody stability on surfaces .... 29
  2.7  Antibody interactions with small antigens .......... 30
    2.7.1  Interaction forces ............................ 30
    2.7.2  Interactions in bulk solution ............... 31
7.1 Review of findings ................................................. 117
7.2 Future work ....................................................... 119
  7.2.1 Different affinity molecules .............................. 119
  7.2.2 Simulating antibody and fab arrays .................. 121
  7.2.3 Simulating other protein complexes .................. 122
7.3 Conclusions ..................................................... 123

REFERENCES .......................................................... 124
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Height of barriers to binding in kJ mol(^{-1}) for the different fab-lysozyme treatments. AVG represents the average barrier height while ERR represents the standard error.</td>
<td>75</td>
</tr>
<tr>
<td>5.2</td>
<td>Binding strength in kJ mol(^{-1}) for the different fab-lysozyme treatments. AVG represents the average binding strength while ERR represents the standard error.</td>
<td>76</td>
</tr>
<tr>
<td>5.3</td>
<td>Change in entropy upon binding ((\Delta S_{binding})) in kJ mol(^{-1}) for three fab-lysozyme systems.</td>
<td>79</td>
</tr>
<tr>
<td>6.1</td>
<td>Sites used to calculate (\xi_1) and (\xi_2) in bulk simulations.</td>
<td>88</td>
</tr>
<tr>
<td>6.2</td>
<td>RMSD values for the HA1 and HA2 domains in HA.</td>
<td>93</td>
</tr>
<tr>
<td>6.3</td>
<td>Excluded volumes of the HA1-1, HA1-2, and HA1-3 domains in Å(^3).</td>
<td>95</td>
</tr>
<tr>
<td>6.4</td>
<td>Binding energies between the fab and each HA1 domain for each bulk case where HA is rigid.</td>
<td>100</td>
</tr>
<tr>
<td>6.5</td>
<td>Binding energies between the fab and each HA1 domain for each 2D bulk case where HA is flexible.</td>
<td>101</td>
</tr>
<tr>
<td>6.6</td>
<td>Summary of binding strength as a function of HA angle relative to the surface and where HA is rigid.</td>
<td>102</td>
</tr>
<tr>
<td>6.7</td>
<td>Summary of binding strength as a function of HA angle relative to the surface where HA is flexible.</td>
<td>108</td>
</tr>
<tr>
<td>6.8</td>
<td>The percentage of simulation trajectory frames where flexible HA was found in an open configuration near the surface.</td>
<td>108</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

2.1 Examples of protein structural components. Panel (a) is the amino acid sequence, Panel (b) shows common secondary structure motifs, and Panel (c) shows the tertiary structure. The protein shown in Panel (c) is lysozyme. [1] ........................ 6

2.2 Structure of a typical IgG antibody. Panel (a) highlights the three primary domains and along with other structural elements, and Panel (b) shows the subdomains that comprise each primary domain. The structure comes from PDB ID 1IGT. [2] ... 8

2.3 Crystal structures of three full-length antibodies in the PDB. Corresponding peptide chains have the same color. Panel (a) is 1IGT, Panel (b) is 1IGY, [3] and Panel (c) is 1HZH. [4] ................................................. 8

2.4 Layout and use of a planar antibody microarray. Panel (a) shows the initial manufacturing process, Panel (b) shows the microarray after the sample is added, and Panel (c) shows the microarray after washing and labeling steps. ................. 11

2.5 Size comparison of (a) a viral capsid, [5] (b) an antibody, (c) a fab, (d) lysozyme, and (e) hemagglutinin. [6] ................................................................. 38

4.1 Radius of gyration (Å) of the antibody at different temperatures in bulk solution as a function of simulated time (ns) for a typical simulation. Here, blue represents snapshot 1 (260 K), red represents snapshot 2 (308 K), yellow represents snapshot 3 (340 K), and purple represents snapshot 4 (390 K). The grey area highlights the equilibration portion of the simulation. ................................. 48

4.2 Example of how the fraction native curves for the two V_L domains do not exactly match. ............................................................... 50

4.3 Relative heat capacity as a function of temperature for the antibody in (A) bulk solution, (B) on a hydrophobic surface, and (C) on a hydrophilic surface. For panels (B) and (C) solid lines represent tethering in a flat orientation and dashed lines represent tethering in an upright orientation. Roman numerals correspond to antibody configurations shown in Figure 4.4. ................................. 51

4.4 Representative configurational snapshots of the antibody on different surface types. 52

4.5 Melting temperatures of the ABS under different conditions. Error bars represent standard errors from six simulation replicates. ................................. 55

4.6 Depictions of the fab for the two tethering configurations investigated in this study. [A] The flat tethering configuration, and [B] the upright tethering configuration. ... 62

5.2 The reaction coordinates for the umbrella simulations. ............................... 63

5.3 Fab-lysozyme PMF for bulk solution. The black line represents the average PMF, the gray region represents the standard error, and the dotted line divides the vertical axis into positive and negative regions. The labels correspond to the configurational snapshots. Images (A), (B), and (C) show representative snapshots of the proteins’ behavior at short, intermediate, and long separation distances, respectively. ................................. 66

5.4 Panels (A) and (B) show the PMFs for the fab-lysozyme system with the fab in the upright and flat configurations, respectively, on a purely-repulsive surface. Panels (C) and (D) show configurational snapshots and binding mechanisms for the upright and flat cases, respectively. ................................. 69
5.5 Panels (A) and (B) show the PMFs for the fab-lysozyme system with the fab in the upright and flat configurations, respectively, on a hydrophobic surface. Panels (C) and (D) show configurational snapshots and binding mechanisms for the upright and flat cases, respectively. ................................................................. 72

5.6 Panels (A) and (B) show the PMFs for the fab-lysozyme system with the fab in the upright and flat configurations, respectively, on a hydrophilic surface. Panels (C) and (D) show configurational snapshots and binding mechanisms for the upright and flat cases, respectively. ................................................................. 74

5.7 Theory showing how surfaces stabilize protein-protein interactions. Panel (A) represents the fab-lysozyme system in bulk solution and Panel (B) represents the system when the fab is tethered to a surface. ................................................................. 80

6.1 Structural setup of the fab and HA. Panel (a) shows the fab from two viewpoints, Panel (b) shows HA in the rigid configuration from two viewpoints, and Panel (c) shows an example configuration of flexible HA after relaxation by molecular dynamics. ................................................................. 85

6.2 Setup of $\xi_1$ and $\xi_2$ for simulations in bulk solution. Panel (a) is for Rigid/Flex #1, Panel (b) is for Rigid/Flex #2, and Panel (c) is for Rigid/Flex #3. In each Panel HA1-1 is colored red, HA1-2 is blue, and HA1-3 is green. ................................................................. 89

6.3 Setup of $\xi_1$, $\xi_2$, and HA angle for simulations involving a surface. ................................................................. 90

6.4 Bulk solution PMF for the fab-HA system for the control Rigid #1. Panel (a) depicts the reaction coordinate setup, Panel (b) is the PMF overlaid with potential binding pathways, and Panel (c) shows representative configurations along the binding pathways. ................................................................. 94

6.5 Bulk solution PMF for the fab-HA system for the control Rigid #2. Panel (a) depicts the reaction coordinate setup, Panel (b) is the PMF overlaid with potential binding pathways, and Panel (c) shows representative configurations along the binding pathways. ................................................................. 96

6.6 Bulk solution PMF for the fab-HA system for the control Rigid #3. Panel (a) depicts the reaction coordinate setup, Panel (b) is the PMF overlaid with potential binding pathways, and Panel (c) shows representative configurations along the binding pathways. ................................................................. 97

6.7 Bulk solution PMF for the fab-HA system for the treatment Flex #2. Panel (a) is the reaction coordinate setup, Panel (b) is the PMF and potential binding pathways, and Panel (c) shows representative configurations along the binding pathways. ................................................................. 99

6.8 PMFs and reaction coordinate setups for fab-rigid-HA simulations on a hydrophilic surface. Panels (a) through (e) are the PMFs and Panels (f) through (j) are the reaction coordinate setups. ................................................................. 103

6.9 Fab-HA configurations along the proposed binding pathways where the fab is on a surface and HA is rigid. ................................................................. 104

6.10 PMFs and reaction coordinates for each surface treatment where HA is flexible. Panels (a) through (e) are the PMFs and Panels (f) through (j) are the reaction coordinate setups. ................................................................. 109

6.11 Fab-HA configurations along the proposed binding pathways where the fab is on a surface and HA is flexible. ................................................................. 110
6.12 Percentage of trajectory frames where HA is in an open configuration as a function of $\xi_2$. Dots represent data from simulations, the red dashed line is a parabolic trend line, and the blue dashed line represents the percentage of trajectory frames in bulk solution that have open HA configurations.

7.1 Coarse-grain intact-antibody models that bind antigens developed for this dissertation. Panel (a) is the antibody-lysozyme model and Panel (b) is the antibody-hemagglutinin model.

7.2 Comparison of a fab, scFv, and camelid antibody. Panel (a) shows the fab as translucent with the scFv portion on top as a ribbon diagram and Panel (b) shows the structure of a camelid antibody.

7.3 Example of a coarse-grain $5 \times 5$ fab array that recognizes hemagglutinin.

7.4 Structure of Xylose Isomerase. This enzyme is composed of 4 symmetric subunits.
CHAPTER 1. INTRODUCTION

Antibody microarrays are a type of biosensor that could revolutionize how molecular detection and identification is accomplished in many scientific areas. The enzyme-linked immunosorbent assay (“ELISA”) serves as the current gold standard for molecular detection and has been used auspiciously for decades. However, the miniaturized layout of antibody microarrays enables them to have the potential to provide faster results, be cheaper to manufacture, be easier to use, require smaller sample volumes, and permit much higher degrees of parallel screening compared to ELISA arrays. Potential applications of this microarray technology could be cheap and comprehensive personalized medicine, [7–12] creating larger and more-intricate molecular profile data sets for researchers, [10, 13, 14] and serving as portable biohazard sensors for bioterrorism responses and military groups exposed to biowarfare. [11, 15, 16] Additionally, the design technologies that go into making advanced microarrays could be applied to other surfaces decorated with proteins and could enable a host of improvements in other areas [17–26] that rely on or could benefit from immobilized proteins.

Unfortunately, despite their great potential, current antibody microarrays have been unable to find mainstream use because of reliability problems. The issues include poor replication of results both within arrays and between arrays, [27–29] poor signal quality from microarray spots, [8, 9, 28, 30, 31] unbalanced antibody binding performance, [28, 32–34] and cross-reactivity between antibodies not present during the antibody discovery process. [9, 10, 34–36] Some of the issues stem from difficulties with procuring high-quality antibodies against molecules of interest, micro-spotting techniques, and surface defects. However, a major problem is from unintended interactions between the proteins and the microarray substrate that alter or destroy protein shape and function. Many studies have been directed toward developing substrate materials and immobilization strategies that are better at avoiding unwanted protein-surface interactions, leading to significant improvements in microarray function. However, because performance issues still re-
main limiting, more details about the molecular-level behavior of antibodies and their antigens near the surface are needed to lead to further improvements in microarray technologies.

A significant amount of effort has been made to delve deeper into the microscopic environment at the microarray surface. Unfortunately, typical laboratory methods for obtaining protein structures, X-ray crystallography and Nuclear Magnetic Resonance (NMR), cannot be applied to heterogeneous environments such as surface-bound proteins. [37–39] Researchers have been able to apply a variety of other laboratory methods with differing degrees of success, and such work has provided much information about how proteins adsorb to different surface types, how structural stability is affected by the new environment, and ultimately how well the proteins continue to function after attachment to the surface. However, even the highest resolution method (atomic force microscopy) is unable to resolve atomically-detailed protein structures on the surface—information that would be particularly useful for understanding relevant molecular dynamics and improving the immobilization setup. As a result, knowledge from these laboratory studies needs to be augmented by methods that can circumvent the current equipment limitations.

*In silico* methods can provide information unobtainable by laboratory means because individual proteins can be modeled in atomic, or near-atomic, detail while their molecular motions can be examined with a high degree of precision. As a result, molecular simulation has emerged as the primary tool for studying protein-surface interactions. [17, 37, 40] Advances in computing power, improved model parameterization, and the availability of efficient configurational sampling methods are making it easier to obtain accurate thermodynamic metrics of molecular-level events such as protein binding and protein folding that are fundamental to microarray function. However, despite the many examples of antibody-surface and antibody-antigen simulation-based studies in the literature, several questions related to microarray performance remain mostly, or completely, unanswered. As a result, there is a need for completing new molecular simulation work that does more to address the factors that impact microarray performance.

The goal of the work explained in this dissertation was to discover the biophysics involved in antibody stability after attachment to a surface and how the surface impacts antibody-antigen binding. Once these fundamental behaviors are understood, new design principles for antibody microarrays can be formulated which will lead to improvements in next-generation devices and move these towards mainstream use. Additionally, because the factors that affect antibody microarrays
are similar to those that affect other materials decorated with proteins, the knowledge gained from this study should also lead to improvements in technologies beyond microarrays.

The remainder of this dissertation is structured as follows. An overview of the relevant literature is presented in Chapter 2. This is followed (Chapter 3) by an overview of the objective of this work and an outline of the tasks done to accomplish the objective. The following three chapters (Chapters 4-6) describe three studies done to understand the fundamental behavior of antibodies on surfaces. Chapter 4 explains how antibody stability is affected by attachment to different types of surfaces and the orientation of it relative to the surface. Chapter 5 explains how antibody-lysozyme binding operates changes upon attachment of the antibody to a surface as well as its orientation relative to the surface. Chapter 6 explains how three characteristics of hemagglutinin (its large size, multiple binding sites, and flexibility) affect antibody binding while on the surface, and compares this behavior to the behavior of the antibody-lysozyme complex. Each of these chapters lay out a specific hypothesis that was tested, the methods used to test the hypothesis, the results of the simulations, and a discussion of the impact of the results. The final chapter (Chapter 7) summarizes the work and offers ideas on where future work lies. The work presented in this dissertation resulted in two peer-reviewed publications (Chapter 4 [41] and Chapter 5 [42]) and a third publication (Chapter 6) in progress.
CHAPTER 2. LITERATURE REVIEW

2.1 Basic antibody concepts

2.1.1 Value of proteins

Proteins are biological macromolecules that perform a myriad of functions within living organisms. Examples of these functions include serving as structural supports, catalysts, defense mechanisms, mechanical/optical/chemical sensors and signalers, and molecular transporters. Proteins often perform these tasks in ways that current man-made materials cannot match. For example, while typical industrial turnover rates are approximately 1 s\(^{-1}\), the enzyme catalase has a turnover rate upwards of \(4 \times 10^7\) s\(^{-1}\). [43] Additionally, many enzymes only bind to a specific stereoisomer or produce a specific stereoisomer without byproducts. Such abilities create exciting opportunities for enhancing or enabling various scientific, medical, and industrial technologies by combining man-made devices with naturally-occurring and engineered proteins. Such technologies will increase the accessibility to rapid personalized medicine, [7–12] enable the assembly of comprehensive data sets for proteomics and big data research, [10, 13, 14] provide soldiers and rescue workers with robust biowarfare and biohazard agent detection, [11,15,16] improve medical implant biocompatibility, [17–19] add antimicrobial properties to surfaces, [20–24] and expand opportunities for industrial enzyme catalysis. [25,26] Antibodies are a specific class of proteins tasked with attaching themselves to a specific molecule, usually foreign invaders within a host organism. As such, they have been used as sensing molecules for diagnostic arrays and therapeutic treatments [44–46] for various human conditions.

2.1.2 Structure-function relationship

The three dimensional structure of a protein is absolutely critical to its function. Figure 2.1 shows the different levels of organization within proteins. Proteins start off as polymers (called
“peptides”) of amino acids (also called “residues”), shown in Panel (a) of Figure 2.1. There are 20 common, naturally-occurring amino acids whose properties vary according to their individual chemistries. Some of these amino acids have distinct positive or negative charges, some are polar, and the rest are apolar (hydrophobic). The sequence of amino acids in a peptide chain creates the conditions needed to make the peptide fold upon itself properly and to create the final, functional shape. As the folding process begins, the first three dimensional structures to form are various “secondary structures,” which consist of frequently seen motifs (Figure 2.1 Panel (b)). Some motifs have repetitive shapes, such as the “α-helix” and “β-sheet,” while “loops” are amorphous but rigid. Groups of secondary structures in turn combine to produce the “tertiary structure” of the protein, which is the overall shape of the molecule (Figure 2.1 Panel (c)). This third level of structure can be relatively simple or composed of several groupings (called “domains”) of different tertiary structures. Beyond this, groups of individual peptides can come together, forming what is called “quaternary structure” (not shown).

*Tertiary (and sometimes quaternary) structure is crucial to protein function.* If the shape of the protein becomes distorted or destroyed the protein will lose its function. As it relates to antibodies on microarrays, the structures most important to their function are the two antigen binding sites (ABS) on each antibody.

### 2.1.3 Protein folding

The protein folding process is analogous to a chemical reaction. The number of steps involved varies from protein to protein, while the entire process typically takes \( O(10^{-6}) \) to \( O(10^3) \) s. [47] The folding process often starts spontaneously and it can be sped up greatly (upwards of \( 30\times \)) with the help of “chaperones” (which act like catalysts) that guide the process. [48] Some proteins, such as antibodies, require the help of chaperones to achieve their correct finished state. [49] Thus, commercial antibodies must be produced in eukaryotic cell systems with these chaperones, such as Chinese hamster ovary (CHO) cells. [49]

Folding is typically initiated by the association of hydrophobic residues with one another. This results in a “hydrophobic collapse” that allows the unfavorable negative change in entropy within the protein as it folds upon itself (\( \Delta S_{\text{folding}} \approx -7.5 \text{ kJ mol}^{-1} \)) to be overcome by the favorable increase in entropy of the water molecules that are freed from surrounding the hydrophobic
Figure 2.1: Examples of protein structural components. Panel (a) is the amino acid sequence, Panel (b) shows common secondary structure motifs, and Panel (c) shows the tertiary structure. The protein shown in Panel (c) is lysozyme. [1]

residues ($\Delta S_{\text{water}} \approx +45 \text{ kJ mol}^{-1}$). As a result, hydrophobic residues are often hidden at the center of proteins while hydrophilic (polar) and charged residues are left on the surface to facilitate solubility and function. The rest of the steps required to complete the folding process are usually difficult to predict from just the amino acid sequence. This results from the astronomical number of possible atomic configurations available to the peptide chain. [50]

Protein unfolding ("denaturation") is a significant issue that degrades or limits overall protein function (for enzymes, this is similar to catalyst poisoning). Proteins can denature in response to various chemical stimuli, such as extreme pH and salt concentrations, or non-aqueous solvents. Physical stimuli, such as nearby solid surfaces, fluid shearing, and excessive heat exposure can also break protein structure. These conditions interfere with the hydrophobic and electrostatic interactions that hold the protein together. Sometimes unfolded proteins refold back into their proper shape and continue to function once their environment improves. However, even partial denaturation can significantly degrade protein function and result in permanent aggregation and/or precipitation out of solution. Antibodies, the type of protein most important to this work, typically cease functioning entirely after denaturation.


2.1.4 Physical characteristics of antibodies

Figure 2.2 shows the molecular structure of a typical IgG antibody, which is the most common type of antibody. [32] Panel (a) highlights the three primary domains of IgG antibodies, the location of the antigen binding sites, examples of disulfide bonds (shown in yellow), and the location of attached sugar molecules. Panel (b) highlights the smaller subdomains that make up each primary domain. IgG antibodies are composed of four independent peptide chains that are connected by disulfide bonds. There are two identical “light” chains (approximately 25 kDa each) and two identical “heavy” chains (approximately 50 kDa each). Each light chain associates with one end of a heavy chain to form a “fragment, antigen binding” (Fab) domain while the other end of each heavy chain associates with each other to form a “fragment, crystallizable” (Fc) domain. The middle part of each heavy chain remains amorphous and forms a “hinge” region which enables the fabs to move independent of one another and the Fc domain. Each light chain and each heavy chain have a variable region (V_L and V_H, respectively) whose amino acid sequence can be modified by immune cells to enable the antibody to bind to new targets. These variable regions form the ABS of the fabs (also called the “complementarity determining region,” or CDR). Intact IgG antibodies form an equilateral triangle shape where each side is approximately 160 Å across [2] and are about 150 kDa (1 Dalton = 1 AMU) in mass. [32] In addition to the IgG type, other single-antibody molecular types are IgD and IgE. Complexes of two (IgA) or five (IgM) antibodies that are linked together are also found naturally. [43] This dissertation will only involve single IgGs or IgG fragments, as these are most commonly used in antibody microarrays.

2.1.5 Antibody crystal structures

The importance of antibodies has motivated many researchers to attempt to crystalize them to obtain their atomic structures. The significant flexibility of antibodies has, until 20 years ago, largely hampered these attempts. However, improvements in crystallization techniques have lead to the assembly to three full-length antibody structures that are available in the Protein Data Bank (PDB). [2–4] These structures are shown in Figure. 2.3 It is important to note that these antibody structures are a result of the antibodies binding to themselves, and not to a separate antigen molecule.
Figure 2.2: Structure of a typical IgG antibody. Panel (a) highlights the three primary domains and along with other structural elements, and Panel (b) shows the subdomains that comprise each primary domain. The structure comes from PDB ID 1IGT. [2]

Figure 2.3: Crystal structures of three full-length antibodies in the PDB. Corresponding peptide chains have the same color. Panel (a) is 1IGT, Panel (b) is 1IGY, [3] and Panel (c) is 1HZH. [4]

While entire antibodies have been difficult to crystallize, it is much easier to crystallize fabs bound to their antigen. In fact, fabs are sometimes used to help crystallize proteins that are difficult to crystallize by themselves. As a result, many fab-antigen complexes can be found in the PDB. While the ABS varies from one antibody to the next, the rest of the fab structure and the Fc domain are largely the same among all antibodies. This creates the possibility of assembling a chimeric antibody-antigen model, as was done for the work discussed in Chapters 5 and 6.
2.1.6 Summary of protein biochemistry

Proteins are crucial to the maintenance of life. They perform a variety of functions in ways and with precision that are difficult or impossible to match with current man-made technologies. They represent a fine balance of atomic forces that enable a peptide polymer to adopt a three dimensional shape that enables the function of the protein. If this shape is distorted or destroyed the protein will cease to function. As such, it is imperative that attaching proteins to a surface is done in a way that maintains or stabilizes the typical three dimensional shape of the protein seen in bulk solution.

Antibodies are large and complex proteins tasked with identifying and capturing specific molecules in complex biological mixtures. A few crystal structures of intact antibodies exist in the Protein Data Bank that can be used for generating models for this work. However, there are no structures of intact antibodies bound to an antigen other than themselves; instead, structures of antibody fragments bound to their antigens are more common. The availability of these structures enables the development of the chimeric protein models in this work used to study antibodies and antigens near surfaces for improving antibody microarrays.

2.2 Current antibody microarrays

2.2.1 Designs

Figure 2.4 shows an example of an antibody microarray. The array consists of a base plate onto which antibodies specific to different target molecules are deposited in unique, known locations on the array surface (Panel (a)). Once the array is finished, a sample is incubated with the array (Panel (b)), followed by specific washing steps. During these steps, unbound proteins are removed while labels are attached to the captured antigens (Panel (c)). In essence, if a spot is fluorescent at the end of the process, the location of the spot tells what antigen was found and the intensity of the light indicates the concentration of that molecule in the sample. Protein microarrays like this were originally developed in parallel with DNA microarrays to provide additional information about protein expression levels, protein-protein interactions, and biochemical activities of proteins in samples that could not be achieved by DNA microarrays alone. [33, 51]
Protein microarrays come in several different basic formats. Microarrays can be either “forward phase” in which antibodies are immobilized to the array surface, or “reverse phase” in which the target proteins are immobilized on the surface. However, antibody microarrays are almost always forward phase. Additionally, they can come in either a planar or micro-bead format. [27] Planar microarrays consist of a flat substrate onto which antibodies specific to each target molecule are immobilized to a unique grid spot on the chip surface. When used, a sample is first added to the array surface, followed by various washing steps, after which the signals are acquired. Another common setup is a micro-bead layout. These arrays consist of a solution of nano- or micron-sized polymer beads onto which antibodies have been attached. The bead solution is first mixed with the sample, then the beads are collected and washed, and finally the signals are acquired.

Each design has its own advantages and disadvantages. Planar formats are easier to manufacture through micro-spotting techniques and allow for greater multiplexing (the number of targets searched for in the assay) than micro-bead systems. Micro-bead designs allow for better sample mixing so that the antibodies have a better chance to come into contact with low-concentration target molecules. However, regardless of the design, the affinity molecules are are significantly affected by the surface properties. As such, both formats are susceptible to poor performance if the antibodies become denatured or oriented incorrectly.

Some research indicates that using only the antigen-binding portion of the antibody (the fab) as the affinity molecules for microarrays is a way to improve microarray performance. [31,52,53] Using only the fab domains of antibodies is believed to minimize antibody-surface interactions and also harness other phenomena that maximizes binding. One reason for this is that it avoids undesirable interactions between the Fc domain and the surface or target proteins. [53] Another is that it can be easier to control fab orientation relative to the surface compared to complete, intact antibodies. [52] Moreover, the smaller size of the fab, compared to the whole antibody, allows for more sensing elements per surface area, which aids signal acquisition when miniaturizing microarray spots. [31] The confluence of all of these factors, and others likely unknown, can yield large increases in antigen binding performance. For example, one study showed that a surface coated with well-oriented fabs resulted in 20× increased binding capacity compared a surface coated with intact antibodies. [54] Because fab-based microarrays generally demonstrate better performance,
The antibody microarray is manufactured by spotting each antibody onto a specific location on the array surface. A sample of interest is added to the finished microarray. Afterward, several washing and labeling steps are performed. Spots where the immobilized antibody captured an antigen produce a signal from which the identity of the antigen and its concentration in the sample can be determined.

Figure 2.4: Layout and use of a planar antibody microarray. Panel (a) shows the initial manufacturing process, Panel (b) shows the microarray after the sample is added, and Panel (c) shows the microarray after washing and labeling steps.

the antibody-antigen binding studied in Chapters 5 and 6 only involve the fab portion, and not the Fc region, of the antibody molecule.

Once the antibodies have had a chance to capture their target molecules, it is necessary to use some method to determine whether the antibodies were successful or not. Common methods include chemiluminescence, [36] fluorescence, [55] or surface plasmon resonance (SPR). [56] The chemiluminescence and fluorescence methods are based off a traditional sandwich ELISA format (see Figure 2.4). After the immobilized antibodies capture their target antigens a second layer of antibodies specific to each target molecule is added. Then, a third antibody (which binds to the second-layer antibody) labeled with an enzyme or a fluorescent molecule is added on top of the second layer of antibodies. The labels produce light whose macroscopic intensity is proportional to
the number of antigens that were captured. In contrast, the SPR method does not produce a direct signal from the antibody-antigen complexes. Instead, as the sample incubates with the array the accumulation of target molecules on each microarray spot is periodically checked by detecting how the additional mass changes the resonance frequency of the substrate surface plasmons (resonating oscillations in the substrate’s electrons). This means that SPR can detect binding events without additional antibodies and it can obtain binding as a function of time. However, it can only be used with planar arrays and has difficulty detecting small molecules because of their small masses.

Regardless of the detection strategy, the quality of the resulting data is strongly dependent on the stability and binding capacity of the antibodies on the array surface. It is important to note that issues related to producing the detection signals themselves are beyond the scope of this dissertation, as it is focused on protein stability. The next two sections focus on microarray elements more-closely related to the stability and function of the immobilized capture antibodies.

2.2.2 Substrate materials

Researchers have developed a large number of materials to act as the substrate for protein microarrays. Initially, polyvinylidene difluoride (PVDF) membranes [28] or polystyrene [57] were used because of their ability to effectively bind proteins. However, to accommodate even smaller microarray sizes, different support materials had to be developed. Specifically, glass slides often serve as the substrate for microarrays due to their rigidity and the relative ease of adding desirable chemical moieties to the surface. Two other prominent substrates for protein adhesion include gold [33,58] and metal oxides. [33,59] These materials provide very smooth surfaces, which helps prevent protein aggregation [60] and mitigate interference with antibody binding specificity. [31] On top of this base layer another molecular group is usually added that ultimately serves as the surface that the nearby proteins interact with. Often this new layer consists of an alkane chain where one end is capped with a silane (for glass) or thiol (for gold) reactive group that binds to the substrate while the other end presents a different reactive group that is visible to the proteins and modulates their interactions with the new layer. Such molecular layers are referred to as self-assembled monolayers (SAM). For antibody microarrays, protein A or protein G (which bind to the Fc region) can be immobilized either directly to the substrate or on top of the alkane group and serves to retain the capture antibodies on the surface. [54, 58] The possible combinations of
functional group type and concentration, alkane chain length, and base layer materials create a large configurational space, but the current consensus is that the ideal protein microarray setup should utilize only a single surface type. [33] Whatever that substrate is, it needs to allow high protein binding without affecting protein structure or function. [9, 28, 33] However, the diverse nature of proteins makes it difficult to design a substrate that can accommodate all of these design requirements. More information about how antibodies interact with different types of surfaces would aid substrate and SAM selection.

2.2.3 Immobilization techniques

There are three primary ways that are currently used to immobilized proteins to surfaces. These are (1) physical adsorption, (2) physical entrapment, and (3) covalent linkage. [61] Physical adsorption relies on hydrophobic and electrostatic attraction to retain proteins on the surface. It is by far the easiest method for depositing proteins onto surfaces [54] and tends to be the most common method used as a result. [31] When proteins adsorb to surfaces in this manner, they do so in random orientations. [30, 31, 62] This results in lower microarray performance and reproducibility because (1) it allows unstable parts of the proteins to interact more with the surface and (2) causes a portion of their active sites to be pointed toward the surface rather than presented to bulk solution for ligand binding. [31, 54, 57] For antibody microarrays, adsorption and orientation can be aided by first depositing a layer of protein A or protein G, which will bind to the Fc region of the antibodies. The end result is that the antibodies tend to be oriented better then they would be if adsorbed directly onto the surface.

Physical entrapment relies on some kind of physical mesh, such as cross-linked hydrogel polymers, to trap proteins near the surface. The proteins are allowed to diffuse into the gel matrix until they eventually become entangled too much to move further. The gel ensures that the proteins are surrounded by water molecules [34, 57] which helps to maintain their functional shape. Additionally, the three dimensional structure of the gel enhances protein loading on the surface by providing more space for proteins to gather than a two dimensional surface. Moreover, the entrapping polymer stands help retain the proteins in their designated location during washing steps. The obvious drawback to this method is that the entrapping polymer strands can also block lig-
Covalent linkage is more complicated than the previous immobilization strategies, but it can have significant advantages, including the possibility to control the orientation of the protein on the surface (contrary to the previous two methods described). Sometimes proteins are conjugated to the surface following regular physical adsorption by targeting a specific functional group in the protein to create chemical bonds between it and the surface. Functional groups that are typically targeted include the terminal amine of lysine, carboxyl groups in aspartic and glutamic acid, and the thiol groups of cysteines. The covalent bonds ensure that the protein remains on the surface and cannot diffuse or be washed away. However, the reaction conditions to make the bonds are typically harsh and can therefore damage protein structure and function. Moreover, these functional groups exist in multiple copies on the surface of the protein. This means that the locations where bonds form are random to each deposited protein and multiple protein-surface bonds are possible.

Control over protein orientation during immobilization has been achieved using site-specific covalent linkage. It requires inserting an unnatural amino acid (UAA) containing a unique chemical group (i.e. one not normally found in biology) into the protein sequence. This substitution can have deleterious effects on protein stability and currently requires significant trial and error to find a location that is both surface-accessible and that minimizes negative impacts on the protein. After the modified protein is produced, it can be conjugated with a complementary chemical moiety located on a prepared substrate so that there is a single bond attaching the protein to the surface. This process has shown promise to improve protein stability and function compared to other immobilization methods.

While UAAs could be incorporated into antibodies, their large size and production requirements may make the incorporation process impractical. However, targeting the sugar residues in the Fc domain for chemical attachment provides another method with some orientation control and may be a more viable option than UAA incorporation. It is important to note that this dissertation focuses on the overall effects of the surface and protein orientation on antibody stability and function, rather than on the effects of different immobilization strategies.
2.2.4 Summary of microarray designs

Antibody microarray formats include many different designs. Some designs are easier to manufacture while others provide better mixing with a sample. In all cases, protein-surface interactions are of paramount importance. When depositing proteins on the surface several methods exist providing different design trade-offs. The attachment method must have as little negative impact on the protein’s stability and structure as possible, to ensure that the deposited molecules will perform their intended jobs.

Researchers have attempted to optimize all of the aforementioned design variables. However, they have yet to agree upon a substrate material and immobilization method combination that is best for microarray applications. [9, 35] This disagreement arises from the fact that a variety of designs have been tested, but poor antibody binding performance continues to plague microarray technologies.

2.3 Microarray performance issues

2.3.1 Excessive variability

One of the issues with current microarray technologies is high variability between replicate spots on the same microarray or between replicate tests done with copies of the same array. High variability is problematic because it casts doubt on microarray results and obscures the trends in the data that the users are looking for. This issue can result from manufacturing defects [8,9,12,27,29] which cause microarray spots to be smeared or faint even if the antibodies are functioning properly. This is a real concern, but prior experience has shown that defects are usually not the primary problem.

It is often assumed that all of the proteins to be deposited have similar and reproducible interactions with the surface. [12] While this is obviously untrue for a mixture of proteins, it does not even hold true for antibodies, despite the homogeneity of their overall structures and amino acid sequences. [33, 57] One study by Ellington et al. highlighted an example where the majority of duplicate regions (where each region contained multiple replicate spots) on a microarray showed at least 30% variability. [27] Angenendt et al. tested 13 different substrate materials and saw intra-substrate variabilities between 11 and 37%. [28] Another study by Angenendt et al. saw variabili-
ties has high as 43% with epoxy-based substrates. [29] A study searching for Alzheimer’s disease biomarkers found that its initial results could not be replicated on additional microarrays. [10] This variability arises despite efforts to use the exact same manufacturing procedures for replicate arrays. [29]

Such results show that poor reproducibility is a major issue that prevents microarrays from finding mainstream applications. It also shows that no single substrate material (currently in existence) alleviates this problem. Existing manufacturing methods can be fined-tuned to avoid physical defects with the microarray spots, but more fundamental understanding is needed to assess how the antibodies react to different kinds of surface materials.

### 2.3.2 Underperformance of immobilized antibodies

All of the methods for depositing proteins on surfaces result in different levels of macroscopic performance. Underperforming antibodies can cause poor signal-to-noise ratios in the microarray results. Antibodies that have been passively adsorbed to a surface show as little as 10% activity. [30] The cause of such low performance remains elusive since some antibodies function well for normal ELISAs (where they are also deposited on a surface) but do not display activity under smaller-scale microarray conditions. [28, 35]

Many attempts have been made to aid antibody binding by properly orienting the affinity molecules on the array surface. Experimental [58] and simulation [62] studies have shown that properly oriented, intact antibodies are expected to be more efficient at binding than those immobilized randomly. Proper orientation is important for keeping the active sites of proteins available to bulk solution, [38] but doing so does not necessarily mitigate aberrant protein-surface interactions. Even if structural distortions are minimal, they can significantly degrade antibody performance. [9] This issue can be dealt with by adding an excess of antibodies to each microarray spot, [8] although this method is limited by the saturation limits of the antibodies in the spotting solution, acceptable spot sizes, and spot quality.

Loss of antibody activity is a particular problem as devices become smaller in size because the number of sensing elements per spot decreases as well. [31] Inadequate activity impacts the sensitivity of microarrays, which, depending on the intended application, can create a significant barrier to mainstream use. For example, cytokines (proteins that modulate inflammation in the
body) can have concentrations below 3 pg mL\(^{-1}\) in healthy humans. [65] However, some microarrays are only able to detect antigen levels 100 to 1000\(\times\) higher than this. [58, 66] Without a better understanding of the phenomena affecting antibody function on microarray surfaces, little can be done to rationally design microarrays except for \textit{a posteriori} testing to determine how well each antibody responds to immobilization to a given surface in the presence of other antibodies intended to be used on the same microarray. [8]

2.3.3 **Unbalanced antibody performance**

Even when antibodies function adequately on a given surface they often demonstrate dissimilar levels of performance. Unbalanced antibody binding makes microarray calibration difficult and complicates interpretation of the results. Because antibodies as a group are considered to be very homogeneous in structure, it would be expected that different antibodies would react similarly to the same surface. However, studies have shown that differing behavior among antibodies is not unexpected [32] and that there is a need for different environments for different antibodies to remain stable. [28, 33] This is corroborated by recent simulation work that found antibodies often interact with a polystyrene surface through its ABSs. [67] Since the ABS is designed to be highly variable between antibodies, such interactions could be a reason different antibodies behave differently on the same surface. More recent work involving fabs also found that, despite fabs having similar structures, sites near the ABS of different fabs displayed differing sensitivity to the UV light that was used to immobilize the fragments to the surface. [31]

Often these differing behaviors show up as differences in levels of detection (LOD). The results of Angenendt \textit{et al.} showed that the LOD varied upwards of 42\% between different microarray coatings. [28] Hucknall \textit{et al.} described their success using a poly(oligo(ethylene glycol) methacrylate) polymer brush to entrap antibodies onto a surface to achieve a LOD of 100 fg mL\(^{-1}\) for several interleukins (another type of inflammation regulator). [34] However, for one of these targets, osteoprotegerin, the smallest LOD was 1 pg mL\(^{-1}\), 10\(\times\) higher than the LOD of the other molecules, demonstrating that their method does not necessarily transfer to all antibodies equally. Another study found that non-covalent immobilization methods can perform as well as covalent ones, but the results showed that the LODs varied upwards of 10,000\(\times\) depending on the surface type used. [66]
Researchers have noted that discrepancies in LOD readings for different antibodies sometimes arise from differences in their affinities for their target molecules. [34] However, during microarray setup care is usually taken to account for these discrepancies. Additionally, only antibodies with high affinities for their targets are suitable for microarray use because of the expectation that the antibody-antigen bond must survive several washing steps before the microarray is read. Otherwise, there is a significant threat of promiscuous binding if weaker-affinity antibodies are used.

2.3.4 Cross-reactivity

Although DNA microarrays have progressed from searching for hundreds to millions of targets simultaneously, antibody microarrays typically remain in the tens of targets at most. [10] This poor scale up in multiplexing is a result of significant cross-reactivity between the immobilized antibodies. Cross-reactivity occurs when an antibody promiscuously binds to antigens other than its intended target. It is believed that this is perhaps the largest barrier to producing high-performing microarrays. [9, 10] A study by Juncker et al. noted that upwards of 95% of more than 11,000 quality-controlled antibodies developed against human proteins for diagnostic purposes displayed cross-reactivity. [10]

Other studies highlight the difficulty researchers face finding high affinity antibodies in the first place. This is especially true for difficult targets such as phosphorylated sites on proteins (difficult because of the small sizes of the sites). [68] It has been suggested that using polyclonal antibodies (a mixture of antibodies that recognize different epitopes on the same antigen) may improve cross reactivity problems, [69] but this broad recognition strategy may ultimately contribute more to exasperating the problem than fixing it. [57] Usually, antibodies that are certified for use with standard ELISAs are transferred to microarrays. [35] However, that same study noted that just because an antibody pair works well for a standard ELISA (where each well targets a single antigen) does not mean that they will function similarly well in a more complex environment, such as on a multiplex microarray.

Moreover, microarrays should perform well even with “dirty” samples such as whole blood, but this is still not the case. For example, the polymer brush method devised by Hucknall suffered a $3 \times$ increase in the smallest LOD when they tested whole blood instead of purified solutions. [34]
Ultimately, there may always be decreased performance when working with complex samples simply because more protein combinations are possible. [36] However, work needs to be done to ensure that the capture antibodies on the microarray surface are functioning optimally so as to mitigate cross-reactivity.

2.3.5 Summary of performance issues

The examples cited above show that antibody microarrays have significant performance issues to overcome before reaching their theoretical potential. [66] As such, there is sufficient room for making improvements. Some issues can be mitigated by developing better micro-spotting techniques to ensure that all antibody spots are uniform. Larger changes, such as modifications to the substrate materials will also be needed. Until these challenges are overcome, microarrays will likely remain either highly simplified [10, 27, 35] or unable to find mainstream use. These challenges provide the motivation to probe the biophysics occurring at the antibody-antigen level on the surface to help rationally guide microarray design changes.

2.4 Laboratory methods for studying proteins on surfaces

2.4.1 Methods

Because a more complete picture of how surface attachment affects protein structure and stability is needed to rationally design next-generation microarray devices, much experimental work has been done to understand protein behavior on surfaces. Researchers typically seek atomic-level details of proteins to best understand their structure and function. Unfortunately, typical laboratory techniques used to obtain atomistic structural details of proteins, X-ray crystallography and nuclear magnetic resonance (NMR), cannot be applied to the heterogeneous environments created by the surface. Other laboratory methods amenable to surfaces have found some success, but their resolution is low. Surface plasmon resonance (SPR) [70–72] and quartz crystal microbalance (QCM) [73, 74] have been used to determine protein binding activity on the surface because of their ability to discriminate when additional mass builds up on the substrate. This can be done as a function of time, allowing researchers to watch the antibody-antigen binding process occur
in real time. Circular dichroism (CD) [73, 75] can tell researchers the relative amounts of different secondary structures that are present, allowing an indirect comparison between known protein structural details and how those structures change in the presence of a the surface. Fourier transform infrared spectroscopy (FTIR) [18, 76, 77] can provide protein adsorption rates and orientation of proteins on surfaces by detecting changes in secondary structure (similar to CD). X-ray photoelectron spectroscopy (XPS) [52, 59, 70, 71, 78] can determine how the elemental composition and atomic bonding observed on the surface changes as proteins adsorb to the surface. Polarized neutron reflectometry (PNR) [77, 79] can be used to determine the thickness and density of protein layers adsorbed to surfaces. Time-of-flight secondary ion mass spectroscopy (TOF-SIMS) [80–87] can also determine how much protein has adsorbed to a surface as well as gross estimates of protein orientation. However, TOF-SIMS needs to be coupled with other techniques to verify the results. Atomic force microscopy (AFM) [62, 71, 76] can probe protein binding strength, [88–97] protein orientation and rough protein structural details, [62] and surface morphology. [70] Other techniques can be found in the literature, but these methods are most commonly used. When combined together these techniques provide different viewpoints of general protein behavior after immobilization (e.g. whether or not structural changes happen or if they remain functional).

2.4.2 Resolution limitations

The main problem with the aforementioned laboratory methods is that they cannot give researchers the atomic-level details that are needed to better understand why protein function degrades on a surface. In absence of this information, researchers are left to surmise the finer details only based on their knowledge of bulk solution protein structures. Except for AFM, all of the laboratory methods only quantify average metrics of protein structures or behavior on the surface. As such, the numerous molecular configurations that contribute to each measurement end up blending together, causing their unique attributes to remain invisible. Even AFM, which can see molecules at near-atomic resolution and was heralded as a game-changing technology for studying proteins on surfaces, [39] lacks the ability to discriminate fine details in protein structure (e.g. it can discern that antibodies have three primary domains, but not more than that). Moreover, none of these techniques are capable of capturing the molecular motions of individual immobilized proteins. As
a result, molecular simulation has emerged as the primary method for studying protein behavior and structure on surfaces. [37–39]

2.4.3 Summary of laboratory methods

Laboratory methods have been, and will continue to be, necessary to quantifying and understanding protein behavior on surfaces. However, current available tools are limited in the atomic-level details that they can provide researchers. However, laboratory methods can be coupled with computational tools to provide a better overall picture of protein stability and function on surfaces.

2.5 Molecular simulation of proteins

2.5.1 Methods

The increased availability of powerful, high-speed computers has allowed molecular simulation techniques to become an important toolset to complement laboratory-based methods of studying proteins. Supercomputers are now sufficiently powerful to enable researchers to study molecular systems with a million or more atoms. [98, 99] Moreover, a variety of robust simulation software packages, parameter sets, and analysis and visualization tools are available. Molecular dynamics is a frequently-used simulation method that uses Newton’s laws of motion to calculate particle positions as a function applied forces over short, discrete steps in time. This method inherently seeks the lowest energy configuration of the system, but simulated systems can become trapped in meta-stable states. This entrapment can create problems with configurational sampling, which in turn can bias the simulation results.

Researchers have formulated different ways to enhance the configurational sampling of the basic molecular dynamics method. For example, the work in Chapter 4 involved computing the constant volume heat capacity (C_v) of the antibody as a function of temperature under different conditions. These kinds of calculations can be challenging because they require that the protein go through its folding process many times to obtain accurate energetic metrics of the behavior. The replica exchange method [100] enhances the configurational sampling by simultaneously simulating replicas of the protein at different temperatures and periodically swapping the atomic coordi-
nates. As the simulation progresses this swapping ensures that folded copies of the protein have the opportunity to unfold while unfolded copies have the opportunity to re-fold into their original three dimensional structures. As a result, the protein folds and unfolds many times, providing enough data to accurately calculate the heat capacity as a function of temperature using statistical analysis programs such as pyMBAR. [101]

The work in Chapters 5 and 6 required that the fab and antigen go through the binding process multiple times. These kinds of calculations are also difficult because they require that the two proteins go through the process of aligning with one another and moving towards and away from each many times to accurately record the binding energetics. The sampling of this process can be enhanced by using a harmonic restraint to hold the two proteins at a specific distance from one another, and repeating this process over a range of separation distances. This modified molecular dynamics method is generally referred to as “umbrella sampling.” [102–104] During each simulation a histogram keeps track of the distances between the two proteins. Once all of the simulations are finished, the histograms are analyzed together using a method such as the Weighted Histogram Analysis Method (WHAM), [103] which accounts for the effects of the restraint in each simulation and determines the potential of mean force (PMF) between the two proteins without the effects of the original restraints.

### 2.5.2 All-atom models

Protein models with varying levels of complexity are currently available. Many studies of protein behavior have involved all-atom models of proteins where atomic interactions are governed by a set of simplified energetic potentials (i.e. not electronic wave functions) such as harmonic potentials (for bonds and angles), torsion potentials (for rotations around bonds of four consecutive atoms), and Lennard-Jones and Coulombic potentials (for non-bonded interactions). In some cases, the proteins are simulated along with a bath of distinct water molecules (an “explicit solvent”). Other studies have employed an “implicit solvent” potential that reproduces the average interactions between the solvent and the protein. These equations can incorporate van der Waals and electrostatic effects, [105] as well as Brownian motion. [106]

Regardless of the solvent model that is used, all-atom models are typically expected to capture all relevant protein behavior due to their high level of detail. Sometimes these details
are particularly important, such as during drug design, because individual atoms often play critical roles in protein-ligand binding. [107, 108] Many studies using all-atom models of proteins on surfaces are found in the literature and have provided insights into how proteins adsorb to surfaces, [17, 19, 71, 109–116] how they orient themselves on surfaces, [19, 109–111, 115, 116] and how their tertiary and secondary structures change on surfaces. [19, 109, 111, 117] In many cases, all-atom simulation studies lay the initial groundwork for studying proteins under different conditions.

However, while all-atom simulations can overcome experimental resolution limitations, they have limitations of their own. In particular, they have difficulties sampling highly-complex protein behavior such as protein folding and unfolding or ligand binding and unbinding. This is because typical simulation time steps are often 0.5-1.0 fs long and therefore only small atomic movements occur during each simulation step. However, complex and large-scale movements (i.e. protein folding and ligand binding), in contrast, can occur on timescales of milliseconds to seconds. Thus, current computers simply cannot perform the calculations fast enough to capture (if they capture it at all) the relevant behavior enough times to accurately calculate useful thermodynamic metrics. [37,118] Compounding this problem is the fact that the number of water molecules needed to fully encompass the protein increases at a cubic rate with increasing protein size. Eventually this means that virtually all of the simulation time is spent moving water molecules around. Additionally, there is no guarantee that all of the atomic interaction parameters (particularly for nonbonded interactions) used in such simulations are accurate or work under all circumstances. [17, 40, 119] These issues have lead researchers to develop new types of protein models that can better replicate the desired behaviors for study.

### 2.5.3 Coarse-grain models

The limitations inherent with all-atom models have motivated researchers to formulate even simpler “coarse-grain” (CG) protein models designed to better capture specific aspects of protein behavior. The reduced complexity of CG models can help them circumvent several of the aforementioned problems with all-atom simulations. Foremost, CG models can capture folding/unfolding and ligand binding events easily and fast enough to see them multiple times within a single simulation. [37–39] It is believed that the simplified interactions between particles (both
because there are fewer particles and because the interaction potentials themselves tend to be simpler) smooths the energy landscape available to the CG protein. [120] The result is that CG proteins can more-easily explore their configurational phase space. Second, their simplicity leaves fewer variables to parameterize, which can make it easier to develop the proper parameters to accurately replicate a desired behavior. As a result, a variety of CG protein models have been developed in an attempt to find the optimal balance between structural detail and simulated behavior.

CG models do have important limitations. Besides lacking atomistic details, they usually lack explicit solvent particles, which can eliminate the entropic effects of the solvent on protein structures. [105, 120] As with any other computer model, the CG parameters are only valid under circumstances similar to those used to generate the parameters. For example, this dissertation involves the CG “Gō-like” protein model of Karanicolas and Brooks [120, 121] which is based off of data at typical biological conditions (bulk solution, 150 mM salt, pH of 7). While this particular model has demonstrated good agreement with experimental protein folding studies at these conditions, a new set of parameters would be needed to obtain accurate protein dynamics at non-standard conditions. However, since microarrays almost always operate at typical biological conditions, new parameter sets (based on a different salt concentration or pH) were not needed. Additionally, an advanced surface model has been developed to work with the protein model of Karanicolas and Brooks which is suitable for studying the stability of a variety of proteins on different surface types. [39]

2.5.4 Summary of molecular simulation

The limitations in experimental methods have led to the use of molecular simulation techniques to study proteins on surfaces. These computer tools allow proteins to be modeled at various levels of detail on any surface imaginable. While all-atom models were expected to capture all of the relevant behavior, their interaction parameters are imperfect and their high level of detail precludes them from properly sampling large-scale molecular movements. Coarse-grain models can overcome some of these limitations because they reduce the complexity of generating viable parameter sets, simplify the energy landscape available to the simulated molecules, and ultimately require fewer calculations to move particles around. Importantly, coarse-grain models of proteins
2.6 Antibody behavior on surfaces

2.6.1 Antibody stability

The \( V_L \) and \( V_H \) domains must not denature if an antibody is to bind to its intended target. Antibodies have a higher thermal stability than most proteins and typically do not denature completely until they reach 70-100°C. [32,49] This high denaturation point indicates that antibody structures should mostly remain intact after the molecules are deposited on the surface. Unfortunately, a variety of experimental studies have shown that this is not the case. For example, Butler et al. somewhat famously noted that 90% of antibodies passively adsorbed to a polystyrene surface became inactive. [30] A similar study showed that antibodies passively adsorbed to polystyrene particles did so irreversibly. [122] While the molecules were difficult to wash off of the surface (which was good), only about half of the antibodies continued to function, indicating that significant structural perturbations had occurred. Similar irreversible aggregation of antibodies was found by Bee et al. on stainless steel surfaces, which are ubiquitous in biotechnology and pharmaceutical manufacturing. [45] However, Bee noted that 63% of antibodies immobilized to glass surfaces could be brought back into solution, indicating that most of the antibody structure remained intact.

Such studies show that antibody interactions can vary wildly between different surfaces. Not only is high temperature stability important, but in some cases low temperature stability may come into play as well (since biological solutions are usually stored at or near freezing temperatures). Interestingly, some antibodies (called “cryoglobulins”) denature at cold temperatures. [123] Researchers are unsure why this occurs, but in either case, denaturation of antibodies tends to be irreversible because the unfolded molecules aggregate together. [32,49,124]

While most of the antibody structure is composed of the “immunoglobulin fold,” [32] the different primary domains retain distinct characteristics. In particular, the upper portion of the Fc region, called the \( C_{H2} \) domain, is glycosylated (contains covalently bound sugar molecules) when the antibodies are produced by eukaryotic cells. This glycosylation is important for maintaining structural rigidity of the \( C_{H2} \) domain. [49,125,126] As a result, the \( C_{H2} \) domains of deglycosylated
antibodies demonstrate lower melting points than those with glycosylation. It is believed that these sugar molecules also modulate the binding affinity of the Fc domain (the C₃H₃ domain in particular) with cell surface receptors. [44, 125, 126] However, the glycosylation is not essential for the antigen binding activity of the fabs. The protein model for this study will neglect these sugar molecules because there are no parameters for conjugated sugar molecules. Although the lack of glycosylation may destabilize the antibody structure some, it provides a reasonable approach to model development, since manufactured antibodies are often left deglycosylated due to difficulties developing cell lines that produce the correct glycosylation patterns. [46]

2.6.2 In silico antibody-surface studies

A few prior studies of antibodies interacting with surfaces can be found in the literature. Sheng et al. used a 12-site CG antibody model to study how these molecules interact with charged surfaces. [127] They were able to show that the preferred orientation of the antibody on the surface was a consequence of the balance between van der Waals and electrostatic interactions. Specifically, the antibody preferred to lay flat on the surface to maximize van der Waals interactions when they were dominating, but it would more likely be found upright on a strongly charged surface. They noted that their study did not consider solvent-induced effects near the surface. Moreover, they suggested that having one dipole moment for the entire antibody may miss aspects of the adsorption process, since electric dipole moments for the fabs can be distinct from the Fc domain.

A later study by Zhou et al. repeated the work of Sheng et al. using a united-residue model of two antibody structures in an attempt to better understand the protein-surface interactions. [128] Their results also suggest that antibodies, because of their dipole moment, can display preferred orientations on charged surfaces. However, this behavior is dependent upon the strength of the dipole moment, the charge density of the surface, and the ionic strength of the solution. At low charge densities antibodies had freedom to adopt various orientations on the surface while stronger electrostatic interactions with the surface caused antibodies to adopt a specific orientation. Their higher-resolution model provided more structural flexibility than the 12-site model of Sheng et al., but still did not take into account the effects of protein unfolding. Thus, this work misses essential behavior that can happen if the protein structure has the opportunity to break down near the surface.
Javkhlanlugs et al. used all-atom molecular docking to study the details of how the antibody 1IGY adsorbed to polystyrene surfaces. [67] They found that the antibody preferred to adsorb in either a fab-fab-on, fab-fc-on, or an fc-on orientation, with the fab-fab-on orientation being the most-energetically stable configuration. They also found that the C\textsubscript{H}2 domains strongly interacted with the polystyrene (hydrophobic) surface. Their results suggest that when the fabs were in contact with the surface that the V\textsubscript{L} and V\textsubscript{H} domains mediated many of the interactions. This finding could help explain why antibody-surface behavior varies so much, since these portions of antibodies are intended to vary from one molecule to the next. However, this meant that their findings were possibly not applicable to all antibodies. Additionally, they were unable to observe any large-scale structural changes because of the complexity of their model system.

Finally, De Leo et al. also used all-atom docking methods to study how the antibody 1IGT formed complexes with different sized carbon nanotubes. [129] They found that the fabs interacted more with the smaller nanotubes while larger nanotubes would mainly interact with the Fc domain. Again, these results suggest that fab-surface interactions may be one of the factors that affects how individual antibodies react to surfaces. However, they concluded from their sequence analysis that most antibodies should interact similarly to the carbon nanotubes. They also saw that the configurations of the antibody with the nanotubes changed as the tube diameter increased, such that larger diameters encouraged the antibody to increase its contact with the surface. Again, this highlights the fact that protein molecules can be significantly attracted to particular surfaces, which can lead to unexpected behaviors. Ultimately, though, their study they did not see any significant secondary structure perturbations, and they suggested that their all-atom approach possibly precluded them from seeing such events.

2.6.3 \textit{In silico} studies of small proteins on surfaces

CG protein models with residue-level detail have proven useful for understanding the thermodynamics of small proteins on surfaces. Initial work by Knotts et al. showed that attaching protein A to either a neutral or weakly attractive surface degrades its structural stability at higher temperatures or when mechanical stress is applied. [130] Moreover, attachment to an attractive surface encouraged the protein to become trapped in a meta-stable, partially-unfolded state that would be non-functional. A later study by the same author investigated the enthalpic and entropic
effects of protein adsorption to a surface. [131] They demonstrated that attaching the protein to a surface can entropically stabilize the protein relative to bulk solution. In such cases, the surface reduces the entropic cost of achieving the proper folded state because it prevents a number of unfolded configurations from forming, simply because the surface reduces the available volume near the protein. The result is that folded configurations form a higher proportion of all of the available configurations and the folded state is more stable. This can be seen by comparing their entropy calculations while on the surface with those in bulk solution. However, they concluded that the surface may ultimately destabilize the protein by interfering with the favorable intra-protein enthalpic contacts between residues that maintain tertiary and secondary structures.

Friedel et al. also looked at the energetic contributions to the stability of a β-barrel protein on a surface. [132] Their results showed that the vibrational entropy of the protein on the surface has the greatest influence on protein stability. While they also concluded that the surface inhibits the formation of unfolded configurations, they suggested that this line of thought cannot explain why decreased protein stability sometimes happens. As a result, several thermodynamic components must be considered together to better understand protein behavior. Wei et al. also showed that the tether should be placed such that it allows the protein room to rotate and vibrate on the surface to maintain its stability. [37] This freedom of movement is necessary because at ambient temperatures there are vibrational, rotational, and translational states want to be occupied by the atoms inside of the protein. If the protein is tethered such that it becomes overly-restricted on the surface, it will denature to make these energetic states available. These studies highlight the idea that surface tethering need not always be destabilizing to small proteins, but it is unknown if similar behaviors also apply to larger antibodies as well.

Knotts et al. initially suggested that it is unlikely that the optimal tether site on the protein can be determined a priori. [130] However, work by Wei et al. found that stability can be correlated to certain secondary structural elements for a subset of proteins called all-alpha orthogonal proteins. [37] There are recognizable structural elements that, when used as a tether point, consistently impact the protein’s stability. Essentially, any point on an exterior loop that points away from the protein body serves as a stabilizing tether point while loop segments that point inward will be destabilizing. In a later study Wei et al. showed how the tether location affects the stability of a larger, more complicated protein (lysozyme). [38] They found a particular tether
location which eliminated a partially-folded intermediate state typically seen in bulk solution (thus improving the stability of the protein on the surface). Similar work by Loong et al. found a second tethering site on lysozyme that (when combined with the first site) further increased the stability of the molecule on the surface by eliminating a second meta-stable folding intermediate. [133] This caused lysozyme’s folding behavior to become more like a two state mechanism rather than a three state mechanism. Friedel et al. looked at tether placement on a β-barrel protein, and saw that it significantly impacted the folding mechanism of that protein, too. [134] It was again shown that different tethering sites impact protein stability in different ways. Ultimately, such studies show that the tether placement is often important for small protein on the surface, because the entire molecule is close to the surface. However, the inherent flexibility and large size of antibodies might mean that they are relatively unaffected by the exact placement of the tether on the surface, so long as the fabs are available to bulk solution.

CG studies have also looked at how protein orientation relative to the surface impacts its function. Besides accounting for thermal stability, the tether site must orient the active site away from the surface. [38] Otherwise, the surface blocks the substrate from approaching the active site to become bound. As previously mentioned, Wei et al. found a particular site on lysozyme that both protected its stability and oriented it in a functionally-meaningful way on the surface. It demonstrates that it is possible to find a tether point that maximizes both the stability and activity on the surface. An experimental study reproduced this configuration and agreed that this particular site was better than other surface-bound configurations (although none were as active as the bulk solution control). [135] While simulations of fabs on surfaces are difficult to find, experimental studies have shown that proper orientation can increase binding performance upwards of 193× when the ABS is directed toward bulk solution. [31] While orientation makes a significant difference for small protein behavior, it remains uncertain how orientation affects antibody stability and function given their large size.

2.6.4 Summary of antibody stability on surfaces

Experimental and simulation studies indicate that proteins often strongly interact with man-made surfaces. Knowing more details about how antibodies behave on a microarray surfaces is crucial to improving their function. Studies of small proteins have shown that immobilization
need not be detrimental to a protein’s stability, as long as it does not affect structurally-stabilizing intra-protein contacts. Otherwise, the surface can eliminate unfolded configurations and thereby enhance the stability of the protein. Antibodies have been shown to have stable interactions with some surfaces; however, the number of simulation studies are limited. These studies either lacked important structural details or their all-atom models hampered the sampling of important configurational changes in the antibodies. As such, work involving protein models with intermediate complexity could help researchers better understand how antibodies behave on different surface types.

2.7 Antibody interactions with small antigens

2.7.1 Interaction forces

Besides antibody interactions with the surface itself, the impact of the surface on antibody-antigen binding is also critical to microarray performance. The interactions between the antibodies and their target proteins must also be stabilized so that microarray spots will produce strong signals with minimal background noise. As such, understanding how proteins bind to one another is the next major consideration.

The fundamental forces that govern the tertiary structure of proteins in bulk solution—electrostatic, van der Waals (hydrophobic effects), and hydrogen bonding—also modulate the formation of protein complexes. [136, 137] All of these forces arise from charge imbalances on atoms or entire molecules that attract atoms/molecules with (fixed or temporary) opposite charges. However, the forces are distinguished from one another by their strength and range of effect. [105, 108, 138] Electrostatic interactions involve fixed charges and can be felt over large distances. They follow Coulomb’s Law, which means the potential energy between the charges (as a function of separation distance, \( r \)) has a \( r^{-1} \) dependance. It is believed that they are a major contributor to protein-protein complexes [139] and that they induce proteins to come close to one another while they are separated by 30-80 Å. [138] This enables the formation of an “encounter complex” in which the proteins are close together but not correctly aligned to form the final structure. [136, 140–142] Further alignment requires additional forces to guide the rest of the binding process.
There are two main types of short-range interactions. One of them is van der Waals effects. These arise when the fluctuating electron cloud of a molecule creates a temporary dipole that induces an attractive dipole in nearby molecules. Such interactions are typically modeled with a Lennard Jones (12-6) potential. These interactions are strongest when molecules are within about 10 Å of one another. It has been suggested that they account for roughly 10% of the total interaction energy [138] but some results indicate that they may account for much more than that. [108] However, they can strongly influence the final configuration of the protein complex, especially because hydrophobic patches on each protein strongly prefer to become buried within the interfacial surfaces. [136] As it relates to molecular simulation, van der Waals calculations make up the bulk of the calculations, as they increase roughly as \( O(N\log(N)) \) (when using a neighborhood list) with the number of particles, and have by far the greatest impact on the behavior of the modeled systems.

The second type of short range interaction that occurs is hydrogen bonding. Hydrogen bonds involve a positively-charged hydrogen atom interacting with a nearby electronegative atom displaying lone pairs of electrons. They are often modeled with a Lennard Jones (12-10) potential. They are stronger than van der Waals effects, but their strength can be dependent upon the geometry of the atoms. Additionally, hydrogen bonds (like electrostatic interactions) quickly become diluted by the presence of nearby water molecules, which readily participate in up to four hydrogen bonds simultaneously. However, hydrogen bonds (as well as the other forces) are greatly amplified by good geometric complementarity between the interacting surfaces, [136] which can involve water molecules that become trapped between the two proteins. [1] The good fitting allows for specific amino acid features in both proteins to be properly aligned with one another and close together to maximize their attractive effects. [139, 143] Specific residues, called “hot spots,” can be of particular importance to the overall protein complex energetics because they optimize the impact of the different attractive forces. [136, 137, 139, 144]

2.7.2 Interactions in bulk solution

Antibody-antigen complexes are often seen as the prototype for protein complexes generally, so they have been studied by many researchers. Antibodies have evolved to have a very high specificity for a particular target molecule and to bind strongly to it. For example, some antibody-
antigen dissociation constants are on the order of $10^{-10}$ M (approximately 65 kJ mol$^{-1}$ of binding energy). [43] Hanasaki et al. used steered molecular dynamics to investigate the dissociation of lysozyme and from a fab fragment. [140] Their work was one of the first to study antibody binding to a protein instead of a small molecule. Their results show that allowing the protein to be flexible is important because the binding process often goes through an “induced fit” mechanism. When this is the case, both protein structures must become somewhat distorted before they find the most stable interfacial configuration (similar to reactants overcoming an energetic activation barrier). In contrast, small molecules typically bind in a “lock-and-key” style without such distortions. Significantly, Hanasaki showed that the deformations were greatest in lysozyme rather than in the fab fragment. It must be noted that their use of steered molecular dynamics may have enhanced any actual distortions that take place because of the unrealistically large separation speed that was used (4 m s$^{-1}$ acting on molecules approximately $10^{-9}$ m in diameter).

Yamashita et al. compared different simulation methods for sampling fab-lysozyme binding properties. [145] They concluded that the steered molecular dynamics method overestimates the binding strength compared to the traditional method of umbrella sampling, in a way similar to how excessive loading rates affect unbinding force measurements made by AFM. Although they attempted to fix this issue, they noted that even their new method pulls the molecules apart faster than reality. They noted part of the problem is that as the proteins separate, time must be given to allow water molecules to intercalate between them, or a vacuum will form. This penetration process cannot be sped up without excessively increasing the temperature of the simulation.

Other studies have shown that including the dynamic motions of proteins in a complex should be included in the energetic analysis. Sinha et al. found that the dynamic motions of the crystal structures of the fab-lysozyme complex are important to observing all of the interactions that contribute to antigen binding. [142] Their work highlights a significant limitation to X-ray crystallography techniques (even when they are applicable)—the crystal structure is only one snapshot of a collection of configurations that the two proteins experience in their bound state. In contrast, molecular simulation allowed them to compare the binding behaviors of different fab-lysozyme structures and recognize interactions that were similar to several of the complexes. Acchione et al. also suggested that dynamic motions are important for understanding protein complexes. [146] They used $^{19}$F NMR to study how lysozyme binding influenced the chemical en-
environment around $^{19}$F-labeled residues in different locations within the fab. Their results indicate that the impact of residue substitutions cannot be fully anticipated based off of the crystal structure, since the substitutions often had little impact on the overall structure of the fab (they still noticeably impacting the binding performance). As such, they showed that it was the dynamic motions of the fab structures that were being affected most by the substitutions and thereby affecting the binding forces.

Antibody-antigen interactions in bulk solution appear to have not been studied with CG models, but other protein-protein interactions have been. They demonstrate that CG models, despite their simplistic natures, can still be used to probe highly-specific processes such as protein recognition and binding. Basdevant et al. developed a CG protein model that also includes a CG solvent model. They chose to study the binding behavior of barnase/barstar, cohesion/dockerin, and an $\alpha$-complex. [105] Their protein model used two or three sites to represent each amino acid, and solvent particles consisted of a single site that represented three water molecules, each with an inducible dipole at its center. Both components modeled van der Waals interactions and electrostatic interactions with separate potentials. They noted that this separation is important for accurately capturing intra-molecule interactions because of the different length scales involved with each interaction. The results showed that the explicit CG solvent also replicated general hydrophobic effects (where non-solvent particles tend to stick together). The results from the barnase/barstar complex showed that both hydrophobic effects and electrostatic interactions were important for complex formation. They suggested that the particular pathway to the final bound state taken by the two molecules was guided by electrostatic attractions at long-range and by hydrophobic interactions at short-range. This concept agrees with prior Brownian dynamics work by others. [106] They noted that protein folding/unfolding was not allowed during their simulations. Doing so (1) allowed them to run simulations at very high temperatures to accelerate phase space sampling (needed because of the explicit solvent), and (2) maintain the proper electrostatic charges on each amino acid, since these were based off of the original crystal structures.

Ravikumar et al. provides a second example of CG molecular dynamics of protein-protein interactions in bulk solution. [108] They used a Gō-like model (similar to the model used later in this dissertation) with a single site representing each amino acid. Like Basdevant, Ravikumar also split interactions into van der Waals and electrostatic effects. Unlike Basdevant, they opted for
an implicit solvent effect only on surface-accessible residues. Their results also show that a CG protein model is capable of discriminating between meta-stable and crystal-like configurations, even if they are similar to one another. Moreover, meta-stable complexes with imperfect fitting were important in guiding the proteins to eventually find their crystal-like structure. They devised a “push-pull-release” system that would forcibly separate the proteins, then push them together a certain distance, and then release them to freely interact with each other before the cycle starts over again. While their method did enhance the configurational sampling, it significantly biased the forces affecting on the proteins in ways that could not be accounted for thermodynamically. Like Basdevant, they noted that protein folding was not allowed during their simulations (although it was possible with their model). While these studies show that CG models can be used for protein-protein interactions (and therefore antibody-antigen interactions), neither study involved the proteins on a surface.

2.7.3 Interactions on a surface

Surprisingly, it appears that little work has been directed toward understanding any kind of protein-protein interactions on or near a surface using molecular simulation. Of the two studies that can be found in the literature, both involved very-coarse models of antibodies (7-10 sites per antibody). They do, however, provide insights relevant to antibody-antigen binding being investigated in this dissertation. De Michele et al. studied antibodies interacting with surface-immobilized antigens (like a reverse phase microarray). [147] They looked at binding behavior as a function of antigen density on the surface. Their numerical analysis indicated that at low antigen densities the strength of the antibody-antigen bond dominated the thermodynamic behavior of the system. However as antigens became more dense on the surface the steric repulsion between the antibodies started to dominate the system. It is important to note that their antigen model was extremely simple—they were represented as a single binding site located on the surface. This simplicity precludes the ability to discern what impact finer antigen structural details might have on the binding process near the surface. Additionally, since microarrays are almost always forward phase, with the antibodies on the surface, this study does not replicate those conditions.

However, a second recent study by Della-Ventura et al. did look at immobilized antibodies interacting with a layer of antigens. [62] They simulated antibodies randomly deposited onto the
surface and showed that four predominate orientations exist on the surface. Unfortunately, three of the four configurations significantly hinder the bulk solution accessibility of the fabs. When compared to well-ordered antibodies on the surface (with their fabs pointed away from the surface), the disorderly antibodies had significantly lower (about 25%) binding capacity. Moreover, they suggested (from experiments) that proper ordering of the molecules on the surface enhances their packing ability so that the density of antibodies on the surface can be higher (although their simulation results do not reflect this). Their simulations neglected molecular dynamics of the molecules, as they considered these motions to be negligible compared to the binding energy. However, as seen with other studies, the molecular dynamics are crucial to understanding all of the interactions that take place between two proteins. Additionally, the particularly coarse nature of the molecules in both studies is unable to capture binding effects caused by the finer details of the proteins’ structures that may be significant when they are placed near a large obstruction, such as a surface. Moreover, neither study by De Michele or Della-Ventura investigated the impact of different surface properties on the antigen binding process.

2.7.4 Summary of antibodies and small antigens

Antibody-antigen interactions are the same as any other protein-protein interaction, but the importance of antibodies in medicine and research have made such complexes frequent targets of study. Protein complexes come together because of a balance of electrostatic and van der Waals forces, as well as hydrogen bonding. All of these forces arise from charge distribution imbalances on different atoms and molecules, although they differ in their range of effect as well as their specificity. Good complementarity of the contacting molecular surfaces is crucial for maximizing the attractive power of these different forces.

Of antibody-antigen complexes that have received attention, those involving lysozyme appear to be most frequently studied. Studies in bulk solution provide insights into the finer details of antibody-antigen interactions, but they have trouble accurately looking at larger-scale thermodynamics when two molecules come together. While CG models have been used, these have not been directed toward antibodies specifically. Studies of antibody binding on surfaces are insightful, but have remained highly simplified. As such, additional studies are needed that look closer at antibody interactions with small antigens, with a level of detail closer to the atomic scale but
that can enable accurate calculations of thermodynamic metrics of their behavior. Moreover, such work needs to be done for bulk solution and different surface types.

2.8 Antibody interactions with large antigens

While substrate materials and immobilization methods have received significant attention in past studies, there are many design variables that affect microarray performance. In particular, the diverse nature of antigen targets suggests that more attention should be directed towards understanding how this diversity affects antibody binding. It is noteworthy that essentially all microarray designs employ the same substrate and the same immobilization strategy for all of the affinity molecules on the chip. This design provides significant manufacturing simplicity, but it may result in poorer performance because it does not optimize affinity molecules to bind targets that come in a nearly limitless diversity of shapes, sizes, flexibilities, symmetries, epitope accessibilities, and chemical properties. Such diversity implies there may be a need to use multiple substrate materials and/or attachment strategies, for at least different categories of antigens, on the same chip to equalize binding performance across the entire array and to further reduce reliability problems. Three properties of particular interest include antigen size, number of binding sites, and antigen flexibility.

2.8.1 Effects of antigen size

Prior studies have demonstrated that the size of the antigen impacts antibody binding behavior on surfaces. Oda et al. specifically studied the stoichiometry of antibody-antigen complexes in bulk solution and compared these to complexes that form on a surface. [72] Using lysozyme, ovalbumin, and bovine serum albumin as examples, they demonstrated that antibodies can bind two antigens regardless of antigen size while in bulk solution. Conversely, when the antibody is first tethered to a surface through the Fc domain, it could only bind a single large antigen at a time (but it could still bind two smaller antigens). The authors suggested that the immobilization reduced the freedom of movement of the fabs so that a large, bound antigen would create enough steric interference to prevent a second molecule from being captured.
Figure 2.5 helps visualize the magnitude of this steric interference when dealing with large antigens. Panel (a) shows that an intact virus capsid is so large compared to an intact antibody (Panel (b)), that it would not be possible for an immobilized antibody to capture two capsids simultaneously. In contrast, antigens such as lysozyme (Panel (d)) are smaller than antibodies, so that two antigens could be captured without clashing with a nearby surface. Finally, hemagglutinin (Panel (e)) is about the same size as an antibody, so that it may or may not be possible to bind two such antigens. Panel (c) shows the size of a fab by itself, and indicates that these may have more difficulty than antibodies capturing moderately large antigens such as hemagglutinin, let alone huge structures such as intact viruses.

Later work by Oda et al. reconfirmed the finding that immobilized antibodies are unable to capture more than one large antigen. [148] They suggested that to be fully functional, antibodies attached to cell surfaces must therefore interact with antigens differently than antibodies in free solution. However, the molecular details of antibody interactions with large antigens while on a surface remain unknown. Additionally, there are no design guidelines that suggest how antibody placement on the surface could be optimized for binding large target molecules.

### 2.8.2 Effects of antigen symmetry

The number of symmetric binding sites on the antigen (binding site valency) has also been shown to impact antibody binding behavior, at least in bulk solution. Molecules composed of several symmetric domains create opportunities for more than one antibody to bind or for a single antibody to bind in two places. Oda et al. also studied the formation of antibody-antigen complexes when the antigens have multiple binding sites. [148] For example, while antibodies typically show a binding stoichiometry of Ab1:Ag2, a menagerie of complex stoichiometries such as Ab2:Ag1, Ab2:Ag2, Ab3:Ag2, etc. can form when the antigens have multiple binding sites. In the case of low-affinity antibodies, complexes were often arranged as Ab1:Ag1. This happens because both fabs must bind to the same antigen molecule to stabilize the complex. Arthur et al. studied the binding characteristics of a therapeutic antibody and its multi-valent target and also demonstrated that a variety of antibody-antigen complexes can form. [149] Significantly, the authors noted that it was important to use bulk solution techniques to study these complexes because immobilization of the affinity molecules can reduce their antigen binding capacity. As such, while
Figure 2.5: Size comparison of (a) a viral capsid, [5] (b) an antibody, (c) a fab, (d) lysozyme, and (e) hemagglutinin. [6]

Various multi-antibody, multi-antigen complexes can form in solution, it is uncertain what impact symmetric binding sites may have on antibody binding under microarray conditions.

2.8.3 Effects of antigen flexibility

Antibodies are inherently flexible molecules. [2, 72, 150] This ability is crucial to their function, but also makes crystallizing whole antibodies difficult. The first attempts to crystallize and analyze whole antibodies in the past could not resolve all parts of the antibody structure [151] and the mobile hinge segments had to be removed [152, 153] to stabilize the rest of the structure enough to produce usable crystals. Improved biochemical methods have made it possible for whole antibody structures to become available in the Protein Data Bank (PDB) [2–4] (see Figure 2.3).
The results of De Michele et al. indicate how important antibody flexibility is to their function. They simulated both rigid and flexible antibodies interacting with immobilized antigens. Flexible antibodies were $7 \times$ more likely to bind two antigens instead of one. At the same time, rigid antibodies (held in a T-shape) were only 30% more likely to capture two antigens. Huang et al. measured antibody flexibility in bulk solution using antibodies bound to gold nanoparticles approximately one third of the size of the antibody. [154] Their measurements indicated that the fab-Fc flexibility is greater than the fab-fab flexibility, complementing the finding that immobilization of the Fc domain significantly impacts antibody binding performance. In contrast, Della-Ventura et al. assumed that such flexibility is unimportant, suggesting that the energy to accomplish such large-scale movements is greater than the thermal energy available at ambient conditions. [62] In any case, work has demonstrated that antibody flexibility is limited on the surface [72, 148] and it may therefore be important to consider the impacts of antigen flexibility on the binding process near the surface, as it may help reduce steric interference with the surface.

2.8.4 Antibody-hemmaglutinin studies

A suitable fab-antigen complex is needed to study the aforementioned antigen characteristics. Hemagglutinin (HA) could serve such a role, given its large size (see Figure 2.5), three-fold symmetry, and potential large-scale flexibility. [155, 156] This influenza protein has also received particular interest because of its potential as a medical target to prevent/treat influenza infections. Xia et al. attempted to discern the binding characteristics of two broadly-neutralizing antibodies against HA and how certain mutants of the latter molecule are capable of escaping antibody binding. [157] A couple of residue hot spots provide significant contributions to the total binding energy, and if these are changed HA will be able to avoid being captured by the fab. A similar study performed by Zhou et al. looked at the mechanism of antibody deterrence by an experimentally-discovered HA mutant. [158] They determined that the mutant displaced water molecules that bridged important hydrogen bonds. Their simulation techniques allowed them to check other potential mutants, of which several caused even greater decreases in the binding affinity of the antibody. In both cases, the systems were composed of all-atom protein models comprised of only one of three HA subunits, to save computational power. No large-scale atomic movements were simulated. Moreover, neither study looked at fab-HA interactions near the surface.
2.8.5 Summary of antibodies and large antibodies

Significant emphasis has been placed on the properties of the substrate materials and immobilization methods used for microarrays. While these are important, there are many potential design variables that could be optimized. Specifically, it would be helpful to understand how different antigen characteristics affect binding performance, especially when dealing with large antigens. Three basic characteristics—antigen size, number of binding sites, and antigen flexibility—can be studied simultaneously by investigating fab-hemagglutinin interactions both in bulk solution and on a surface. Fab-HA interactions have received attention in past studies, but these were very limited in scope. Moreover, the atomic models were incomplete, no large-scale thermodynamics were measured, and no surfaces were present. Studying fab-HA interactions with a computationally efficient model with near atomic resolution could provide useful insights into how antibodies interact with such large and complex molecules under microarray conditions.

2.9 Literature review summary

Technologies involving surface-bound proteins have the potential to enable vastly improved devices that could impact medicine, scientific research, national defense, and chemical processes, making them worthy of continued development. Antibody microarrays are one such device whose current iterations continue to demonstrate serious failures in reliability. Several variables can be optimized, but the greatest unknown at this time involves the biophysics of the antibody-surface interactions and the effects of the surface on antibody-antigen binding mechanisms.

Experimental methods can probe antibody-surface and antibody-antigen interactions in a variety of ways. Some methods provide data as a function of time while others can visualize the proteins at molecular resolution. However, all current methods currently fail to provide the atomically-detailed structures of proteins and molecular dynamics that researchers need to best understand protein behavior. Molecular simulation has emerged as the primary method for providing this information. All-atom models have found extensive use because of their high level of detail, but they are limited in their ability to capture large-scale atomic movements in thermodynamically-valid ways. As such, coarse-grain models have become popular and have demonstrated their ability to simulate protein behaviors with sufficient detail but in a computationally efficient manner.
Antibody behavior on different surfaces have been extensively studied by many methods, but the details of the interactions remain elusive. Those studies that involve molecular simulation have been limited by either using highly-coarse molecular models or atomistic models that preclude proper configurational sampling. Moreover, these studies have mostly looked at antibodies on hydrophobic surfaces. As such, using an antibody model with intermediate resolution could provide a better idea of the actual thermodynamic behavior of the molecules on different types of surfaces.

Antibody binding to antigens has also received significant attention from past studies. Studies involving all-atom models have largely only been able to probe fine atomic or residue-residue interactions between the proteins. Efforts to look at larger-scale thermodynamic metrics of the overall antibody-antigen binding process have suffered from biases related to the simulation methods. However, examples of coarse-grain studies of protein-protein interactions indicate that such models can be useful for studying how these complexes come together. Antibody-antigen binding on a surface has been studied, too. However, these studies relied on highly-coarse models, did not probe different surface types, and sometimes did not even take molecular motions into account. Again, it appears that using a protein model with intermediate resolution may be helpful at providing a better view of the binding process both in bulk solution and, more importantly, on different types of surface materials.

Substrate materials and immobilization methods have received the most attention in efforts to improve microarray function. However, protein targets come in a diverse array of sizes, shapes, flexibilities, and chemistries. While it is desirable to make microarrays out of a single substrate material and deposit all of the affinity molecules in the same way, this may not ultimately achieve the best or most-balanced performance across the entire array. As such, work is needed to study some of these characteristics. Specifically, hemagglutinin provides an opportunity to study the effects of antigen size, multiple binding sites, and antigen flexibility. These characteristics cannot be tackled using all-atom models because of their complex nature, but models without residue-level resolution will not have enough detail to properly understand the antibody-antigen-surface interactions.

This literature review suggests three areas of antibody behavior that can be addressed through coarse-grain molecular simulation. It is anticipated that addressing these topics will pro-
vide new insights that will help researchers not only understand what is happening on microarray surfaces, but why proteins are misbehaving. Ultimately, the hope is that such work will show that technologies involving proteins on surfaces can be improved through rational design choices.
CHAPTER 3. OBJECTIVE AND TASKS

3.1 Objective

The primary purpose of this study was to obtain a better understanding of the residue-level interactions between antibodies and surfaces and how they affect antibody stability and antigen binding. The results should inform researchers and manufacturers about generalized antibody behaviors and help them make rational design decisions to improve the performance of their devices. Such improvements will enable antibody microarrays to find mainstream use in fields such as medical care, scientific research, national defense. In addition, the design guidelines gained from this study should be applicable to other protein complexes that are candidates for immobilization to surfaces for applications such as biochemical reactors, medical implants, and cellular signaling.

3.2 Tasks

This objective was fulfilled by accomplishing the following tasks:

1. Investigate individual antibody stability on different types of surfaces.
   (a) Develop a full-length, coarse-grain antibody model.
   (b) Simulate antibody behavior in bulk solution, on a hydrophobic surface, and on a hydrophilic surface.
   (c) Simulate the antibody on the surface in different orientations relative to the surface.
   (d) Develop a theory to explain why antibody stability differs between bulk solution and on the different kinds of surfaces.

2. Investigate antibody binding of small antigens on different types of surfaces using lysozyme.
   (a) Develop a chimeric, antibody-lysozyme coarse-grain model from two different crystal structures.
(b) Simulate antibody-lysozyme binding in bulk solution, on a hydrophobic surface, and on hydrophilic surface.

(c) Simulate the antibody on the surface in different orientations relative to the surface.

(d) Develop a theory to explain why the binding behavior changes when placed on different kinds of surfaces, and determine the optimal surface type and orientation configuration.

3. Investigate the impacts of antigen size, multiple binding sites, and antigen flexibility on antibody binding when on a surface using hemagglutinin.

(a) Develop a chimeric, antibody-hemagglutinin coarse-grain model from two different crystal structures.

(b) Simulate antibody-hemagglutinin binding in bulk solution and on a surface (single best surface and orientation configuration, as determined by Task 2).

(c) Simulate hemagglutinin as either a rigid or flexible molecule and held at different angles relative to the surface.

(d) Develop a theory to explain how these antigen characteristics impact the antibody binding performance on the surface and compare it to the behavior with lysozyme.

The remainder of this dissertation is organized as follows. The development of an antibody model and its application to study antibody stability on surfaces is discussed in Chapter 4 (Task 1). The modification of this antibody model to bind to lysozyme and its use to study the impact of surface immobilization on antibody binding of small antigens is then discussed in Chapter 5 (Task 2). The methods developed in Chapter 5 are reused in Chapter 6 to create an antibody model that binds hemagglutinin. This model is used to study the impacts of different antigen characteristics on antibody binding performance and compares this to binding behavior with lysozyme (Task 3). Chapter 7 summarizes the results of this work and provides examples of future research directions.
CHAPTER 4. ANTIBODY STABILITY ON SURFACES

4.1 Introduction

This chapter discusses work designed to study the thermal stability of individual antibodies on surfaces. The thermal stability of a protein, often measured as heat capacity ($C_v$) as a function of temperature, is indicative of the resiliency of the protein’s structure under different conditions. For example, a microarray setup that decreases the overall thermal stability of antibodies (compared to bulk solution) would likely display a shorter shelf life and poorer signal quality because many of the antibodies would be dysfunctional. As such, one goal for microarray designers is to ensure that antibody thermal stability on the surface is at least similar to their stability in bulk solution to ensure that the molecules retain their functional shape.

The hypothesis of this work was that interactions between surfaces and antibodies disrupt antibody structure, which contributes at least partly to the poor performance of current antibody microarrays. The hypothesis is tested by focusing on single antibody behavior in bulk solution (the control) and a series of conditions where the antibody is tethered to different types of surfaces in different orientations (the treatments). Specifically, the antibody was tethered to either a hydrophobic or a hydrophilic surface in either an “upright” or “flat” configuration. The upright orientation is often considered the ideal setup for antibodies on a surface because the fabs are directed away from the surface while the flat configuration often occurs when the antibody is passively adsorbed to the surface.

This chapter is organized as follows. The first section describes the antibody and surface models, simulation protocols, convergence criteria, $C_v$ calculations, antibody-surface tethering, and ABS melting criteria. The next section reports the results, which are presented as $C_v$ curves, antibody configuration snapshots, and fab melting temperatures. Finally, the results are analyzed and discussed in light of microarray design goals. The methods, results, and discussion can also be found in article 061101 of *The Journal of Chemical Physics*. [41]
4.2 Methods

4.2.1 Antibody and surface models

The antibody model used in this work was based on the Gō-like model of Karanicolas and Brooks with input files generated using the MMTSB Gō-model Builder [120, 121]. In this formalism, each residue is represented as a single site located at the Cα position of the residue in the crystal structure. Pairs of residues are assigned an attractive interaction (Equation 4.1) if there is a hydrogen bond between them or a non-hydrogen atom on one of the residues located within 4.5 Å of a non-hydrogen atom in the other residue. In this case, \( \varepsilon_{ij} \) is based on the identities of the interacting residues \( i \) and \( j \), \( \sigma_{ij} \) is the equilibrium distance between their Cα atoms in the crystal structure, and \( r_{ij} \) is the instantaneous distance between the two sites during the simulation. All other interactions (Equation 4.2) are purely repulsive. Here, \( \varepsilon_{ij} \) is the same for every interacting pair of residues and \( \sigma_{ij} \) is calculated from the two residue radii. In this model solvent effects are included implicitly as an energetic barrier included in the attractive interactions that replicates the need for two residues to displace water molecules in order to come into close contact. This model has been shown to produce folding mechanisms that both qualitatively and quantitatively agree with experiments and has been used extensively in simulation studies of protein folding in the bulk and on the surface.

\[
V_{ij} = \varepsilon_{ij} \left[ 13 \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 18 \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{10} + 4 \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (4.1)
\]

\[
V_{ij} = \varepsilon_{ij} \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} \quad (4.2)
\]

The interaction of the protein with the surface was represented using the recently-developed model of Wei and Knotts [39]. The model was designed to be compatible with the Karanicolis and Brooks protein model previously described. It was carefully parameterized using experimental data [119] to yield qualitative and quantitative agreement to experimental adsorption energies. It has been validated for large proteins not used in the parameterization. In the model, the surface is represented as a solid, homogeneous, flat plane that is either hydrophobic, moderately hydrophilic,
or hydrophilic. The potential energy between a particle and the surface is calculated using Equation 4.3.

\[
V_{\text{surface}} = \sum_{i} \pi \rho \sigma_{i}^{3} \epsilon_{i} \left[ \theta_{1} \left( \frac{\sigma_{i}}{z_{is}} \right)^{9} - \theta_{2} \left( \frac{\sigma_{i}}{z_{is}} \right)^{7} + \theta_{3} \left( \frac{\sigma_{i}}{z_{is}} \right)^{3} - (\theta s \chi s + \theta p \chi p_{i}) \left( \frac{\sigma_{i}}{z_{is}} \right)^{3} \right]
\]  

Here \( \rho = 1.0 \text{ Å}^{-3} \), \( \sigma_{i} \) and \( \epsilon_{i} \) are the Gō-like model parameters for residue \( i \), \( z_{is} \) is the distance between residue \( i \) and the surface, \( \theta_{1} = 0.2340 \), \( \theta_{2} = 0.4936 \), \( \theta_{3} = 0.1333 \), \( \theta_{s} = 0.0067 \), and \( \theta_{p} = 0.0333 \). The terms \( \chi s \) and \( \chi p_{i} \) represent the hydropathy indices for the surface type and residue \( i \), respectively, and are available in the original reference [39].

This surface model has two important components. The first is an attractive well that varies with each surface type/residue combination. The second is an entropic “desolvation” penalty that replicates the reality that proteins or residues must displace water molecules next to the surface in order to adsorb. This desolvation penalty also varies with each surface type/residue combination.

### 4.2.2 Simulation protocols

Simulation work for this paper involved five conditions: bulk solution, hydrophobic surface flat orientation, hydrophobic surface upright orientation, hydrophilic surface flat orientation, and hydrophilic surface upright orientation. Each condition was simulated with six independent replicates for statistical analysis. All simulations were conducted with the replica exchange method [100] using the Nosé-Hoover Chain integration method [159–161] involving three thermostats of mass \( 10^{-26} \text{ kg Å}^{2} \). Each simulation consisted of 80 replicas spaced two Kelvin apart starting at 242 K and ending at 400 K. The large number of replicas was used to ensure that the complex folding behavior of the multi-domain molecule was adequately captured. The time step size was 3 fs and each simulation was run for 10 million equilibration steps and 30 million production steps. The overall simulation time used to generate each heat capacity curve was therefore 720 nanoseconds. Swaps between replicas were proposed every 2,000 steps. Previous studies [37, 38, 133, 162] have shown that this replica exchange protocol produces well-converged
and accurate estimates of the thermodynamic properties for systems modeled using the formalisms described above.

4.2.3 Convergence of structures

Figure 4.4 contains representative snapshots of the antibody configurations under different conditions. However, to help demonstrate convergence Figure 4.1 shows the radius of gyration in Å as a function of simulated time (ns) for the four simulations from which the bulk snapshots were taken. Here, blue represents snapshot 1 (260 K), red represents snapshot 2 (308 K), yellow represents snapshot 3 (340 K), and purple represents snapshot 4 (390 K). The grey area highlights the equilibration portion of the simulation. As can be seen, the average radius of gyration quickly converges to an equilibrium value during the first part of the simulation. The jagged appearance of each plot is a result of configuration swaps that are part of the replica exchange method. Similar behavior is seen in the simulations of surface/orientation conditions.
4.2.4 Heat capacity calculations

The heat capacity, $C_v$, is related to the fluctuations of the potential energy $\langle U \rangle$ according to

$$C_v(T) = \frac{\langle U^2 \rangle_T - \langle U \rangle_T^2}{RT^2} \quad (4.4)$$

where $R$ is the gas constant, $T$ is the temperature, and the $\langle \rangle$'s denote the average of the corresponding quantities. Ensemble averages were generated using the multistate Bennett acceptance ratio (MBAR) method implemented in the program pyMBAR [101].

4.2.5 Tethering

Tethering to the surface was accomplished using a harmonic restraint with an interaction potential, $U_{\text{restraint}}$, of the form

$$U_{\text{restraint}} = \frac{1}{2}kr^2 \quad (4.5)$$

Here $k = 100$ kcal mol$^{-1}$ is the strength of the restraint and $r$ is the distance of the restrained site from its original position. Tether sites were placed 5.8 Å above the surface. The two sites used to tether the antibody to the surface were site 1158 (His 286, chain D) for the flat orientation and site 1292 (Glu 420, chain D) for the upright orientation.

4.2.6 ABS melting temperatures

The melting temperatures of the antigen binding sites (ABS) shown in Figure 4.2 were calculated using the native contact maps produced by each simulation. A native contact map is a list of all of the attractive interactions in the system and indicates whether each one is present at a particular snapshot in the simulation. As per the model of Karanicolos and Brooks, a native contact is present if the distance between the participating residues is within 1.2x the distance obtained from the crystal structure. The native contact map can be segmented according to antibody substructures ($V_L$, $C_H$, etc.) and these pieces are processed to produce the fraction native curves for these subdomains. The melting temperature is the temperature at which the fraction native value equals 0.5.
To calculate the ABS melting temperature for a particular surface type/orientation combination, the native contact map data for the two V<sub>L</sub> subdomains were extracted from those simulations and processed into fraction native curves. An example of the fraction native curves is shown in Figure 4.2. Notice that the V<sub>L</sub> subdomain produces a single folding transition. The temperatures reported in Figure 4.2 were chosen from the lower of the two melting temperatures for each condition. The melting temperatures are not equal as the C<sub>H</sub>2 and C<sub>H</sub>3 subdomains are not exactly symmetric with respect to the Fabs. Despite this asymmetry, the two melting temperatures agree to within 10 K.

4.3 Results

4.3.1 Figure descriptions

Figure 4.3 shows the C<sub>v</sub> for the antibody as a function of temperature under different conditions. The curves were normalized to make the highest value in each curve equal to one to facilitate comparison of the results from different surface types. The temperature range of Figure 4.3 is limited to 330 K to 390 K because outside of this region the C<sub>v</sub> becomes flat. Panel (A) corresponds to the antibody in bulk solution with no surface present (the control), Panel (B) the antibody tethered to a hydrophobic surface, and Panel (C) the antibody tethered to a hydrophilic surface. In Pan-
Figure 4.3: Relative heat capacity as a function of temperature for the antibody in (A) bulk solution, (B) on a hydrophobic surface, and (C) on a hydrophilic surface. For panels (B) and (C) solid lines represent tethering in a flat orientation and dashed lines represent tethering in an upright orientation. Roman numerals correspond to antibody configurations shown in Figure 4.4.
Figure 4.4: Representative configurational snapshots of the antibody on different surface types. The time evolution of the radius of gyration for the bulk case is included in Figure 4.1 to demonstrate convergence of structures to equilibrium configurations. Panel (A) represents configurations from the antibody in bulk solution, Panel (B) represents the antibody tethered to a hydrophobic surface, and Panel (C) represents the antibody tethered to a hydrophilic surface. Each configuration has matching roman numeral labels on Figure 4.3.
4.3.2 Antibodies in bulk solution

The folding behavior shown in Panel (A) of Figures 4.3 and 4.4 represents the behavior of the antibody in its native aqueous environment. Notice that three peaks are present in the bulk case. This behavior agrees with experimental $C_v$ curves of other antibodies [49, 126, 163] which have been shown to contain two or three peaks. Before denaturation, the Fabs and Fc domains are free to move and rotate as seen in Figure 4.4A-i. The lowest-temperature peak, located at 353 K, represents the denaturation of the $C_{H2}$ subdomains of the Fc domain and the $C_{H1}$ and $C_L$ subdomains of the Fabs (Figure 4.4A-ii). The tallest peak, located at 359 K, represents the denaturation of all of the $V_L$ and most of the $V_H$ subdomains, which contain the antigen binding sites (Figure 4.4A-iii). The highest-temperature peak, located at 374 K, represents the denaturation of the $C_{H3}$ subdomain of the Fc domain, after which the remainder of the $V_H$ subdomains finally denatures (Figure 4.4A-iv).

4.3.3 Antibodies on a hydrophobic surface

Panel (B) of Figures 4.3 and 4.4 shows the folding behavior of the antibody tethered to a hydrophobic surface. Unlike the bulk case, and regardless of attachment geometry, the $C_v$ curves now show only a single folding transition that contains contributions from the folding behavior of all of the domains in the antibody. The peak occurs at 351 K for the flat orientation and 347 K for the upright orientation. The simplified folding behavior and decreased overall stability of the antibody results from the destabilization of the $V_L$, $V_H$, and $C_{H3}$ subdomains which causes them to denature at the same time as the weaker subdomains $C_L$, $C_{H1}$, and $C_{H2}$, which themselves have also been destabilized (compared to the bulk case) by the attraction to the surface. Figure 4.4 shows that regardless of the attachment geometry of the bulk case) by the attraction to the surface. Figure 4.4 shows that regardless of the attachment geometry of the antibody, the molecule ultimately adopts a collapsed conformation where the $C_{H2}$ subdomains have denatured and are adsorbed on the surface (Figure 4.4B-i). The data show that denatured portions of the antibody have a strong affinity for the hydrophobic surface. At higher temperatures the entire molecule adsorbs flat onto the surface (Figure 4.4B-ii).
4.3.4 Antibodies on a hydrophilic surface

Panel (C) of Figures 4.3 and 4.4 show the folding behavior of the antibody tethered to a hydrophilic surface. When the antibody is tethered in a flat orientation, the folding behavior of the different subdomains become distinct from one another similar to the behavior in bulk solution. Four peaks are found in the $C_v$ curve, located at 355 K, 362 K, 377 K, and a final shallow peak at 381 K. The first peak corresponds to the denaturation of most of the Fabs and all of the $C_{H2}$ subdomains (Figure 4.4C-ii), the second and fourth peaks the denaturation of the $V_H$ subdomains, and the third peak the denaturation of the $C_{H3}$ subdomains. Figure 4.4C-i shows that at lower temperatures the antibody maintains its tertiary structure and has more freedom of movement compared to the antibody on the hydrophobic surface. In contrast to the hydrophobic surface, denatured parts of the molecule are not tightly adsorbed onto the surface and can easily move away from the substrate bulk phase (Figure 4.4C-iii). This change in behavior is due to the decreased attractiveness between the hydrophilic surface and the antibody compared to the hydrophobic surface. The $C_v$ curve of the antibody in the upright position continues to have a single, though broad, peak located at 355 K, but a low peak at 377 K is also present. The broadness of the peak at 355 K indicates that most of the individual domains comprising the antibody fold over a more narrow temperature range than in bulk solution. The peak at 377 K corresponds to the folding of a portion of the $V_H$ subdomains that are stable at high temperatures.

4.3.5 Fab melting temperatures

Because antibody microarrays function on the principle of antigen binding, the stability of the antigen binding sites (ABS) of the Fabs on the surface is of paramount importance. Figure 4.5 depicts the melting temperature of the ABS. The term “melting temperature”, as used in this chapter, is a biological term that does not refer to a thermodynamic heat of fusion melting temperature. The ABS were considered melted when 50% of the native contacts in the $V_L$ subdomains were broken, and the temperature at which this occurs is called the melting temperature. See the Supplementary Information for more details on how the melting temperatures were calculated.

When no surface is present the ABS melt at 359 K. However, when the antibody is placed on a hydrophobic surface the melting temperature dramatically decreases to 346 K for both the
flat and upright orientations. Moreover, to within statistical error, the stability of the ABS on the surface is the same regardless of attachment geometry. The stabilities of the ABS for the flat and upright tethering are similarly equal on the hydrophilic surface, but in contrast to the hydrophobic surface the ABS on a hydrophilic surface are almost as stable as they are in the bulk.

### 4.4 Discussion

#### 4.4.1 Antibody behavior on the surface

The hypothesis of this work was that *interactions between surfaces and antibodies disrupt antibody structure, which contributes at least partly to the poor performance of current antibody microarrays*. The results indicate that antibody stability is a strong function of surface hydrophobicity, but not attachment geometry. This stands in contrast to previous work showing that the stability of small proteins on surfaces is strongly affected by the placement of the tether. For small proteins, the entire molecule is close to the surface and this close proximity impacts the rotational and vibrational freedom of the entire molecule [37, 38]. In contrast, the size of the antibody and the flexible loops that connect the Fabs and Fc domains impart a high degree of rotational and vibrational freedom even when tethered to the surface. Rotational and vibrational freedom also correlates to the changes in stability among different types of surfaces. The decreased attraction
between the hydrophilic surface and the antibody, compared to the hydrophobic surface, allows most of the molecule to move freely above the surface. In contrast, the stronger attraction between the hydrophobic surface and the antibody encourages the weaker C$_{H2}$ subdomains to denature and become adsorbed onto the surface, which in turn causes the entire molecule to remain close to the surface. As has been reported elsewhere [37,132,134], the result is a decrease in stability driven by the system desiring to access rotational and vibrational states that are not available until unfolding provides increased conformational freedom.

4.4.2 Effects of different surface types

Perhaps the most important result is that the antibody tends to collapse onto the hydrophobic surface even at temperatures below the main melting transition of the antibody. This insight is significant because it contradicts the notion that when tethered in an upright position the antibody stays in a more stable or more accessible conformation [54]. Collapse occurs because some portion of the structure, usually the C$_{H2}$ subdomains, partially denatures and then becomes adsorbed onto the surface. The destabilization of the antibody and the subsequent adsorption of the random coil state on the surface may explain some of the reasons for the unreliable performance of antibody microarrays. Unfolding and adsorption of the antibody onto and steric hinderance from the surface likely disrupts antigen recognition and binding. In this case, antigens would need to maneuver to the surface to interact with Fabs that are restricted to move in the plane of the surface. Moreover, extensive denaturation could lead to a loss of function of the Fabs or non-specific binding of proteins. The latter occurs because the surface becomes more attractive to proteins in solution as denatured antibody coils cover it. In short, several mechanisms leading to decreased and highly variable signals from antibody microarray spots are supported by the results.

4.4.3 Applications to microarrays

In light of this, the data suggest that antibody microarray performance could be improved by using a hydrophilic substrate. Traditionally, hydrophobic surfaces are used as deposition of antibodies is easily done by spotting. Using a hydrophilic surface would require site-specific covalent linkage of the antibodies to the surface to prevent leaching during microarray storage and
use. Recent developments are making this highly-controlled covalent tethering possible [25], and future work in this area is promising [26].

4.5 Conclusion

In summary, this work was done to investigate how surface hydrophobicity and tether orientation affect the overall stability of antibodies to aid antibody microarray design. The results show that surfaces can have a dramatic effect on the folding behavior and stability of antibodies while the attachment geometry has little impact due to the size and flexibility of the molecule. The individual domains of antibodies collapse onto and denature more easily on hydrophobic surfaces but are more likely to desorb and remain stable on hydrophilic surfaces.
CHAPTER 5. ANTIBODY-LYSOZYME BINDING ON SURFACES

5.1 Introduction

Investigations of antibody behavior on surfaces [41, 62, 67, 129, 147] and antibody-antigen binding in bulk solution [140, 142, 145, 164] using molecular simulation have yielded valuable insights into the behavior of antibodies by themselves or with their antigens. However, few details exist regarding how antibodies bind to their antigens while attached to a surface. As such, moving microarray performance forward now requires a better understanding of how surface immobilization affects binding. The work in this chapter examines how the surface affects the ability of a fab to bind to its antigen.

The hypothesis is that the presence of the surface disrupts normal fab-antigen binding through a combination of steric interference and attractive forces between the proteins and the surface. This hypothesis is tested by using molecular simulation to focus on single fab-antigen dynamics in bulk solution (the control) and a series of cases where the fab is immobilized to a planar surface (the treatments). The fab was attached to either a purely-repulsive, hydrophobic, or hydrophilic surface in either a “flat” or “upright” configuration. A flat configuration keeps the antigen binding site (ABS) of the fab close to the surface while the upright configuration allows the body of the fab to move and largely avoid contact with the surface.

The rest of this chapter is organized as follows. The first section describes the fab-antigen and surface models, fab tethering, simulation setups for bulk solution and on the surface, the methods used to calculate the results, simulation protocols, and the setup for replicates. The next section presents the results, which are free energy landscapes, or potentials of mean force (PMF), for the binding of the fab to the antigen and representative snapshots of the proteins’ configurations. They show that the tethering configuration and surface hydrophobicity significantly impact antigen binding. Analysis of the results suggests that the optimal tether-site/surface-type combination is an upright fab on a hydrophilic surface because such minimizes negative entropic changes and
5.2 Methods

5.2.1 Experimental design

The goal of this work was to determine how surface tethering affects fab-lysozyme binding. Because antibodies are naturally found in aqueous environments, behavior in bulk solution was chosen as the control to which surface-bound treatments were compared. Surface-induced effects were studied by tethering the fab to one of three kinds of surfaces: purely-repulsive, hydrophobic, and hydrophilic. The purely-repulsive surface was used to help elucidate whether changes in binding behavior resulted from steric interference or attractive interactions with the surface. Also, as many protein immobilization techniques can result in random orientations of the protein relative to the surface, two representative configurations were tested—the “flat” configuration, where the fab is tethered close to the ABS, restricting its motion, and the “upright” configuration, where the tether is placed at the end of the molecule so that the remainder of the fab has a large range of motion relative to the surface.

5.2.2 Protein model

To achieve the adequate sampling of phase space needed to calculate the relevant thermodynamic properties of interest, the fab and lysozyme were modeled using the Gō-like model of Karanicolas and Brooks. This model yields protein folding mechanisms and stabilities that are in good quantitative and qualitative agreement with experimental data, but does so in a computationally permissible manner. [120, 121] In this coarse-grain formalism, each residue is represented as a single site located at the Cα position of the residue in the crystal structure, which for this study is the crystal structure 1YQV as found in the Protein Data Bank. [1]

As with all Gō-like models, the tertiary structure of the protein is maintained by defining native contacts. For this model, pairs of residues form such contacts if they comprise a hydrogen
bond in the native structure or if one of the heavy atoms in one residue is found within 4.5 Å of a heavy atom of another residue. The energy of a native contact, $V_{ij}$, takes the form:

$$V_{ij} = \varepsilon_{ij} \left[ 13 \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 18 \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{10} + 4 \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

(5.1)

Here, $r_{ij}$ is the instantaneous distance between the two sites comprising the native contact, $\varepsilon_{ij}$ is the interaction energy between the two residues forming the native contact and is based on the experimental energies reported by Miyazawa and Jernigan, [165] and $\sigma_{ij}$ is the distance parameter of the native contact and is set equal to the distance between the two $\alpha$-carbons of the residues comprising the contact. Solvent effects are included implicitly in this potential as an energetic barrier (the $10^{th}$ power term) that replicates the cost of displacing water molecules as two residues come into close contact. All other interactions, those that are not native contacts, are purely repulsive. The reader is referred to the original papers for more details on the model. [120, 121]

### 5.2.3 Surface models

Interactions between the protein and the surface were represented by two different surface models. The first model was a purely-repulsive surface with a potential function as shown in Equation 5.2. [37] Here, $\varepsilon_s = 0.0363$ kcal mol$^{-1}$, $\sigma_{is}$ is residue-specific, $z_{is}$ is the distance between site $i$ and the surface, and $c$ is also residue-specific such that $V_{\text{surface}}$ falls to zero when $z_{is} = \left(\frac{2}{5}\right)^{1/6} \sigma_{is}$. All residues interact with this surface in the same way, namely, a short-ranged, purely-repulsive potential, regardless of residue chemistry.

$$V_{\text{surface}} = \sum_i^N \varepsilon_s \left[ \left( \frac{\sigma_{is}}{z_{is}} \right)^9 - 7.5 \left( \frac{\sigma_{is}}{z_{is}} \right)^3 + c \right]$$

(5.2)

The second model of the surface, used for the hydrophobic and hydrophilic treatments, is that of Wei and Knotts. [39] The mathematical form of the potential is shown in Equation 6.2. Here $\rho = 1.0$ Å$^{-3}$, $\sigma_i$ and $\varepsilon_i$ are the model parameters for residue $i$, $z_{is}$ is the distance between residue $i$ and the surface, $\theta_1 = 0.2340$, $\theta_2 = 0.4936$, $\theta_3 = 0.1333$, $\theta_5 = 0.0067$, and $\theta_P = 0.0333$. The terms $\chi_S$ and $\chi_P$ represent the hydropathy indices for the surface type and residue $i$, respectively. Specifically, the indices are $\chi_S = 4.5$ for the hydrophobic surface and $-1.0$ for the hydrophilic surface.
The values of $\chi_P$ for each amino acid type can be found in the original reference. This surface model was designed to work with the coarse-grain protein model of Karanicolas and Brooks, and, as is shown in the original reference, demonstrates both qualitative and quantitative agreement with experimental protein adsorption studies.

$$V_{surface} = \sum_i^N \pi \rho \sigma_i^2 \epsilon_i \left[ \theta_1 \left( \frac{\sigma_i}{z_{is}} \right)^9 - \theta_2 \left( \frac{\sigma_i}{z_{is}} \right)^7 ight] + \theta_3 \left( \frac{\sigma_i}{z_{is}} \right)^3 - (\theta_S \chi_S + \theta_P \chi_P) \left( \frac{\sigma_i}{z_{is}} \right)^3 \right]$$

(5.3)

5.2.4 Fab surface tethering

As mentioned above, the fab was attached to the surface at one of two different sites to create a “flat” and “upright” configuration (see Figure 5.1). The tethering site was SER 286B for the flat orientation and LYS 431B for the upright configuration. For both configurations the tethering site in the fab was held at the coordinate (0, 0, 5.8), or 5.8 Å above the surface, using a harmonic restraint with a spring constant of 100.0 kcal mol$^{-1}$. This approach of using a harmonic restraint as a tethering method has been done previously. [37, 38, 41, 133]

To aid interpretation, a consistent color scheme has been used to depict the fab and lysozyme throughout this work. As shown in Figure 5.1, the fab is divided into three colors where red represents the $V_L$ and $V_H$ domains which together contain the ABS, orange represents the $C_L$ and $C_{H1}$ domains, and white represents loops that connect these together. Lysozyme, as shown in Figure 5.2, is depicted in ice blue.

5.2.5 Bulk solution simulations

Simulations of fab-lysozyme interactions in bulk solution were done using 1D umbrella sampling. The reaction coordinate, $\xi_1$, was defined to be the distance between the ABS and the antigen as pictured in Figure 5.2. The specific sites used to define this distance were GLU 35B of the fab and SER 50 of lysozyme. The umbrella restraint was a harmonic potential of the form
Figure 5.1: Depictions of the fab for the two tethering configurations investigated in this study. (A) The flat tethering configuration, and (B) the upright tethering configuration.

\[ U_{\xi_1} = \frac{1}{2} k_{spring} (\xi_1 - \xi_{1,0})^2 \]  

(5.4)

where \( U_{\xi_1} \) is the potential energy of the restraint, \( k_{spring} \) is the strength of the restraint, \( \xi_{1,0} \) is the equilibrium separation distance for a given simulation, and \( \xi_1 \) is the instantaneous value of the reaction coordinate. Simulations were performed for \( \xi_{1,0} = 10.0, 11.0, 12.0, ..., 110.0 \ \text{Å} \).

The choice of GLU 35B and SER 50 was made because it allows the \( \xi_1 \) restraint to serve as an axis around which the best-fit planes containing the ABS sites for the fab and lysozyme, respectively, are able to rotate when the fab and lysozyme are close to one another. An example of this rotation is shown in Figure 5.3, Snapshot (B) (discussed later). For reference, \( \xi_1 = 14.7 \ \text{Å} \) in the bound state as described by the crystal structure.

Simulations started from the PDB crystal structure, so extensive equilibration was done to ensure thermal equilibrium was reached before acquiring data for evaluation. Each simulation consisted of two equilibration phases and one production phase. In the first equilibrium phase, \( k_{spring} \) was set to 10.0 kcal mol\(^{-1}\) Å\(^{-2}\) and the umbrella restraint served to pull the two molecules to their assigned value of \( \xi_{1,0} \). This phase consisted of \( 5 \times 10^5 \) steps. In the second phase, \( k_{spring} \) was set to 0.25 kcal mol\(^{-1}\) Å\(^{-2}\) and the two molecules were allowed to thermodynamically equilibrate for \( 3 \times 10^7 \) steps. In the production phase, from which data were taken for analysis, \( k_{spring} \) was the same as the second phase and a total of \( 5 \times 10^7 \) steps were run. In total, each single simulation consisted of 60 ns of equilibration followed by 100 ns of production. As mentioned above, 51 total values of \( \xi_{1,0} \) were simulated, so each 1D PMF presented below consists of just over eight microseconds of simulation time.
5.2.6 Surface-bound simulations

Both 1D and 2D umbrella simulations were done with the fab tethered to a surface. The 1D simulations were done following the same scheme as that described for the bulk (e.g. $\xi_1$ was the distance between GLU 35B and SER 50, three simulation phases were done, etc.) As will be described below, the results from these simulations were difficult to interpret because a single reaction coordinate could not distinguish between separate states of the system. To overcome this difficulty, a second reaction coordinate was added, and 2D umbrella simulations were performed. The second reaction coordinate, $\xi_2$, serves to identify the distance between lysozyme and the surface. Specifically, $\xi_2$ is defined as vertical distance of SER 50 from the XY plane placed at $Z = 0$, and is depicted schematically in Figure 5.2. The form of this second umbrella restraint was

$$U_{\xi_2} = \frac{1}{2} k_{\text{spring}} (\xi_2 - \xi_{2,0})^2$$

(5.5)

where $k_{\text{spring}} = 0.25$ kcal mol$^{-1}$ Å$^{-2}$, $\xi_{2,0}$ is the equilibrium value for the umbrella, and $\xi_2$ is the Z-position of SER 50 at a given time step. The second restraint maintained the height of lysozyme above the surface but allowed lysozyme to move freely in the X and Y directions. The range of values for $\xi_{2,0}$ depended on whether the fab was in a flat or upright orientation. For the flat orientation $\xi_{2,0}$ was 10.0, 12.5, 15.0, ..., 50.0 Å and for the upright orientation it was 10.0, 12.5, 15.0, ..., 90.0 Å. For all 2D umbrella simulations $\xi_{1,0}$ was 10.0, 12.5, 15.0, ..., 110.0 Å. These setups yielded 697 simulations for each flat orientation case and 1,353 simulations for each upright orientation.
5.2.7 PMF calculations

The PMFs were calculated from the production data using the Weighted Histogram Analysis Method (WHAM). [103] Because this method produces the PMF to within an arbitrary, additive constant for all values of $\xi_1$ and $\xi_2$, comparisons are made easier by shifting the entire energy landscape to be zero where protein-protein and lysozyme-surface (for treatments) interactions are negligible. This occurs at high values of $\xi_1$ and $\xi_2$, so a point in each PMF at this location was chosen to be the reference point. The PMF value of the reference point was then subtracted from the entire PMF, making the PMF value at this location approximately zero. With this approach, non-interacting regions have PMF values near zero, energetic barriers have positive PMF values, and energetically favorable regions have negative PMF values.

5.2.8 Simulation protocols

To prevent the molecules from unnaturally unfolding due to forces placed on them by the umbrella restraints, additional harmonic restraints were placed on each of the intra-protein native contact pairs. [166, 167] The spring constant for these additional restraints was 10.0 kcal mol$^{-1}$ Å$^{-2}$ for the first equilibration phase and 1.0 kcal mol$^{-1}$ Å$^{-2}$ for the second equilibration phase and production. All simulations had a time step of 0.002 ps, were done at 300 K, and used the the Nosé-Hoover Chain integration method [159–161] involving three thermostats of mass $10^{-26}$ kg Å$^2$.

5.2.9 Replicates

Independent replicates of the PMFs for the fab-lysozyme system in bulk solution and on the surface were done to estimate the uncertainties in the results. Replicates were created by randomly assigning the initial velocities using a different seed value for the random number generator in each case. Uncertainties were calculated as $\sigma/\sqrt{N}$ where $\sigma$ was the standard deviation of the PMF at a specific value of $\xi_1$ and $\xi_2$ and $N$ was the number of independent replicate simulations performed. For each case three replicates were performed.
5.3 Results

5.3.1 Fab-lysozyme interactions in bulk solution

Figure 5.3 shows the PMF for the fab-lysozyme system in bulk solution (the control) along with representative configurational snapshots of the proteins’ behavior. The horizontal axis shows the separation distance between the fab and lysozyme (\( \xi_1 \)) in Å while the vertical axis shows the PMF in kJ mol\(^{-1}\). The black line represents the average PMF and the gray region is the standard error. The dotted line divides the vertical axis into positive and negative regions. The labels (A), (B), and (C) indicate approximate values of \( \xi_1 \) for the configurational snapshots.

At long separation distances the PMF is relatively flat indicating that the fab and lysozyme do not interact. As the two molecules approach each other a barrier appears at a separation distance of about 55 Å. This barrier is a combination of solvation effects and entropy loss as the two proteins approach one another. It smoothly increases until it reaches a maximum height of 9.9 kJ mol\(^{-1}\) when the separation distance is about 20 Å. Immediately after the peak the PMF drops down into a single free energy well with a depth of -18.7 kJ mol\(^{-1}\) that corresponds to the bound state. This minimum occurs at a separation distance of 14.9 Å.

Analysis of the trajectory files provided a molecular-view of the dynamic motion of the fab-lysozyme system. The proteins’ behavior can be divided into three states. The first state, (A), is the bound state and is essentially the same as the crystal structure. The second state, (B), is a partially bound configuration in which lysozyme can rotate in the plane parallel to, but offset from, the best-fit plane containing the ABS contacts in the fab. Although lysozyme can rotate a full 360°, the rotation is hindered by energy barriers arising from excluded-volume interactions between the two molecules. The partial binding (state (B)) is mediated by a group of three heavy chain contacts that include LEU 263B of the fab and TYR 53, SER 81, and LEU 84 of lysozyme. These contacts are located close together and create a pseudo-axis around which lysozyme rotates. State (B) occurs often at separation distances of 22 to 24 Å. The final state, (C), is an unbound state observed at increasing frequency as the separation distance increases. In state (C) lysozyme is able to rotate and move freely relative to the fab.
Figure 5.3: Fab-lysozyme PMF for bulk solution. The black line represents the average PMF, the gray region represents the standard error, and the dotted line divides the vertical axis into positive and negative regions. The labels correspond to the configurational snapshots. Images (A), (B), and (C) show representative snapshots of the proteins’ behavior at short, intermediate, and long separation distances, respectively.

5.3.2 Interactions on a surface, 1D PMF results

Initially, surface-bound simulations were run with the same umbrella setup used for the control (only the fab-lysozyme separation distance was controlled). However, unlike the control, PMFs from replicate runs never seemed to converge to a single curve. It was hypothesized that the differing behavior of the replicates indicated that multiple pathways were available for the system to follow as lysozyme bound to the fab. To separate the different binding mechanisms a second reaction coordinate, $\xi_2$, was introduced to separate cases where lysozyme is close to the surface as
it approaches the fab from cases where lysozyme approaches from the bulk as it binds. The results of the 2D umbrella simulations used to test this hypothesis are now described.

5.3.3 Interactions on a purely-repulsive surface

Figure 5.4 shows the 2D PMFs for the fab-lysozyme system where the fab is tethered to a purely-repulsive surface. Panel (A) represents the fab in the upright tethering configuration and (B) the fab in the flat configuration. The color bar on the right provides the PMF values in kJ mol\(^{-1}\). Lines and dots represent potential paths that lysozyme could travel to bind to the fab and are presented to guide the discussion. Panels (C) and (D) contain configurational snapshots corresponding to the labeled points in Panels (A) and (B), respectively. The purely-repulsive surface serves to separate the excluded-volume surface effects from other surface effects to be discussed later.

In Panel (A), the PMF is flat (Point 1) until lysozyme either comes close to the surface or the fab. As lysozyme approaches the fab a barrier is encountered in the region of \(\xi_1 = 18\) to 40 Å (see Points 3 and 5). The barrier is relatively flat with a value of 4.4 kJ mol\(^{-1}\) in this region. When \(\xi_2 < 15\) Å, the barrier is very high because both molecules experience excluded-volume repulsion from the surface. After traversing this barrier lysozyme can fall into the bound state potential well (Point 6) where the minimum is -23.6 kJ mol\(^{-1}\). For the flat case (Panel B), the PMF is flat (and close to zero) for most combinations of \(\xi_1\) and \(\xi_2\), but unlike the upright configuration, no barrier exists as lysozyme lines up to bind to the fab (Points 2 and 4). Also for this case, the bound state comprises only a small range of \(\xi_2\) values (due to the limited movement of the fab), compared to the upright configuration, but the depth of the bound-state well is stronger with a free energy of -28.6 kJ mol\(^{-1}\) (see Point 5).

Panel (C) shows snapshots of the system along the example binding pathways indicated in Panel (A). One pathway starts lysozyme close to the surface while the other places lysozyme farther away. Path (1,2,3,6) shows lysozyme approaching the fab while close to the surface. At Point (2) lysozyme must begin to move away from the surface as it continues approaching the fab to align its binding site with that of the ABS. As seen in Snapshots (1) and (6), the fab prefers to stand upright to minimize contact with the surface; however, when lysozyme is sufficiently close, the fab bends over to interact (Point 2). This occurs when \(\xi_1\) ranges from approximately 18 to
40 Å (See Panel A) and likely helps encourage lysozyme to move away from the surface. Once lysozyme reaches Point (3) it can easily move between several values of $\xi_2$ while interacting with the fab. That is, the partially-bound complex is free to rotate through space and position itself in both upright and flat positions. As lysozyme completes binding with the fab, Point (6), the fab-lysozyme complex prefers to stand upright again to minimize contact with the surface with the lowest energy well of the PMF extending from $\xi_2 = 75-85$ Å.

Path (4,5,6) shows another possible binding mechanism where lysozyme starts far from the surface such that the lysozyme-surface interactions are negligible. This pathway shows how lysozyme can approach the fab in a way that allows the fab to stay upright during the process (Snapshots 4 and 5). When lysozyme approaches this way both it and the fab are able to avoid clashing with the surface, which helps the two molecules align themselves properly for binding.

Panel (D) shows snapshots from simulations with the fab in the flat orientation. Path (1,2,5) shows that lysozyme does not have to change its height above the surface as it approaches the fab until it reaches Point (2), at which time it must move away from the surface to fully connect with the fab. Comparing Snapshots (2) and (5) shows that lysozyme does not have to move far from the surface; however, the limited mobility of the fab means that it can do little to change the Z-coordinate of its ABS to accommodate lysozyme at a different Z-coordinate. In addition, the non-ABS (orange) side of the fab lifts off of the surface, to minimize contact, and causes the ABS to be angled toward the surface.

Path (3,4,5) shows that when lysozyme comes in from a high Z-coordinate it must descend towards the surface to reach the ABS. Unlike the previously-mentioned pathways, lysozyme binding along this pathway begins with formation of light-chain contacts with the fab rather than the heavy-chain contacts as seen in bulk solution. Here, the main interacting residues include ASN 31A, TYR 32A, and TRP 91A of the fab and ARG 45, THR 47, ASP 48, ARG 68, and PRO 70 of lysozyme. This is shown in Snapshot (4). By interacting with light chain contacts first, lysozyme can minimize its contact with the surface until the ABS can draw it down further and the final bound state, Point (5), is achieved.
Figure 5.4: Panels (A) and (B) show the PMFs for the fab-lysozyme system with the fab in the upright and flat configurations, respectively, on a purely-repulsive surface. Panels (C) and (D) show configurational snapshots and binding mechanisms for the upright and flat cases, respectively.
5.3.4 Interactions on a hydrophobic surface

The 2D PMFs and binding pathways for the fab-lysozyme system on a hydrophobic surface are found in Figure 5.5. Again, Panel (A) represents the fab in the upright configuration and (B) the fab in the flat configuration. Notice that the free energy landscape, and thus binding pathways, are different than those for a purely-repulsive surface. This is due to the attractive nature of the former. One new feature found in both upright and flat orientations is the presence of a low potential energy trough located near the surface at $\xi_2 = 25-30 \, \text{Å}$ for most values of $\xi_1$. In Panel (A) the deepest part of this new low energy trough is located between Points (1) and (2) where the PMF = -12.3 kJ mol$^{-1}$. Compared to Point (2), the bound state has a free energy of -29.1 kJ mol$^{-1}$ when $\xi_2 = 70 \, \text{Å}$ (Point (7)) and -10.9 kJ mol$^{-1}$ when $\xi_2 = 28.5 \, \text{Å}$ (Point (3)). These two bound states, the former upright and the latter flat, are separated by a barrier at $\xi_2 \approx 35 \, \text{Å}$. The main barrier to binding is located from $\xi_1 = 15$ to 40 Å and $\xi_2 = 30$ to 70 Å and has a maximum height of 18.9 kJ mol$^{-1}$. As $\xi_2$ increases this barrier significantly diminishes to a height of only 2.5 kJ mol$^{-1}$ near Point (6).

In Panel (B) the PMF also contains a low energy trough at $\xi_2 \approx 27.5 \, \text{Å}$ with the deepest point being -16.2 kJ mol$^{-1}$ located near Point (2). In comparison, the deepest part of the bound state, Point (5), is only somewhat deeper at -22.6 kJ mol$^{-1}$. No barrier must be crossed for lysozyme to reach the bound state as the low energy trough along the surface can guide lysozyme directly to the ABS.

Two potential binding pathways, with representative snapshots taken from the simulations, are highlighted in Panel (C) of Figure 5.5. If lysozyme approaches the fab while close to the surface it could follow path (1,2,3,7). This pathway allows lysozyme to stay within the low energy trough produced by the surface, bind with the fab, and then rotate to an upright position by crossing the barrier at $\xi_2 \approx 35 \, \text{Å}$. As shown in Snapshots (1), (4), and (7), the fab prefers to stand upright and away from the surface with the non-ABS domain (orange) staying in contact with the surface. To bind with a surface-adsorbed lysozyme the fab must bend over, which moves it away from this preferred position. Moreover, once bound in the flat position, the attractive nature of the hydrophobic surface inhibits rotating the complex to the upright position, which is the cause of the 6.5 kJ mol$^{-1}$ barrier located at $\xi_1 = 15 \, \text{Å}$ and $\xi_2 \approx 35 \, \text{Å}$ that must be traversed to move from Point (3) to Point (7). A different path, (1,2,4,6,7), allows a free lysozyme molecule to start close to the surface until it reaches Point (2), where, if it escapes the low energy trough, it can go around the
barrier found at lower $\xi_2$ values and join path (5,6,7). Once at Point (6) lysozyme is able to drop directly into the most stable bound state, located at Point (7), while crossing over a more-shallow barrier.

In Panel (D), which shows representative snapshots for the flat case corresponding to Panel (B), only a single binding pathway is apparent. The low energy trough causes both the fab and lysozyme to be held close to the surface, so that lysozyme must follow path (1,2,3) to reach the fab. Point (2) is a local minima that occurs because it is possible for lysozyme to feel attracted to both the surface and the fab at the same time without actually binding. The depth of this trough likely makes it difficult for lysozyme to approach the fab and reach the correct bound state. Instead, it raises the possibility that lysozyme becomes stuck in a meta-stable configuration that makes it appear that a binding event has taken place when it has not. This is supported by the standard error of the free energies for these two states ($\pm1.7$ kJ mol$^{-1}$ for the low energy trough and $\pm2.4$ kJ mol$^{-1}$ for the correct bound state), which shows that, statistically, they have very similar stabilities. However, if lysozyme continues to approach the fab, it can become bound (Point (3)) without encountering an energetic barrier.

### 5.3.5 Interactions on a hydrophilic surface

The 2D PMFs for the fab-lysozyme system on a hydrophilic surface are shown in Figure 5.6. Panel (A) is for the fab in the upright tethering configuration and Panel (B) the fab in the flat configuration. An immediate difference between the hydrophobic and hydrophilic surfaces is the lack of a low-energy trough near the surface for the latter. Instead, the hydrophilic surface repels lysozyme, similar to the purely-repulsive surface but the effect is more long-ranged and extends to a further distance away from the substrate. In Panel (A), a single energetically favorable well is present that represents the bound state with a minimum of -27.7 kJ mol$^{-1}$ located at Point (5). A barrier is present at $\xi_2 \leq 75$ Å; however, in the region of $\xi_2 \geq 75$ Å (Point (4)) lysozyme can reach the fab without traversing any barrier.

In Panel (B), the flat configuration also lacks a low energy trough running along the surface and again repels lysozyme. There is also a small barrier just before reaching the bound state. The potential energy well that corresponds to the bound state encloses a small range of $\xi_1$ and $\xi_2$ values.
Figure 5.5: Panels (A) and (B) show the PMFs for the fab-lysozyme system with the fab in the upright and flat configurations, respectively, on a hydrophobic surface. Panels (C) and (D) show configurational snapshots and binding mechanisms for the upright and flat cases, respectively.
and only has a depth of -11.9 kJ mol$^{-1}$, which is the least deep of all of the bound states previously discussed. In this treatment, lysozyme has no choice but to approach the fab from bulk solution.

Panel (C) shows representative snapshots corresponding to the data in Panel (A). In path (1,2,4,5) lysozyme starts off close to the surface and far from the fab. When lysozyme begins relatively close to the surface, and starts its approach to the fab, it reaches Point (2) and encounters a substantial barrier. This forces lysozyme further away from the surface ($\xi_2$ increases) as it continues to reduce its distance from the binding site ($\xi_1$ decreases) until the system reaches Point (4) from which it easily proceeds to the bound state located at Point (5). Snapshot (2) shows that the reason why lysozyme must move upward is that the fab is unwilling to bend over to interact with a lysozyme molecule close to the surface unlike it does on the hydrophobic surface (See Figure 5.5, Panel (C), Snapshots (2) and (3)). Path (3,4,5) shows that if lysozyme approaches the fab while far from the surface that it can reach the bound state without encountering any barrier. By moving above the fab at Point (4) lysozyme is able to accommodate the fab’s reluctance to bend over towards the surface. All of the snapshots show that the fab remains in an upright position and, unlike for the hydrophobic case, extended away from the surface.

In Panel (D), only a single binding pathway is shown. Following path (1,2,3), lysozyme must descend from bulk solution to reach the bound state. Despite Path (1,2,3) being energetically favorable, Snapshots (1) and (2) show that the fab is also greatly affected by the surface when tethered in the flat orientation. Specifically, the portion of the fab not tethered (the orange domain) significantly lifts off of the surface. This causes the ABS to be pointed down toward the surface rather than to the bulk, as is optimal for binding. At this angle, it is difficult for lysozyme to find the ABS which accounts for some of the reason why the binding strength in this treatment is so low.

### 5.3.6 Binding energies

Tables 5.1 and 5.2 provide a summary of the binding behavior of the control and the treatments. Specifically, Table 5.1 contains the height of the energetic barriers and Table 5.2 contains the free energy of the most stable bound state observed in Figures 5.3, 5.4, 5.5, and 5.6. These two properties influence the kinetics and thermodynamics of the binding process, respectively. Transition state theory states that a higher barrier will slow down the binding process while a more
Figure 5.6: Panels (A) and (B) show the PMFs for the fab-lysozyme system with the fab in the upright and flat configurations, respectively, on a hydrophilic surface. Panels (C) and (D) show configurational snapshots and binding mechanisms for the upright and flat cases, respectively.
negative free energy means the complex is more stable. Understanding this allows comparisons to be made between each treatment and the control to predict the best possible combination of tether orientation and surface hydrophobicity to maximize the fab’s binding ability.

The average values listed in Tables 5.1 and 5.2 were calculated as follows. Averages for the binding strengths were calculated by taking the mean of the minimum values of each replicate for each treatment. For the control, the barrier energy average was found by taking the mean of the maximum barrier height from the three replicates. For treatments, the barrier for each PMF was calculated by taking a 2D mean of the area representing the barrier and then taking the mean of these values from the replicates.

Although all cases had barriers to binding in some region of phase space, only those cases where surmounting the barrier was absolutely required to enter into the optimal bound state have non-zero barrier energies listed in Table 5.1. For example, in Figure 5.6 Panel (A) a barrier is seen where $\xi_2 \leq 75$ Å and $\xi_1 = 18$ to 30 Å, but the bound state (Point (5)) can be accessed through a pathway where the PMF is flat (Point (4)). Because it is possible for lysozyme to approach the fab and become bound while avoiding the barrier, no barrier energy is listed in Table 5.1.

Table 5.1 indicates an unexpected trend in the PMF data. It shows that the bulk case has the highest entropic barrier to binding, followed by the purely-repulsive/upright treatment, the hydrophobic/upright treatment, and finally the hydrophilic/flat treatment. However, all other cases have optimal bound states that can be accessed without crossing an energy barrier. This finding is significant because it means that tethering the fab to either a hydrophobic or hydrophilic surface has the potential to improve the fab-lysozyme binding kinetics compared to bulk solution.

Table 5.2 helps clarify the trends observed in the optimal bound state configurations of the PMFs. The upright tethering configurations all have binding strengths greater than bulk solution
with the purely-repulsive, hydrophobic, and hydrophilic surfaces showing 26%, 56%, and 48% stronger binding, respectively. The binding strength trend of the flat configurations is more complicated. Compared to the bulk case, binding is 53% stronger on the purely-repulsive surface, 21% stronger for the hydrophobic surface, and 36% weaker on the hydrophilic surface. The variability likely arises from the close proximity of the surface having a greater effect on the fab-lysozyme complex than in the upright cases. In summary, the order of the binding strengths from strongest to weakest is: upright/hydrophobic, flat/purely-repulsive, upright/hydrophilic, upright/purely-repulsive, flat/hydrophobic, bulk solution, and flat/hydrophilic.

5.4 Discussion

5.4.1 Configurational effects

The first objective of this study was to provide a better understanding of how tethering a fab to a surface impacts the behavior of the fab-lysozyme system. To aid the discussion, it is helpful to separate different surface-induced effects on the behavior of each protein. The PMFs for the purely-repulsive surface demonstrate that at large fab-lysozyme separation distances ($\xi_1 \gtrapprox 60$ Å) the excluded-volume interactions between the proteins and the surface do not extend far from the surface ($\xi_2 \lesssim 12.5$ Å). This is expected because this surface only serves to prevent molecules from passing through it. For the upright fab, when lysozyme is nearby ($\xi_1 \lesssim 60$ Å) these excluded volume interactions become more significant when lysozyme is also close to the surface ($\xi_2 \lesssim 30$ Å). This happens because the tether location causes the body of the fab to clash with the surface when the fab bends over to interact with the surface-adsorbed lysozyme (see Figure 5.4, Panel (C), Snapshot (3)). In contrast, while the flat orientation generally keeps the body of the fab closer
to the surface than the upright orientation, the former setup allows the body of the fab to avoid clashing with the surface when lysozyme is nearby. As a result, steric interference with the surface is negligible above $\xi_2 \approx 12.5$ Å for all values of $\xi_1$. The results show that, for the given tether sites, excluded-volume interactions only extend about 12.5 Å above the surface.

The hydrophobic surface induces a low energy trough that affects fab-lysozyme interactions more than excluded-volume interactions. The low energy trough, located at $\xi_2 \approx 27.5$ Å, creates a driving force that can trap both molecules close to the surface. Specifically, the orange domain of the fab becomes “seated” on the surface rather than standing away from it as was seen with the purely-repulsive surface. It is much more difficult for an upright fab to bend over toward the surface to interact with a surface-adsorbed lysozyme. The low energy trough also makes it difficult for a bent-over fab-lysozyme complex to pull away from the surface and become upright to achieve the most stable bound configuration. As a result, an upright fab may not be able to adopt the proper orientation to intercept surface-adsorbed lysozyme, or, once it does, may not be able to maximize the complex stability. In contrast, the flat tethering site aligns the fab with the low energy trough so that it is in the proper position to capture surface-adsorbed lysozyme. However, the improved orientation of the fab allows for the creation of a meta-stable state in which lysozyme feels attracted to both the surface and the fab at the same time. This meta-stable state has a stability similar to the correct bound state. Such increases the possibility that, even after appropriate washing steps, a microarray spot using flat-tethered fabs could produce a false positive signal due to the meta-stable lysozyme. In summary, the hydrophobic surface can interfere with fab-lysozyme binding by restricting their ability to become properly aligned and by producing meta-stable states that trap lysozyme away from the fab.

The hydrophilic surface also shows surface effects that extend beyond excluded-volume interactions. Unlike the hydrophobic surface, there is no low energy trough present. Instead, the hydrophilic surface repels the fab and lysozyme, much like the purely-repulsive surface, except that the repulsion extends to $\xi_2 \approx 25$ Å instead of dying off at 12.5 Å. The extended reach of the repulsion significantly restricts the fab’s movement so that in the upright orientation it cannot bend over to interact with a surface adsorbed lysozyme. However, the repulsion also means that lysozyme is more likely to be found at a further distance away from the surface, placing it in a better position to be captured by the fab. The extended repulsion appears to affect the flat-
tethered fab even more than the upright fab. Because the flat tether site is located near the ABS, the untethered (orange) domain moves as far away from the surface as possible causing the ABS to point toward the surface instead of toward bulk solution. As a result, the fab is not oriented correctly to efficiently capture lysozyme molecules that have no choice but to approach the fab from bulk solution. In summary, repulsion created by the hydrophilic surface can interfere with the ability of the fab and lysozyme to align properly, but it also prevents lysozyme from being adsorbed to the surface, which helps it stay available for capture.

The results indicate that tethering fabs to a surface has the potential to increase fab-lysozyme binding strength. According to Table 5.2, all treatments except for a flat fab on a hydrophilic surface produce fab-lysozyme complexes with binding strengths greater than bulk solution. This is particularly true of upright fabs on either hydrophobic or hydrophilic surfaces, which have binding strengths 56% and 48% stronger, respectively, than the control. The binding strength is a function of the quality of the intra-protein contacts; as the contacts become more stable, the binding strength increases. The fab-lysozyme contacts are stabilized when the fab is attached to the surface because the entropy of the molecules is reduced. However, if the complex comes too close to the surface, as it does with the flat-oriented fabs, the surface interactions overcome this stabilizing effect. This suggests that attaching fabs in an upright orientation on a surface is beneficial to their performance.

Trends in the barriers also suggest that fab-lysozyme complex formation can benefit from attaching the fab to a surface. Table 5.1 shows that all treatments have lower barriers than the control. Additionally, several of the treatments have barriers at or near zero kJ mol⁻¹. These barriers arise in part from the displacement of solvent molecules attached to the ABS, but primarily from the entropic cost required to properly align the fab and lysozyme as they approach one another. They represent kinetic barriers that can slow or possibly prevent lysozyme binding. Configurations that eliminate these barriers therefore have the potential to increase binding rates, which would benefit microarray performance.

### 5.4.2 Entropic effects

The aforementioned discussion as to why tethering a protein complex to a surface has the potential to stabilize it agrees with previous studies involving biomolecule stability at surfaces. [37, 38, 102–104, 130, 131, 133] Specifically, prior work shows that surfaces entropically stabilize an
Table 5.3: Change in entropy upon binding ($\Delta S_{\text{binding}}$) in kJ mol$^{-1}$ for three fab-lysozyme systems.

<table>
<thead>
<tr>
<th></th>
<th>Bulk</th>
<th>Hydrophobic/upright</th>
<th>Hydrophilic/upright</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>99.0</td>
<td>52.9</td>
<td>55.6</td>
</tr>
</tbody>
</table>

individual protein’s structure by eliminating some of the random coil configurations that otherwise form in bulk solution. As a result, conformations that stabilize intra-protein contacts are more likely to occur, the entropic cost of protein folding is reduced, and the protein complex is stabilized relative to bulk solution. In an analogous way, attaching the fab to a surface eliminates certain configurations that do not align the binding site of lysozyme with the ABS of the fab as lysozyme approaches the fab. For example, Panel (A) of Figure 5.7 shows a case in which the ABS of the fab rotates away from lysozyme as the two molecules approach, resulting only in a collision and no binding event. However, Panel (B) shows that when the fab is tethered to a substrate, the surface eliminates this freedom of movement and keeps the fab properly oriented for lysozyme to approach and bind. Thus, the unbound, surface-tethered fab has less entropy to lose upon binding than its bulk counterpart.

This theory, that the surface reduces the entropic costs of binding compared to the bulk case, was tested by calculating the change in entropy that occurs as the fab-lysozyme complex forms for the control and the two best surface treatments. The results, listed in Table 5.3, support this theory and show that an upright fab on a hydrophobic surface has a 47% smaller entropy cost, while an upright fab on a hydrophilic surface has 44% smaller entropy cost upon binding, compared to a fab in bulk solution.

5.4.3 Implications on microarray design

The data presented above suggests an optimal design for microarrays in terms of fab placement on a surface and the type of surface used. The fab should be close enough to the surface so that it cannot reorient itself substantially as the antigen approaches but not so close that interference from the surface over-powers inter-protein contact formation. The flat configuration on either a hydrophobic or hydrophilic surface fulfills neither of these requirements leaving upright tethering as the only viable configuration. The hydrophobicity of the surface is also significant.
Figure 5.7: Theory showing how surfaces stabilize protein-protein interactions. Panel (A) represents the fab-lysozyme system in bulk solution and Panel (B) represents the system when the fab is tethered to a surface.

The attractive nature of hydrophobic surface creates an unbound, meta-stable state near the fab, which increases the likelihood of false positive binding events compared to other surface types. In addition, previous simulation studies [39, 41] have shown that hydrophobic surfaces do more to disrupt protein structure than hydrophilic surfaces. As a result, the hydrophobic surface is unable to serve as the optimal substrate for microarray applications.

Taken as a whole, the data indicate that the optimal fab-based antibody microarray design involves fabs tethered in an upright configuration to a hydrophilic surface. This setup has three distinct advantages. First, the binding energy associated with the fab-lysozyme complex, -27.7 kJ mol\(^{-1}\), is more stable than the bulk case at -18.7 kJ mol\(^{-1}\). Second, this setup eliminates the entropic barrier present in the bulk case, meaning that the binding kinetics should be faster than in bulk solution. Third, the hydrophilic surface is the least attractive to unbound antigens which reduces
the chances of surface fouling, protein denaturation, and false positive signals from microarray spots.

5.4.4 Applicability of the results

The results presented in this study offer an unprecedented understanding of the binding of antigens to surface-tethered fabs; however, more work remains to fully understand the phenomena and to determine the universality of the results. For example, this work used a coarse-grain model for both the intra- and inter-protein interactions. One result of this is that the total binding energies shown in Table 5.2 cannot be directly compared to experimental binding energies or those calculated by all-atom simulations. Such is common and expected for coarse-grain models where the degrees of freedom, and thus the entropy, are significantly reduced compared to the all-atom case. Specifically, the interaction energies between two residues comprising a contact, which are composed of several atom-atom interactions in all-atom models, collapse to a single favorable potential in the formalism of Karanicolas and Brooks. To stabilize the correctly-bound state, the energy scales for the native contacts are set to reproduce the internal energy portion of the free energy of the interaction, as reported by Miyazawa and Jernigan, [165] but the simplified entropic description remains. As a result, the absolute values of the binding energies are only a 1st approximation and are of little value alone. However, given that the energy scale is consistent among all of the simulations, comparisons between the control and the treatments are valid as are comparisons between the peaks and valleys within individual free energy landscapes. This comparative approach was used in this work.

Another consideration is the large number of charged residues typically found in antibodies. For example, prior research has shown that antibodies display preferred orientations on charged surfaces, but that this effect decreases when their environment has a high ionic strength. [128] The protonation state of charged amino acids can change with pH and the interaction between these residues, or such residues with a charged surface, can be altered by the salt concentration. This means that in situations far removed from typical biological conditions (e.g. pH ≈ 7, [Na+] < 150 mM), for which the models used in this work were parameterized, the results would likely change. However, since current microarray technologies avoid extremes in pH, salt concentrations, and
surface charges, the fact that the “natural” conditions are hard-coded into the model, and more harsh conditions cannot be simulated, is of little importance to the present study.

Additionally, as pointed out by Wei and Latour, [119] who did careful adsorption experiments with naturally-occurring amino acids and different types of surfaces, the exact chemistries involved are of less importance than the overall nature of the surface. Specifically, surfaces, regardless of their exact chemistries, can be categorized as one of three types: hydrophobic, moderately hydrophilic, or hydrophilic. Because the surface parameterization was based on these data, the model should be sufficient to capture the relevant electrostatic effects.

5.5 Conclusions

In summary, this study was done to investigate how surfaces affect fab-antigen binding to help improve antibody microarray design. The results show that the presence of a surface, the surface hydrophobicity, and the tether site location on the fab can dramatically affect binding performance even for small antigens like lysozyme. It was found that surfaces can stabilize protein-protein interactions by stabilizing inter-protein contacts and reducing the negative entropy change of complex formation. An analysis of the results shows that binding performance of the fab is best optimized by placing it in an upright orientation on a hydrophilic surface. Ultimately, this work provides hope that antibody microarray performance can be improved through rational design.
Although improvements to substrate materials and immobilization techniques for microarrays have improved their performance, but additional variables that impact antigen binding need to be studied in greater detail. In particular, the effects of various antigen characteristics on binding have received limited attention in prior studies. The present work seeks to initialize these studies by probing how antigen size, symmetry, and flexibility increases or decreases the binding performance of immobilized antibody fragments. Three hypotheses were tested: (1) \textit{large antigens experience reduced binding because of greater steric interference from the surface}, (2) \textit{multiple binding sites improve binding by drawing molecules together from larger separation distances}, and (3) \textit{antigen flexibility improves binding by reducing steric interference between the antigen and the surface.}

The hypotheses were tested by using molecular simulation to probe fab-hemagglutinin dynamics in bulk solution and on a surface when hemagglutinin was held rigid or given flexibility. Configurational sampling limitations inherent in all-atom simulations were overcome by using an advanced, coarse-grain, protein-surface model. [39, 120, 121]

The remainder of the chapter is as follows. First, the simulation methods, protocols, and models are described. Then, the results of the fab-rigid-HA system (control) simulations in bulk solution are presented. After this the results of the fab-flexible-HA system in bulk solution are presented and all of the bulk solution results are summarized. Next, the results of the fab-rigid-HA and the fab-flexible-HA systems on a surface are presented. The results consist of free energy landscapes (potentials of mean force, or PMFs) and snapshots of the fab-HA configurations. The paper finishes with a discussion of the three proposed hypotheses, considerations for making future coarse-grain models of symmetric proteins, and recommendations for antibody microarray designs.

The findings show that (1) while large antigens do suffer greater steric interference from the surface it does not negatively impact their binding compared to small antigens, (2) multiple binding sites do help bring antibody and antigen together from long distances, and (3) antigen flex-
ibility reduces steric interference with the surface but reduces antibody-antigen binding strength (compared to a rigid antigen). Ultimately, it is believed that tethering fabs upright on a hydrophilic surface remains the best configuration microarrays. It was also found that the coarse-grain model has enough specificity to detect even small asymmetries between the three binding sites.

6.1 Methods

6.1.1 Experimental design

The goal of this study was to determine how different antigen characteristics affect antibody-antigen binding on surfaces. Hemagglutinin (HA) was chosen as the antigen for this study because (1) it is comparable in size to an intact antibody (and therefore much larger than a fab), (2) HA has three symmetric binding sites, and (3) HA has significant tertiary flexibility. Additionally, influenza proteins are important targets for medical and biohazard microarrays, so understanding the behavior of HA under microarray conditions could be particularly useful. The affinity molecule for this study is the fab portion of an antibody as prior work suggests that using only the fab portion of antibodies enhances microarray performance. [31] The control consists of the fab-HA system in bulk solution where HA tertiary structure is held rigid. The treatments consist of (1) fab and HA in bulk solution where HA is allowed tertiary flexibility, (2) the fab and rigid HA on a surface, and (3) the fab and flexible HA on a surface. The surface is modeled as a flat plane with hydrophilic characteristics while the fab is tethered in an upright fashion that enables the body of the fab to stand away from the surface. No other surface types or tethering methods were investigated as these were found to be less optimal for microarrays. [41, 42]

6.1.2 Overview of hemagglutinin

Several fab-HA crystal structures exist in the Protein Data Bank (PDB). PDB ID 5UGY [6] was chosen for this study because the fab is bound to the HA1 “head” of HA. The head is composed of three identical HA1 chains referred to as HA1-1 (red), HA1-2 (blue), and HA1-3 (green), as seen in Figure 6.1. Each HA1 domain contains an antigen binding site (ABS), colored dark red, dark blue, and dark green. The “tail” and “core” of HA is composed of three identical HA2 chains
denoted as HA2-1 (pink), HA2-2 (cyan), and HA2-3 (lime). Finally, the fab is composed of a constant domain (orange) and a variable domain (red) where its ABS is located (dark red).

HA is cylindrical in shape and about 131 Å long and 63 Å in diameter. As seen in Figure 2.5, these dimensions make HA about the size of an antibody (an equilateral triangle with sides of 160 Å) and significantly larger than a fab (an ellipse with dimensions of 74 × 46 Å) or lysozyme (egg-shaped with an approximate diameter of 14 Å). If HA were attached to a viral capsid (an icosahedron with a diameter of 800 Å), the HA1 head extends perpendicular to the viral surface, making it the likely target for surface-tethered fabs. However, other crystal structures, such as PDB ID 3SDY, [168] demonstrate that other epitopes are possible.

6.1.3 Protein model

To achieve the adequate sampling of phase space needed to calculate the relevant thermodynamic properties of interest, the fab and lysozyme were modeled using the Gō-like model of Karanicolas and Brooks. This model yields protein folding mechanisms and stabilities that are in good quantitative and qualitative agreement with experimental data, but does so in a computationally permissible manner. [120, 121] In this coarse-grain formalism, each residue is represented as a single site located at the Cα position of the residue in the crystal structure, which for this study is the crystal structure 1YQV as found in the Protein Data Bank. [1]
As with all Gō-like models, the tertiary structure of the protein is maintained by defining native contacts. For this model, pairs of residues form such contacts if they comprise a hydrogen bond in the native structure or if one of the heavy atoms in one residue is found within 4.5 Å of a heavy atom of another residue. The energy of a native contact, \( V_{ij} \), takes the form:

\[
V_{ij} = \varepsilon_{ij} \left[ 13 \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 18 \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{10} + 4 \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right]
\]

Here, \( r_{ij} \) is the instantaneous distance between the two sites comprising the native contact, \( \varepsilon_{ij} \) is the interaction energy between the two residues forming the native contact and is based on the experimental energies reported by Miyazawa and Jernigan, [165] and \( \sigma_{ij} \) is the distance parameter of the native contact and is set equal to the distance between the two \( \alpha \)-carbons of the residues comprising the contact. Solvent effects are included implicitly in this potential as an energetic barrier (the 10th power term) that replicates the cost of displacing water molecules as two residues come into close contact. All other interactions, those that are not native contacts, are purely repulsive. The reader is referred to the original papers for more details on the model. [120, 121]

The biological assembly for 5UGY had to undergo two particular modifications before the final coarse-grain model was made. First, the fab bound to HA1-2 (green) and HA1-3 (blue) were deleted, leaving only the fab bound to HA1-1 (red). Second, a six-residue missing loop segment in the fab was patched using coordinates from PDB ID 4WUK (a crystal structure of just the fab found in 5UGY). The patched all-atom structure was minimized in CHARMM [169] before being submitted to the MMTSB Gō-model builder. [120, 121]

### 6.1.4 Rigid vs. flexible HA

The crystal structure of HA starts the molecule in a cylindrical shape. To prevent unrealistic denaturing of the fab or HA due to forces from the umbrella restraints, [42, 166, 167] additional harmonic restraints were placed on each of the native contacts within the fab or HA to maintain the tertiary structure of each molecule. These restraints did not include any of the 84 native contacts between the fab and the three binding sites on HA. The \( k_{spring} \) for these harmonic restraints were set to 10.0 kcal mol\(^{-1}\) for the first equilibration simulation phase, and then 1.0 kcal mol\(^{-1}\) for the second equilibration phase and the production phase of the simulations.
Initial testing of the coarse-grain model in bulk solution indicated that the HA1-1, HA1-2, and HA1-3 domains can break away from the HA2 core as seen in Figure 6.1. While there is debate whether or not the HA1 head actually opens up like that, [155] the observation of this flexibility at ambient conditions with the coarse-grain model (150 mM salt at 300 K) suggested it might be important when HA is near a surface. As such, a second set of harmonic restraints not including the native contacts that hold the HA1 domains next to the HA2 core was made to enable simulation of “flexible” HA. However, this second set of restraints did not allow the bulk of the individual HA1 domains to denature.

6.1.5 Surface model

Interactions between the protein and the surface were represented by the model of Wei and Knotts. [39] The mathematical form of the potential is shown in Equation 6.2. Here $\rho = 1.0$ Å$^{-3}$, $\sigma_i$ and $\epsilon_i$ are the model parameters for residue $i$, $z_{is}$ is the distance between residue $i$ and the surface, $\theta_1 = 0.2340$, $\theta_2 = 0.4936$, $\theta_3 = 0.1333$, $\theta_S = 0.0067$, and $\theta_P = 0.0333$. The terms $\chi_S$ and $\chi_P$ represent the hydropathy indices for the surface type and residue $i$, respectively. The hydropathy index for the hydrophilic surface is $-1.0$. The values of $\chi_P$ for each amino acid type can be found in the original reference. This surface model was designed to work with the coarse-grain protein model of Karanicolas and Brooks, and, as is shown in the original reference, demonstrates both qualitative and quantitative agreement with experimental protein adsorption studies. [23, 24, 170–173]

\[
V_{surface} = \sum_i^N \left[ \pi \rho \sigma_i^3 \epsilon_i \left[ \theta_1 \left( \frac{\sigma_i}{z_{is}} \right)^9 - \theta_2 \left( \frac{\sigma_i}{z_{is}} \right)^7 ight] + \theta_3 \left( \frac{\sigma_i}{z_{is}} \right)^3 - (\theta_S \chi_S + \theta_P \chi_P) \left( \frac{\sigma_i}{z_{is}} \right)^3 \right] \right] \tag{6.2}
\]

6.1.6 Bulk simulations

Fab-rigid-HA behavior in bulk solution was designed to be the control for comparison of fab-flexible-HA behavior or fab-HA behavior near a surface. They also served to determine if
Table 6.1: Sites used to calculate $\xi_1$ and $\xi_2$ in bulk simulations.

<table>
<thead>
<tr>
<th>Condition</th>
<th>$\xi_1$ Sites</th>
<th>$\xi_2$ Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rigid/Flex #1</td>
<td>109 631</td>
<td>109 1127</td>
</tr>
<tr>
<td>Rigid/Flex #2</td>
<td>109 631</td>
<td>109 1623</td>
</tr>
<tr>
<td>Rigid/Flex #3</td>
<td>109 1127</td>
<td>109 1623</td>
</tr>
</tbody>
</table>

the fab showed a preference for any of the three HA1 domains. The sequences of the HA1 and HA2 chains as well as their crystal structures indicate that HA should be behave as a perfectly symmetrical protein. However, analysis of the biological assembly (see Table 6.2) shows that the crystal structure is sufficiently asymmetrical such that fab preference should be tested. As a result, bulk simulations involved three sets of two reaction coordinates where $\xi_1$ was the distance between the fab and one ABS and $\xi_2$ was the distance between the fab and one of the other two ABSs. The sites used to calculate these values are listed in Table 6.1 and visually shown in Figure 6.2. Site 109 is TYR A 109 in the fab while sites 631, 1127, and 1623 are LEU E 191, LEU G 191, and LEU I 191 in HA. In each simulation the equilibrium values of $\xi_1$ and $\xi_2$ were 5.0, 7.0, 9.0, ..., 65.0 Å. This setup created a total of 961 simulations for each PMF shown in Figures 6.4-6.7 plus two additional PMFs that are not shown. Both $\xi_1$ and $\xi_2$ were fixed with a harmonic restraint (see Equation 6.3) where $k_{spring}$ value was 10.0 kcal mol$^{-1}$ for the first equilibration phase and 0.25 kcal mol$^{-1}$ for the second equilibration phase and the production phase.

To help elucidate the asymmetry in HA, the excluded volumes of the HA1 domains were calculated. These values are listed in Table 6.3. They were calculated by summing up the non-overlapping volumes of the residues in each HA1 domain that are closest to the ABSs (residues towards the tail of HA were excluded). The volume of each residue was determined by using the nonbonded, non-native contact radius of the residue to calculate the volume of the corresponding sphere. The sites used for each HA1 domain volume calculation were 485-700 (HA1-1), 981-1196 (HA1-2), 1477-1692 (HA1-3).

$$U_{\xi} = \frac{1}{2}k_{spring}(\xi_i - \xi_{i,0})^2$$  \hspace{1cm} (6.3)
6.1.7 Surface simulations

Simulations involving a surface were designed to determine how height above the surface as well as the angle of HA relative to the surface would impact fab-HA interactions. As a result, treatment simulations also involved two reaction coordinates. However, as described in Section 6.2.5, it was assumed that only $\xi_1$ would go between the fab and HA. Specifically, $\xi_1$ was the distance between the fab and HA1-1. The second reaction coordinate, $\xi_2$, was set as the distance between the HA1 head and the surface. The exact sites used to calculate $\xi_1$ were TRP A 109 of the fab and LEU E 191 of HA while $\xi_2$ was calculated from the Z-coordinate of THR E 211 in
HA. The instantaneous value of $\xi_2$ was maintained using a harmonic restraint only acting on the Z-coordinate of THR E 211, so that it could otherwise move freely in the X- and Y-directions. The equilibrium values of $\xi_1$ were 5.0, 7.5, 10.0, 12.5, ..., 110.0 Å and the equilibrium values of $\xi_2$ were 20.0, 21.5, 23.0, ..., 110.0 Å. This setup created 2,623 simulations used to generate each PMF. The $k_{spring}$ value for the restraints on $\xi_1$ and $\xi_2$ was set up the same as in bulk solution. The fab was tethered to the surface at site LYS B 223 using a harmonic restraint with $k_{spring} = 100$ kcal mol$^{-1}$ such that LYS B 223 would be held at the coordinate (0,0,5.8) Å.

The cylindrical nature of HA and the fact that the ABSs are located at one end of the molecule suggested that the binding behavior would likely be impacted by the angle of HA relative to the surface. To probe this effect, each surface simulation set involved not only a rigid or flexible HA, but also with HA held at a different angle relative to the surface. This setup is depicted in Figure 6.3. The angle was maintained by applying a second Z-coordinate harmonic restraint on ASN F 128 of HA (in the tail). The angles that were tested include 0.0°, 22.5°, 45.0°, 67.5°, and 90.0°. The $k_{spring}$ value for this third restraint followed the same pattern as the restraints on $\xi_1$ and $\xi_2$. This setup allowed HA to move freely in the X- and Y-directions and to rotate freely about its cylindrical axis while holding the HA1 head at a specified distance from the surface and HA at a specified angle relative to the surface such that the HA1 head was pointed toward the surface.


6.1.8 Simulation protocols

Each individual simulation consisted of three phases. The first equilibration phase served to position each molecule so that $\xi_1$, $\xi_2$, and the HA angle were near their intended equilibrium values. This phase was run under the NVE ensemble for $5 \times 10^5$ time steps. The second equilibration phase allowed the molecular positions and structures to relax in preparation for production steps. It was run in the NVT ensemble using the Nosé-Hoover Chain integration method [159–161] involving three thermostats of mass $10^{-26}$ kg Å$^2$ for $1 \times 10^6$ time steps. The last segment was the production phase, from which data were taken for analysis. This was also NVT with Nosé-Hoover Chain integration and consisted of $5 \times 10^6$ time steps. All time steps were 0.002 fs.

6.1.9 PMF calculations and visualization

The PMFs were calculated from the production data using the Weighted Histogram Analysis Method (WHAM) [103]. Because this method produces the PMF to within an arbitrary, additive constant for all values of $\xi_1$ and $\xi_2$, comparisons between different simulated conditions were made easier by shifting each energy landscape to a value of zero where protein-protein and protein-surface interactions are negligible. This occurs at large values of $\xi_1$ and $\xi_2$. As a result, non-interacting regions have PMF values near zero, energetic barriers have positive PMF values, and energetically favorable regions have negative PMF values. In Figures 6.4-6.8 and 6.10, non-interacting regions are green, energetic barriers are yellow to red, and low energy regions are cyan to blue.

Numerous figures representing the structural configurations of the molecules were made to visualize the proteins’ behavior at points of interest on each PMF. Each protein image was generated using Visual Molecular Dynamics (VMD) [174] (found at http://www.ks.uiuc.edu/Research/vmd/) and its internal Tachyon ray tracing engine.

6.1.10 Configurational sampling

The simulation protocols were designed to ensure that sufficient sampling of molecular configurations had been achieved to provide reasonably accurate values for binding energies and energetic barriers. The number of necessary equilibration steps was determined by tracking the
potential energy of the fab-HA system and finding the minimum number of steps required for the potential energy to level off. This occurred, for several \((\xi_1, \xi_2)\) combinations, within \(1 \times 10^6\) time steps. The number of necessary production steps was determined by first calculating the PMF after \(50 \times 10^6\) production steps for a single case, and then comparing this PMF to PMFs calculated after \(1, 2, 3, 4,\) and \(5 \times 10^6\) production steps. It was found that \(5 \times 10^6\) steps produced essentially the same PMF as \(50 \times 10^6\) steps, so this value was used for all of the remaining PMFs. Because of the large number of simulations required for each control/treatment, only a single replicate of each case was performed, and, as a result the binding energies and energetic barrier values do not have confidence intervals. Despite this, it is anticipated that the PMFs are accurate enough to enable a discussion of the hypotheses mentioned above.

6.2 Results

6.2.1 HA asymmetry and bulk solution PMFs

The initial task was to test whether or not the fab would show a preference for any of the three ABSs available to it. The amino acid sequences and the general shape of the crystal structure of HA indicate that it forms a perfectly symmetrical molecule. However, a closer inspection of the crystal structure reveals asymmetry, possibly a result of imperfect fitting of the atomic coordinates to the electron density maps from the X-ray crystallography process. The root mean square deviation (RMSD) of HA1-2 and HA1-3 with respect to HA1-1 and HA2-2 and HA2-3 with respect to HA2-1 are listed in Table 6.2 and quantify the extent of the asymmetry in the crystal structure. The RMSD values are non-zero, meaning that the three fabs (in the original biological assembly) and the three binding sites are not exactly identical. As a result, the coarse-grain model should also exhibit this asymmetry since the native contacts are designed to maintain the original starting structure of the molecule. As such, it was expected that the fab would show a preference for its original ABS (HA1-1) over the other two ABSs (HA1-2 or HA1-3).

The impact of the asymmetry in HA on the fab binding behavior was investigated by simulating the molecules in bulk solution and calculating the free energy of the fab-HA system in relation to two fab-ABS distances \((\xi_1, \xi_2)\). Figures 6.4-6.7 show the results of these simulation sets. In these figures Panel (a) shows the reaction coordinate setup, Panel (b) shows the PMF in
Table 6.2: RMSD values for the HA1 and HA2 domains in HA.

<table>
<thead>
<tr>
<th>HA Domains</th>
<th>RMSD (Å²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1 ↔ 1-2</td>
<td>3.61</td>
</tr>
<tr>
<td>1-1 ↔ 1-3</td>
<td>4.79</td>
</tr>
<tr>
<td>2-1 ↔ 2-2</td>
<td>3.78</td>
</tr>
<tr>
<td>2-1 ↔ 2-3</td>
<td>9.84</td>
</tr>
</tbody>
</table>

kJ mol⁻¹ overlaid with potential binding pathways, and Panel (c) shows representative configurational snapshots of the molecules’ behavior at each numbered point on the binding pathways. The two-way arrows in Panel (c) indicate the anticipated preference for the direction of the binding reactions. The dotted line shows least preference, the solid line moderate preference, and the heavy solid line most preference. The fab is colored so that the variable domains (V_L and V_H) are red, the constant domains (C_H1 and C_L) are orange, and white represents connecting loops. For HA, HA1-1 is colored red, HA1-2 is blue, HA1-3 is green, while HA2-1, HA1-2, and HA2-3 are grey.

### 6.2.2 Rigid HA in bulk solution

Figure 6.4 represents the case where \( \xi_1 \) is the distance between the fab and HA1-1 and \( \xi_2 \) is the distance between the fab and HA1-2 (referred to as “Rigid #1”). The PMF shows that at long separation distances (Point (1)) the fab and HA are free to move independent of one another. As \( \xi_1 \) and \( \xi_2 \) decrease (the fab approaches HA), the fab first finds the ABS of HA1-3. This configuration (Point (2)) represents a local minimum with a binding energy of -4.8 kJ mol⁻¹. Once it reaches this point the fab could remain bound to HA1-3, break away from HA entirely, or move around HA to find either HA1-1 or HA1-2. An energetic barrier is present before reaching the bound state at either Point (3) or Point (4). The height of this barrier varies from 9.6 kJ mol⁻¹ near Point (4) to 9.1 kJ mol⁻¹ near Point (3). If the fab manages to cross either of these barriers and reach Point (3) (bound to HA1-1) or Point (4) (bound to HA1-2), then the system will achieve a free energy lower than at Point (2). Specifically, the binding energies for HA1-1 and HA1-2 are -16.8 and -15.4 kJ mol⁻¹, respectively. Because the energetic barriers are similar in height and the two energetic wells are similar in depth, it appears that under these conditions the fab has a greater, but similar, preference for HA1-1 and HA1-2 and least preference for HA1-3.
Figure 6.4: Bulk solution PMF for the fab-HA system for the control Rigid #1. Panel (a) depicts the reaction coordinate setup, Panel (b) is the PMF overlaid with potential binding pathways, and Panel (c) shows representative configurations along the binding pathways.

Figure 6.5 represents the case where $\xi_1$ is again the distance between the fab and HA1-1 but where $\xi_2$ is the distance between the fab and HA1-3 (“Rigid #2”). As before, starting at Point (1) the fab and HA can move largely independent of one another. However, as the fab approaches HA under these conditions, it sees two local minima at intermediate $\xi_1$ and $\xi_2$ distances, instead of the single minimum seen in Figure 6.4. Because the PMF is flat around both of these minima, it is possible for the fab-HA system to interchange between the two configurations with relative ease.

The two local minima exist because of the asymmetry in HA. Point (2) has a free energy of -3.1 kJ mol$^{-1}$ while at Point (3) it is -5.5 kJ mol$^{-1}$. Because Point (3) has a slightly lower free energy it should be slightly preferable to Point (2). As shown by Configuration (3) in Panel (c), Point (3)
Table 6.3: Excluded volumes of the HA1-1, HA1-2, and HA1-3 domains in Å³.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Excluded Volume (Å³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HA1-1</td>
<td>7321</td>
</tr>
<tr>
<td>HA1-2</td>
<td>6944</td>
</tr>
<tr>
<td>HA1-3</td>
<td>7022</td>
</tr>
</tbody>
</table>

on the PMF arises from the variable domain of the fab fitting into the crevice formed where the HA1-2 and HA1-3 domains meet. In this configuration the fab feels attracted to both HA1-2 and HA1-3, but is not bound to either ABS. Although this position does not maximize the binding force from either HA1-2 or HA1-3, the combined partial attractions cause this configuration to have a slightly lower free energy than when the fab only, but more-directly, sees HA1-2. The fab fits into this crevice because the excluded volume for the HA1-2/HA1-3 combination is smaller than the excluded volumes for either the HA1-1/HA1-2 or HA1-1/HA1-3 combinations. This can be seen in the excluded volumes of each HA1 domain listed in Table 6.3. The asymmetry results in the appearance of a second local minimum that is offset from where \( \xi_1 \approx \xi_2 \). Additionally, the large excluded volume of the HA1-1 domain means that the fab is unable to fit into the crevices formed by the HA1-1/HA1-2 or the HA1-1/HA1-3 domains so that no other unbound local minima exist in the PMF.

After reaching either Points (2) or (3) the fab can find other configurations with lower energy states, after overcoming energetic barriers. From Point (3) the fab can overcome a barrier with a height of 14.7 kJ mol\(^{-1}\) and bind fully to HA1-3 (Point (5)) where the free energy is -9.1 kJ mol\(^{-1}\). However, if the fab moves the other direction, towards Point (2), it can then find its way over to HA1-1, which presents both a lower energetic barrier (10.6 kJ mol\(^{-1}\)) and a more-favorable bound state at Point (4) where the free energy is -14.2 kJ mol\(^{-1}\). Thus, unlike the Rigid #1 case, in the Rigid #2 case the fab has a noticeable preference for HA1-1 over HA1-3 and a significant preference for HA1-1 over HA1-2.

Figure 6.6 depicts the case where \( \xi_1 \) is the distance between the fab and HA1-2 and \( \xi_2 \) is the distance between the fab and HA1-3 (“Rigid #3”). This configuration displays the greatest homogeneity in the binding behavior of the three ABSs. Starting at Point (1), the fab first finds the ABS to HA1-1 (Point (2)), which presents a significant local minimum in the free energy of -9.9
Figure 6.5: Bulk solution PMF for the fab-HA system for the control Rigid #2. Panel (a) depicts the reaction coordinate setup, Panel (b) is the PMF overlaid with potential binding pathways, and Panel (c) shows representative configurations along the binding pathways.

While this state is quite stable, the fab can rearrange itself around HA to bind to either HA1-2 (Point (3)) or HA1-3 (Point (4)), and thereby achieve a lower free energy of either -13.3 kJ mol\(^{-1}\) or -13.1 kJ mol\(^{-1}\), respectively. In addition, HA1-2 and HA1-3 also present similar energetic barriers (11.1 kJ mol\(^{-1}\) or 11.2 kJ mol\(^{-1}\)). As a result, under these conditions, the fab has a similar preference for all three HA1 domains.

6.2.3 Flexible HA in bulk solution

After treating HA as a rigid cylinder, bulk simulations were run where the HA1 domains were given flexibility. These new simulation sets are denoted Flex #1, Flex #2, and Flex #3 because
Figure 6.6: Bulk solution PMF for the fab-HA system for the control Rigid #3. Panel (a) depicts the reaction coordinate setup, Panel (b) is the PMF overlaid with potential binding pathways, and Panel (c) shows representative configurations along the binding pathways.

they have the same reaction coordinate setups as Rigid #1, Rigid #2, and Rigid #3, respectively. This flexibility was investigated because initial testing of the HA coarse-grain model indicated that at typical conditions (150 mM salt at 300 K) these domains displayed a tendency to temporarily break away from the HA2 core. While the HA1 domains do separate from the core of the molecule, they do not denature themselves. It was hypothesized that this additional flexibility could significantly impact the binding behavior of the fab. However, the results of these simulations in most cases showed that the added flexibility only marginally changed the binding behavior from that which was observed in the control PMFs with rigid HA.
There were, however, two noteworthy changes to the binding behavior of the fab-HA system caused by the added flexibility to HA. The first change was a general reduction (6-12%) in the binding strength to each HA1 domain relative to the binding strengths seen with rigid HA. The decreased binding strength results from the additional entropy of the system because it is harder for the fab and any given HA1 domain to align themselves properly for binding.

The second noteworthy change in the binding behavior was the effective elimination of the local minimum at Point (3) seen in the Rigid #2 case (see Panel (b) in Figure 6.5). Figure 6.7 shows the case where $\xi_1$ is the distance between the fab and HA1-1 and $\xi_2$ is the distance between the fab and HA1-3. When approaching HA from long distances (Point (1)), the fab still encounters two local minima. However, this time Point (2), where the fab is bound to HA1-2, is slightly dominant compared to Point (3), because the free energies for these configurations are -3.5 and -1.7 kJ mol$^{-1}$, respectively. Because Point (3) is shallower in the Flex #2 case compared to the Rigid #2 case, it appears that making HA flexible reduces the ability of the HA1-2 and HA1-3 domains to stay within close proximity to one another. This then diminishes the chances that the fab can feel partially attract to both ABSs at the same time. The final result is that it is more likely that the fab-HA system will travel to Point (2) before moving to either Point (4) or Point (5). If the fab overcomes the energetic barrier near Point (5) (13.5 kJ mol$^{-1}$), the system can achieve a free energy of -9.1 kJ mol$^{-1}$. However, as was the case for the Rigid #2 case, the most stable state of the system is found at Point (4) where the fab is bound to HA1-1 with a free energy of -13.5 kJ mol$^{-1}$ after overcoming a barrier of 10.1 kJ mol$^{-1}$.

Although Panel (a) of Figure 6.7 depicts HA in an “open” configuration, all of the HA configurations shown in Panel (c) match the rigid HA case. An open configuration is defined as when either the HA1-1$\leftrightarrow$HA1-2, HA1-1$\leftrightarrow$HA1-3, or the HA1-2$\leftrightarrow$HA1-3 distances (determined by the site 631-1127, site 631-1623, or site 1127-1623 distances) were greater than 50 Å. This cutoff was chosen because in the rigid configuration these distances were all approximately 40 Å. Although a set of tertiary structure restraints were designed to allow HA to be flexible, the molecule was not frequently found in that configuration. This meant that open HA configurations were not available in the trajectories used to depict the representative behaviors of the fab and HA in Figure 6.7. Summing the number of trajectory frames with open HA configurations over all values of $\xi_1$ and $\xi_2$ demonstrates that open HA configurations can be found. However, only 6-9%
Figure 6.7: Bulk solution PMF for the fab-HA system for the treatment Flex #2. Panel (a) is the reaction coordinate setup, Panel (b) is the PMF and potential binding pathways, and Panel (c) shows representative configurations along the binding pathways.

of all trajectory frames across Flex #1, Flex #2, and Flex #3 had HA in an open configuration. This low proportion of open HA configurations indicates that the added flexibility is not important to the binding behavior in bulk solution.

6.2.4 Bulk solution binding summary

The fact that the binding energy to each HA1 domain changes when the reaction coordinate setup changes shows that the asymmetry in HA influences the ABS preference of the fab. The binding energies for each ABS when HA is rigid are summarized in Table 6.4. Organizing the binding energies according to their HA1 domain clarifies that the fab has greatest preference for the
Table 6.4: Binding energies between the fab and each HA1 domain for each bulk case where HA is rigid.

<table>
<thead>
<tr>
<th>Control</th>
<th>Binding Energy (kJ mol⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HA1-1</td>
</tr>
<tr>
<td>Rigid #1</td>
<td>-16.8</td>
</tr>
<tr>
<td>Rigid #2</td>
<td>-14.2</td>
</tr>
<tr>
<td>Rigid #3</td>
<td>-9.9</td>
</tr>
</tbody>
</table>

HA1-1 domain, followed by the HA1-2 and HA1-3 domains, respectively. The reduced preference for the latter two domains can be tied back to the non-zero RMSDs of HA1-2 and HA1-3 relative to HA1-1 as well as the differences in excluded volumes between the three HA1 domains. These structural differences cause the fab/HA to experience a poorer fit and overall greater steric repulsion when bound HA1-2 or HA1-3. Thus, the asymmetry of the starting crystal structure influences the behavior of the derived coarse-grain model.

Introducing flexibility into HA does impact the binding strength of the fab, even if does not significantly change the overall behavior seen in the PMFs. The binding energies for flexible HA are summarized in Table 6.5. As mentioned previously, the binding energy for each HA1 domain is reduced compared to its binding energy when HA is rigid because there is a higher entropic cost to binding. This additional entropy arises from the fact that both the HA1 domain near the fab and the rest HA (which can move independently under these conditions) much orient themselves so that the two molecules can bind. Likewise, the favorable crevice between HA1-2 and HA1-3 is formed less frequently so that the extra minimum seen in Rigid #2 nearly disappears. Despite these differences, the trend of more preference for HA1-1 and less preference for HA1-2 and HA1-3, respectively, remains the same. Thus, in bulk solution, it does not appear that flexibility significantly alters the overall binding behavior.

6.2.5 Fab-HA binding assumption

Efficiently investigating how a surface impacts fab-HA binding requires the introduction of a simplifying assumption regarding the symmetry of HA. The preceding data show that the fab prefers binding to HA1-1 over the HA1-2 or HA1-3 domains. This behavior demonstrates that the Gō-like model of Karanicolis and Brooks [120, 121] can discriminate even small differences
Table 6.5: Binding energies between the fab and each HA1 domain for each 2D bulk case where HA is flexible.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Binding Energy (kJ mol(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HA1-1</td>
</tr>
<tr>
<td>Flex #1</td>
<td>-13.8</td>
</tr>
<tr>
<td>Flex #2</td>
<td>-13.5</td>
</tr>
<tr>
<td>Flex #3</td>
<td>-8.8</td>
</tr>
</tbody>
</table>

...in structure, despite representing each amino acid with only a single site. While this specificity is usually desirable, it adds an unnecessary layer of complexity to the analysis of surface treatments. However, it is expected that under typical laboratory conditions HA would act as a perfectly symmetrical molecule, i.e. the fab would bind to each HA1 domain with equal preference. As such, the remaining simulation results assume that all three HA1 binding sites are equivalent. In other words, the binding behavior is studied by assigning \(\xi_1\) as the distance between the fab and HA1-1 while it is assumed that HA1-2 or HA1-3 could be used interchangeably. Making this assumption allows the PMFs to remain two-dimensional and greatly reduces the number of simulations that are needed.

### 6.2.6 Rigid HA near a surface

The next task was to investigate how tethering the fab to a surface impacts its binding behavior with HA. For these simulations \(\xi_1\) represents the distance between the fab and HA1-1 (depicted in red) and \(\xi_2\) represents the distance between HA and the surface. The HA-surface distance was measured by the Z-coordinate of MET F 77 of HA (the surface plane is located at \(Z = 0 \text{ Å}\)). This site is situated near LEU E 191, the HA site used to calculate \(\xi_1\), but is actually located within HA2-1. This setup allows HA to rotate around its cylindrical axis, ensures that \(\xi_2\) properly represents the distance between the HA and the surface, and allows the HA1 domains to break away from the HA2 core for later simulations. The tail end of HA was fixed at a series of different Z-coordinates to hold HA at specific angles relative to the surface. The angles that were tested include 0.0°, 22.5°, 45.0°, 67.5°, and 90.0°. In all cases the fab is attached to the surface in a manner that allows the body of the fab (including the variable domains) to be directed away from the surface. The surface was configured to have hydrophilic characteristics. This “upright”
Table 6.6: Summary of binding strength as a function of HA angle relative to the surface and where HA is rigid.

<table>
<thead>
<tr>
<th>HA angle (deg)</th>
<th>0</th>
<th>22.5</th>
<th>45</th>
<th>67.5</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binding Strength (kJ mol$^{-1}$)</td>
<td>-13.6</td>
<td>-27.5</td>
<td>-26.8</td>
<td>-85.1</td>
<td>-14.1</td>
</tr>
<tr>
<td>Energy Minimum (kJ mol$^{-1}$)</td>
<td>-13.6</td>
<td>-27.5</td>
<td>-26.8</td>
<td>-85.1</td>
<td>-19.7</td>
</tr>
</tbody>
</table>

configuration on a hydrophilic surface was chosen because previous worked determined it was the best setup for microarrays. [42]

Figure 6.8 depicts the $\xi_1$ and $\xi_2$ setup for each HA angle (Panels (f) through (j)) along with the corresponding PMFs (Panels (a) through (e)). The values of $\xi_1$ and $\xi_2$ both ranged between 0 and 120 Å, but only regions where $\xi_1 < 60$ Å are shown since the most significant fab-HA behavior occurs in the area $5 \lesssim \xi_1 \lesssim 20$ Å. The lines and points on each PMF represent potential binding pathways and are intended to guide the reader’s understanding of how the binding pathways relate to the PMFs. The fab-HA configurations at each numbered point on Figure 6.8 are shown in Figure 6.9 along with reaction arrows that indicate the anticipated progression of the binding process. Note that the PMF color scale has two sets of numerical values; one set (the left side) is for HA angles 0°, 22.5°, 45°, and 90° and a second set (the right side) is for 67.5°. The PMF for 67.5° was scaled so that a value of zero has the same color as zero for all of the other PMFs. This two-scale system was required because the magnitude of the binding energy for the 67.5° treatment was significantly greater than for all other conditions. Finally, a summary of the binding energies and lowest energies for HA angle can be found in Table 6.6.

Figure 6.9 shows two potential binding pathways for the fab-HA system. The first pathway, (1,2,3,4,7), depicts HA approaching the fab while in a horizontal position, parallel to the surface. The second pathway, (5,6,7), shows HA approaching the fab in a vertical position, perpendicular to the surface. Starting at Point (1), HA maintains a distance of about 80 Å from the surface, allowing it to mitigate interactions with the surface. However, it is free to move closer or further from the surface from this point because of the low energy trough from $60 \lesssim \xi_1 \lesssim 90$ Å and $40 \lesssim \xi_2 \lesssim 60$ Å. This trough exists because, while the fab-HA1-1 distance is 45 Å, the fab-HA1-2 and fab-HA1-3 distances can be significantly shorter allowing HA to feel attracted to the fab or even become
Figure 6.8: PMFs and reaction coordinate setups for fab-rigid-HA simulations on a hydrophilic surface. Panels (a) through (e) are the PMFs and Panels (f) through (j) are the reaction coordinate setups.
bound by the fab at these ABSs. This region helps draw HA close enough to the fab so that it will be more likely to cross the energetic barrier (6.8 kJ mol\(^{-1}\)) located along 10 \(\lesssim \xi_1 \lesssim 20\) Å and become bound at \(\xi_1 \approx 7\) Å. At Point (2), HA is still horizontal and relatively close to the surface, but the system experiences a free energy of -13.6 kJ mol\(^{-1}\).

Once HA becomes bound by the fab, it is anticipated that it will remain so while changing its orientation relative to the surface to minimize the free energy of the system. Specifically, if HA rotates away from the surface to an angle of 22.5° it will reach Point (3) where the free energy is -27.5 kJ mol\(^{-1}\). At this point the free energy surface (with respect to HA angle) becomes somewhat flat so that HA is free to stay at 22.5° or to continue to rotate to 45°, since the free energy at Point (4) (-26.8 kJ mol\(^{-1}\)) is similar to Point (3). During this rotation process HA must also increase its distance from the surface by about 16 Å, which requires the fab to stretch further away from the surface also. If HA continues to rotate away from the surface to an angle of 67.5° the system will find a deep global energy minimum located at Point (7) (at approximately 80 Å above the surface). The binding energy at Point (7), -85.1 kJ mol\(^{-1}\), is noteworthy because it is more than 3× stronger than the binding energy at Points (3) and (4), and nearly 6× stronger than the best case binding
in bulk solution. There are two major reasons for this; first, an HA angle of 67.5° makes it easy for the fab to properly align itself to the ABS while optimizing its orientation and distance relative to the surface. Second, Point (7) also represents a state where the entropy of the system increases upon fab-HA binding. This happens because, when the fab and HA are close to each other, the entire complex can rotate about the Z-axis more easily than when HA is far away from the fab.

The second binding pathway in Figure 6.9, path (5,6,7), shows HA approaching the fab while in a vertical orientation. Because Point (7) represents HA in a near-vertical orientation, this path is shorter than the first. It begins at Point (5) where HA is close to the fab but still free to rotate on its own axis. This configuration allows all three ABSs to feel attracted to the fab. Point (5) actually represents the lowest energy state for a perfectly vertical HA with a free energy of -19.7 kJ mol\(^{-1}\). If HA continues to approach the fab then it must sacrifice rotational freedom and begin to change its angle to become captured by the fab (Point (6)). As shown in Configuration (6), HA is unable to remain perfectly vertical and the added strain weakens the binding strength so that it is only -14.1 kJ mol\(^{-1}\). However, once HA becomes bound it can quickly rotate toward the surface to reach Point (7). Once the system achieves Point (7), by following either path (1,2,3,4,7) or (5,6,7), the system will remain in that configuration because of its remarkably low free energy compared to all other states.

### 6.2.7 Flexible HA near a surface

The final task was to study the effects of making the HA1 domains flexible on the behavior of the fab-HA system near a surface. Except for the modified tertiary restraints, all other simulation settings were the same as those for the rigid HA near the surface. The new PMFs and binding pathways for the flexible HA are shown in Figure 6.10. Unlike Figure 6.8, the PMF color scale only has one set of values for all HA angles. The flexibility significantly affects the energetics of HA approaching the fab and the preferred final state of the system. The molecular configurations at the labeled points and the anticipated reaction direction preferences between these configurations are depicted in Figure 6.11. The binding energies and lowest energies for each HA angle treatment are listed in Table 6.7.

The first binding pathway consists of Points (1) through (6), where the flexible HA approaches the fab while in a horizontal position. Starting at Point (1), HA1-1 is far away from the
fab but HA1-2 and HA1-3 can come close to the fab. The flexibility creates difficulty for the fab to remain in close proximity to either HA1-2 or HA1-3, and as a result the low energy trough seen in Panel (a) of Figure 6.8 is no longer present. However, HA can better minimize its contact with the surface because any HA1 domain that gets too close to the surface is free to break away from the HA2 core and move away from the surface. The additional entropy within HA also increases the height of the energetic barrier located from $10 \lesssim \xi_1 \lesssim 30$ Å to 9.7 kJ mol$^{-1}$, likely making the process of reaching the bound state proceed slower than would be seen with rigid HA. However, if the system crosses the barrier the fab can fully bind to HA1-1 (Point (2)) and the system will achieve a free energy of -6.9 kJ mol$^{-1}$.

As was the case for the rigid HA, once the flexible HA is captured by the fab it will likely remain bound while HA adjusts its angle to reach the global energy minimum. This is because the bound state offers a negative free energy while becoming unbound would require the fab-HA system to both leave the low energy well and traverse the same barrier it encountered going from Point (1) to (2). As such, starting at Point (2) HA can rotate 22.5° away from the surface to reach Point (3) where the free energy drops to -12.7 kJ mol$^{-1}$. During this time HA must also move about 10 Å away from the surface. As before, HA sees a relatively flat free energy surface (as a function of HA angle and $\xi_2$) where it can stay close to Point (3) or continue to rotate and move further away from the surface towards Point (4). If HA moves another 20 Å away from the surface and rotates to 45° the system will reach Point (4) where the free energy is -13.0 kJ mol$^{-1}$. During this rotation process, a low energy trough in the region where $\xi_1 \gtrsim 30$ Å and $\xi_2 \gtrsim 100$ Å forms and becomes more prominent (See Figure 6.10, Panels (b) through (e)). Thus, if the fab-HA system were to dissociate while at higher angles, it is likely that HA would end up in this low energy trough, remain close to the fab, and more-easily be recaptured by the fab than at lower angles. This region stretches more so along $\xi_1$, while the low energy troughs in Panels (a), (b), and (c) of Figure 6.8 typically strength more so along $\xi_2$. The new shape of the troughs arises from the ability of the HA1 domains to separate from the HA2 core and to come close to the fab even when the HA1-1 domain is far away from the fab ($\xi_1 \gtrsim 45$ Å). Otherwise, when HA is rigid, there is only a narrow range of $\xi_1$ values in which HA1-2 or HA1-3 can still remain close to the fab.

As with the rigid HA, if the flexible HA rotates to 67.5°, the system can find a more-preferable energetic state than at lower angles. However, here the fab-HA behavior displays the
greatest deviation from the rigid HA. Panel (d) of Figure 6.8 shows that at 67.5° the free energy surface becomes an energy funnel that drives the fab and rigid HA together. In contrast, when HA is flexible the free energy surface no longer looks like a funnel, but instead is largely flat, as it is for 0°, 22.5°, and 45°. The bound state energy located at Point (5), -16.9kJ mol\(^{-1}\), is lower than the free energy of Points (2), (3), or (4), but not drastically different as was seen with the rigid HA at 67.5°. When HA is flexible the system can more-easily achieve the proper alignment to maximize binging at angles different from 67.5° while the entropy of the flexible HA dominates any other sources of entropy, meaning that the system has essentially the same amount of entropy whether it is bound, unbound, close to, or far away from, the surface. The result is that the global minimum of the fab-flexible-HA system moves to where HA is angled at 90.0°.

In contrast to the binding behavior seen in Figure 6.8, the global energy minimum of the fab-flexible-HA system is achieved when HA is perpendicular to the surface. In this case, located at Point (6), the energy surface becomes a funnel that drives the fab and HA together, as seen in Panels (d) and (e) of Figure 6.9. The binding strength is -27.4kJ mol\(^{-1}\). Unlike Panel (e) of Figure 6.8, only a single energy minima exists, where the system is in the bound state. Configuration (6) shows that the HA1-1 domain must break away from the HA2 core to enable the fab to properly align with it while allowing HA to remain perfectly perpendicular to the surface. Moreover, the HA1-2 and HA1-3 domains tend to adopt a similar configuration where they spread outward to minimize interactions with the surface. Given that Point (6) is the preferred state of the fab-HA system, it is possible that HA could simply approach the surface perpendicularly from bulk solution and drop directly into the bound state once the HA1 domains have separated from the HA2 core.

Table 6.8 summarizes the percentage of trajectory frames in which at least one HA1 domain had broken off of the HA2 core compared to the angle of HA above the surface. As mentioned before, simulations of the fab-HA system in bulk solution showed that only 6-9% of trajectory frames had HA in an open configuration. In contrast, when HA is near the surface the percentages increase 10× or more. In addition, as HA angle increases the percentage of frames with an open configuration also increases. This trend likely arises because of a few reasons. First, when HA is horizontal the HA1 domains closest to the surface are less likely to break away because doing so could make them clash with the surface. Second, as HA becomes more perpendicular to the surface the HA1 domains must break away for them to align themselves to the fab for binding.
Table 6.7: Summary of binding strength as a function of HA angle relative to the surface where HA is flexible.

<table>
<thead>
<tr>
<th>HA angle (deg)</th>
<th>0</th>
<th>22.5</th>
<th>45</th>
<th>67.5</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binding Strength (kJ mol(^{-1}))</td>
<td>-6.9</td>
<td>-12.7</td>
<td>-13.0</td>
<td>-16.9</td>
<td>-27.4</td>
</tr>
<tr>
<td>Energy Minimum (kJ mol(^{-1}))</td>
<td>-6.9</td>
<td>-12.7</td>
<td>-13.0</td>
<td>-16.9</td>
<td>-27.4</td>
</tr>
</tbody>
</table>

Table 6.8: The percentage of simulation trajectory frames where flexible HA was found in an open configuration near the surface.

<table>
<thead>
<tr>
<th>HA angle (deg)</th>
<th>0</th>
<th>22.5</th>
<th>45</th>
<th>67.5</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>% frames open</td>
<td>60</td>
<td>70</td>
<td>76</td>
<td>80</td>
<td>82</td>
</tr>
</tbody>
</table>

Finally, when HA is perpendicular there is more space for all three HA1 domains to break away without encountering either the surface or another protein.

6.3 Discussion

6.3.1 Impacts of antigen size

The first hypothesis proposed in the Introduction was that large antigens experience reduced binding because of greater steric interference from the surface. The results show that the large size of HA noticeably impacted fab binding on the surface. A comparison of HA behavior to lysozyme from a previous work [42] helps demonstrate the impact of antigen size. Lysozyme could approach to within about 20 Å of the hydrophilic surface while HA almost always has to remain approximately 40 to 60 Å away from the surface (see Panels (a), (b), and (e) in Figure 6.8 and Panels (a) through (e) in Figure 6.10). The small size of lysozyme allows the entire molecule to fit close to the surface so that it becomes trapped by the low energy trough created by the attractiveness of the surface. As a result, lysozyme can be stable while close to the surface and remain in a position where a fab, whose binding site is also close to the surface (such as with a flat tethering configuration [42]), could capture it. In contrast, HA’s larger size means that a significant portion of it cannot feel attracted by the surface. Instead, these parts only feel an unfavorable free energy
Figure 6.10: PMFs and reaction coordinates for each surface treatment where HA is flexible. Panels (a) through (e) are the PMFs and Panels (f) through (j) are the reaction coordinate setups.
Figure 6.11: Fab-HA configurations along the proposed binding pathways where the fab is on a surface and HA is flexible.

change because of decreased entropy while HA is close to the surface. The outcome is that HA as a whole is repelled by the surface and unable to approach the surface as closely as a small protein. It is therefore critical that the fab be tethered in an upright orientation because this design enables big antigens to avoid the surface and more likely to become bound by the fab.

Not only does the large size of HA affect its binding behavior, but so does the angle of HA relative to the surface. When HA is parallel or nearly parallel to the surface (Panels (a) and (b) of Figures 6.8 and 6.10) more of the molecule is near the surface and therefore able to clash with it. As the angle increases through 67.5°, the parts of HA not directly involved in binding become further away from the surface, reducing the undesirable interactions. The reduced tension between HA and the surface enables the HA1 domains to more-easily approach the fab. When HA is flexible it also adopts a high angle to minimize its interactions with the surface. Since being flexible provides the HA1 domains with additional freedom to align themselves with the fab, HA as a whole is free to remain perpendicular to the surface. These angles are preferred because they represent the optimal balance between forming the maximum number of undistorted contacts between the fab and HA while minimizing interactions with the surface. For rigid HA, angles
further from 67.5° result fewer fab-HA contacts that can form completely. When HA is flexible lower angles simply result in more clashing between the surface and HA, which reduces the overall stability of the fab-HA complex. The conclusion is that the antigen must be able to orient itself to both minimize steric clashes with the surface and be able to rotate its ABS towards the affinity molecules on the surface. If either of these conditions are not met, the fab-antigen binding strength will be greatly reduced. In contrast, Panel (d) of Figure 6.8 indicates that even large antigens can achieve a binding performance far greater than is possible in bulk solution, similar to the 50% increase in binding strength experienced by lysozyme under similar conditions. [42] As such, large antigen do not always experience reduced binding because of steric interference.

### 6.3.2 Impacts of binding site valency

The second hypothesis of this work was that **multiple binding sites improve binding by drawing molecules together from larger separation distances**. Despite the imperfect symmetry of HA, the availability of multiple binding sites did have discernible effects on the binding behavior. The bulk solution PMFs show that multiple binding site options allow the fab to reach a stable state with HA (see Point (2) in Figures 6.4-6.7) without having to overcome an energetic barrier first, even though both molecules have a significant amount of entropy. Even if, or perhaps especially if, the fab is loosely bound to HA, it will then be at a better starting point to then move around HA to find the preferred ABS. The result is that multiple binding sites does help bring the fab and antigen together in bulk solution.

The effect of multiple binding sites becomes even more visible for surface treatments. While the fab does show a preference for binding at HA1-1, the poor fitting with HA1-2 and HA1-3 was problematic only when the fab was very close ($\xi - 1 \approx 7 - 20$ Å). At larger separations distances ($\xi_1 \gtrsim 40$ Å) the fab could feel attracted to any of the HA1 domains without feeling significant steric repulsion. When HA is rigid the attraction from the HA1-2 and HA1-3 domains creates a low energy trough from $40 \lesssim \xi_1 \lesssim 60$ Å $50 \lesssim \xi_2 \lesssim 100$Å. Similar low energy troughs are present when HA is flexible from $\xi_1 \gtrsim 30$ Å and $\xi_2 \gtrsim 100$ Å. As with bulk solution, these troughs help bring the fab closer to HA from long distances. They can also stabilize a loose fab-HA complex when the two proteins are close to one another but not aligned favorably and need to separate some. This is especially important in situations where proper alignment may be difficult.
to achieve, such as near a surface. Because large antigens experience significant steric clashing with the surface, if they separate from the fab they may simply leave the fab altogether. However, multiple binding sites create low energy regions around the antigen that enable the two molecules to stay close to one another and have better chances for coming together properly. In conclusion, not only do multiple binding sites improve fab-antigen binding, but they may be particularly beneficial fab-antigen binding on a surface.

The presence of multiple binding sites also enabled the formation of the nonbonded, but stable state found at Point (3) in Figure 6.5. The asymmetry in the excluded volumes of the three HA1 domains created a crevice into which the fab could get close to both HA1-2 and HA1-3 without encountering the energetic barriers to binding that occur when the fab approaches an ABS straight on. The partial attraction of two ABSs can turn out to be greater than the attraction to a single ABS that displays imperfect fitting. While this issue will not arise for perfectly symmetrical antigens, it could arise in situations where immature antibodies are used as the affinity molecules. Having ABSs in close proximity to one another along with reduced complementarity of immature fabs to the epitope could create situations in which it appears that the antigen has been successfully captured when, in reality, only a non-specific bond is holding the molecules together. This could, for example, be a source of cross-reactivity between different antibodies.

### 6.3.3 Impacts of antigen flexibility

The third hypothesis was that antigen flexibility improves binding by reducing steric interference between the antigen and the surface. The results demonstrate that this hypothesis is mostly untrue as shown by the binding strengths listed in Tables 6.6 and 6.7. When HA is angled at 0°, 22.5°, or 45° the binding energy for the rigid HA is nearly double the binding energy for the flexible HA under each condition. The disparity is greatest when HA is angled at 67.5°—the binding energy of the rigid HA is 5 × greater than that of the flexible HA. Additionally, at this angle the free energy surface of the rigid HA forms a deep funnel that directs the fab and HA to bind while only a small region of $\xi_1$ and $\xi_2$ values allow for fab-HA binding when HA is flexible under similar conditions. Only when HA is held perpendicular to the surface does the flexible HA demonstrate stronger binding (-27.4 kJ mol$^{-1}$) compared to rigid HA (-14.1 kJ mol$^{-1}$). Here, the flexible HA1 domains are able to separate from the HA2 core and maintain a comfortable distance.
from the surface while the HA2 core remains perfectly perpendicular to the surface. The behavior of flexible HA at 90° is significant; however, since if HA were attached to a virus particle only the HA molecules at the bottom of the particle (those closest to and pointed perpendicular or nearly perpendicular to the surface) would be within reach of surface-tethered fabs. As a result, the flexible nature of HA may enable better binding under microarray conditions than would be the case if it were completely rigid.

Another finding from the results is that the presence of the surface does impact the frequency with which HA is found in an open configuration. The data in Table 6.8 show that HA favors the more-frequent formation of open configurations in the presence of the surface compared to bulk solution. The data also show that HA is found in an open configuration more often as the angle of HA relative to the surface increases. However, the data used to generate Table 6.8 can be rearranged so that the percentage of open HA frames is a function of $\xi_2$ (summed over all values of $\xi_1$ and all HA angles) instead of a HA angle (summed over all values of $\xi_1$ and $\xi_2$). Doing so shows how proximity to the surface changes the frequency of open HA configuration formation. This new format is shown in Figure 6.12 where dots represents the percent of trajectory frames with open HA from simulation data, the red dashed line is a parabolic trend line for the simulation data, and the blue dashed line represents the percentage of open HA frames seen in bulk solution. The data show that as HA approaches the surface it is more likely to be found in an open configuration. This trend exists because the surface hinders the movement of HA so that the HA1 domains must break away to align better with the fab. Moreover, when $\xi_2$ is small the HA1 domains not interacting directly with the fab can break away to reduce their interactions with the surface. If the trend line is followed toward increasing values of $\xi_2$, the point at which HA’s behavior in bulk solution dominates can be found. The trend line predicts that when $\xi_2 \approx 123$ Å the surface will have negligible effects on HA such that it will remain largely in its original rigid-like structure. In conclusion, molecular flexibility of the antigen becomes more important to fab-antigen binding and dynamics as the antigen approaches the surface.

6.3.4 Impacts of protein asymmetry

The results also indicate that asymmetry in the crystal structure (atomic PDB coordinates) significantly influences the behavior of coarse-grain models derived from that structure. The amino
acid sequences of the three HA1 domains and the three HA2 domains, as well as the overall structure of HA indicate that it should behave symmetrically. However, the HA1-1/HA1-2 and HA1-1/HA1-3 RMSDs for the biological assembly of 5UGY are significantly greater than zero. While changes in the position of any single the amino acid may seem insignificant, the coarse-grain model of the fab can detect these changes and displays markedly reduced binding to either HA1-2 or HA1-3 because of the poor fitting. This behavior is remarkable given that the coarse-grain protein model represents each amino acid only as a single sphere. This specificity means that coarse-grain protein models, at least those with residue-level detail, are capable of finding specific binding sites on neighboring proteins and can be used to study multi-protein interactions.

The ability of this coarse-grain model to detect small differences in protein structure suggests that additional care may be needed when developing coarse-grain models of proteins with symmetry. It may be necessary for some cases to ensure that all of the binding sites on a protein are sufficiently similar so that its binding partner will not show a preference for any particular site. For example, if an array of fabs were placed on the surface to interact with an HA molecule, all three ABSs would need to be equally accessible to any of the fabs because multiple fabs could simultaneously bind to HA. As such, efforts should be made to adjust the crystal structure so that the
RMSDs between pairs of identical subunits are close to zero before assembling the coarse-grain model.

### 6.3.5 Considerations for microarray design

The results from this study suggest that tethering fabs upright on a hydrophilic surface can work well for both small and large antigens. The hydrophilic surface displays minimal attraction to both the fab and HA, giving the proteins maximum rotational and vibrational freedom near the surface. Such freedom is crucial for large antigens because of the difficulty they can have rearranging themselves to bind to the fab. Additionally, tethering the fabs upright is mandatory for large antigens as they are unable to approach the surface as closely as small antigens. Ultimately, the fab-rigid-HA system at 67.5° indicates that a typical fab-based microarray setup is capable of displaying exceptionally strong binding to large antigens as well as small antigens. Similarly, while HA flexibility reduces the binding strength gains, a 100% increase in binding strength relative to bulk solution was still seen for the system on a surface.

The results of the rigid-HA at 67.5° suggest that the spacing between affinity molecules on the surface could be important. As microarray spots become smaller it may be tempting to make the sensing elements as dense as possible by developing ways to hold them perfectly perpendicular to the surface. However, unless the antigen has significant flexibility, such conditions (similar to Configuration (5) in Figure 6.9) could actually weaken fab binding or lead to cross-reactivity. Instead, if the fabs are allowed some freedom of movement, they will be better able to find the optimal alignment between themselves and their targets, as was seen with the rigid HA at an angle of 67.5°. This necessitates that the fab spacing be big enough to let them adapt to different angles without clashing with one another.

### 6.3.6 Conclusions

In summary, this work was done to study how different antigen characteristics affect fab-antigen binding while the fab is tethered to a surface with the intent to improve antibody microarray design principles. Specifically, this study looked at the impacts of antigen size, symmetric binding sites, and antigen flexibility. The results show that a coarse-grain model fab can discern relatively
small differences between the three binding sites on the antigen resulting from asymmetry in the original antigen crystal structure. Additionally, antigen flexibility has limited impact on the overall binding behavior seen in bulk solution. Large antigens are unable to come as close to the surface as small antigens, but they can still be captured by fabs with remarkable binding strength. Multiple binding sites do help bring fab and antigen together from long distances and provide opportunities for imperfect complexes to loosen up and find a better configuration. Antigen flexibility is particularly important near a surface because it reduces unfavorable interactions with the surface and makes it easier for the binding sites to line up with the fab. This work provides hope that antibody microarray performance can be improved through rational design.
CHAPTER 7. CONCLUSION

7.1 Review of findings

Antibody microarrays are a biotech device that has the potential to revolutionize how molecular detection is accomplished in areas such as medical diagnostics, proteomics research, and national defense. However, current iterations of these devices are unable to find mainstream use because they still exhibit poor replicability as well as low quality, unbalanced, and noisy signals coming from the microarray spots. As such, a goal of current microarray research is to optimize the technology to an extent that a single microarray can screen thousands of molecules in parallel in a rapid, inexpensive, and easy-to-use manner.

Numerous studies based on laboratory and simulation methods have lead to performance improvements, but questions still remain about individual antibody behavior on the surface. Specifically, the literature indicates three questions that could be addressed with intermediate resolution simulation methods to provide new insights into the protein-surface behaviors. These questions are: (1) how is antibody stability affected by attachment to surfaces with different properties, (2) how does surface tethering affect antibody-antigen binding, and (3) how do different antigen characteristics affect antibody binding near a surface? The purpose of the studies presented in Chapters 4 through 6 was to develop a better understanding of how the molecular-level behaviors related to these questions impact antibody microarray performance.

In Chapter 4 the stability of individual antibodies on different surfaces was investigated. The results from simulations in bulk solution indicate that the coarse-grain antibody model qualitatively reproduced the folding behavior of antibodies in aqueous solution seen in experimental studies. This set of baseline behavior was then compared to antibody behavior on hydrophobic and hydrophilic surfaces. The surface results demonstrate that attachment to the surface dramatically affects antibody stability. Specifically, it was found that antibody stability is mostly a function of the surface type rather than the protein’s orientation relative to the surface. Hydrophobic surfaces
are more likely to induce the antibody to collapse onto the surface, resulting in nearly complete denaturation of the antigen binding sites. However, when attached to a hydrophilic surface the antibody structures can remain nearly as stable as seen in bulk solution. These results suggest that antibody microarray substrates should have hydrophilic characteristics to maximize both antibody stability and function.

In Chapter 5 the mechanisms of antibody-antigen binding on different types of surfaces were investigated. Instead of looking at intact antibodies, only a fab-antigen complex (involving lysozyme) was modeled since prior studies have shown that using fabs on microarrays can produce better overall performance. The bulk solution binding behavior was first determined and then compared to different treatments where the fab was attached to hydrophobic or hydrophilic surfaces in two different orientations. The results demonstrate that lysozyme binding is dramatically impacted by both the surface type and fab orientation. Specifically, hydrophobic surfaces have the potential to trap both proteins close to the surface such that they may appear bound together without a binding event actually occurring (potentially creating false-positive signals). In contrast, the hydrophilic surface pushes the proteins away resulting in a beneficial balance of enthalpy and entropy that can strengthen proper lysozyme binding. Finally, it was shown that the fab must be tethered in an upright orientation to give it sufficient mobility to align with lysozyme and to enable lysozyme to mitigate its interactions with the surface. Ultimately, it was concluded that using a hydrophilic surface is beneficial to fab-antigen binding, but only when the fabs are oriented upright on the surface is their binding potential maximized.

In Chapter 6 the impacts of different antigen characteristics on fab-antigen binding near a surface were investigated. This was done because the nearly limitless combination of antigen properties suggests that these may markedly influence microarray performance. Specifically, this project looked at the effects of antigen size, multiple binding sites, and antigen flexibility by studying fab-hemagglutinin binding. The bulk solution results indicate that multiple binding sites help the two molecules stay close to one another as they search for their preferred binding configuration. However, neither antigen size nor flexibility majorly affect binding behavior. Near a surface, the large size of hemagglutinin makes it difficult for it to approach a surface, particularly when it is held in a rigid shape. When hemagglutinin is flexible it can more easily mitigate steric interference with the surface and align itself with the fab for capture. Binding is always stronger on the surface
than in bulk solution, but when hemagglutinin is rigid and held at an angle of 67.5° the complex is stabilized dramatically more by this configuration than any other. The results show that antigen characteristics do influence antigen binding, but suggest that using a hydrophilic surface and tethering fabs upright continues to be the best configuration for maximizing antibody microarray performance.

7.2 Future work

The results from these studies have improved our understanding of antibody behavior on surfaces, but many details remain to be investigated. The study on antibody stability involved just a single antibody while the antibody-antigen binding studies involved a single antibody fragment and a single antigen molecule. However, other proteins with modifiable affinity for variable targets have been discovered or engineered, and could be used in place of antibodies. Moreover, microarray spots consist of many thousands of affinity molecules. Studying the behavior of one or more antigens interacting with many affinity molecules would better represent microarray conditions. Finally, many protein complexes could be used to improve various industrial processes and consumer products. In cases where surface immobilization is needed, such technologies would necessitate the study of the behavior of these protein complexes on surfaces. These potential research directions are described in greater detail below.

7.2.1 Different affinity molecules

The antigen binding studies in this dissertation involved fab fragments, but other affinity molecules could be tested. An initial research direction would be to redo the fab-antigen simulations in Chapters 5 and 6 using intact antibodies instead of fabs. The results from those chapters indicate that immobilized fabs bind well to antigens under several conditions. However, it is still possible that the large size of intact antibodies could prove beneficial for capturing particular antigens. For example, this may happen when the antigen binding site cannot come close enough to the surface to interact with a fab, such as when the antigen binding site is on the side of a hemagglutinin molecule attached to an intact virus particle. Such work would help answer whether or not intact antibodies should ever be used for microarrays. Simulations involving intact antibodies
Figure 7.1: Coarse-grain intact-antibody models that bind antigens developed for this dissertation. Panel (a) is the antibody-lysozyme model and Panel (b) is the antibody-hemagglutinin model.

could be done because the fab-lysozyme and fab-hemagglutinin models presented in Chapters 5 and 6 started off with intact antibodies (see Figure 7.1). For those chapters, the majority of the antibody was deleted from the parameter set to produce the final fab model, but the original parameter sets with the entire antibodies (where both fabs can bind to each antigen binding site) still exist. Thus, the models are already prepared to study the binding behavior of intact antibodies on surfaces in greater detail.

The small size of fabs has proven useful for several applications, but even smaller affinity molecules exist which could yield yet unknown benefits to microarrays. One such protein is the single-chain variable fragment (scFv, see Panel (a) of Figure 7.2). [175] These are made by deleting the C\textsubscript{L} and C\textsubscript{H}1 domains of a fab and then merging the remaining V\textsubscript{L} and V\textsubscript{H} domains into a single peptide chain. They are about half the size of a fab but retain their typical antigen binding properties. Camelid antibodies are even smaller than scFvs (about the size of lysozyme), yet maintain the ability to bind to specific targets (see Panel (b) of Figure 7.2). [176] Both proteins have demonstrated successful binding under microarray conditions, and sometimes outperform antibodies. Moreover, both are easily produced in \textit{E. coli} cells, which makes them easier to engineer than typical antibodies. As such, scFvs and camelid antibodies will likely be seen more frequently in future microarray studies.
7.2.2 Simulating antibody and fab arrays

Antibody-antigen binding in this dissertation was studied using models with a single fab interacting with a single antigen. However, microarray spots always involve thousands of affinity molecules. As such, the binding behavior of a target molecule with many antibodies or fabs remains unknown. A study involving multiple DNA probes on a surface interacting with a target strand recently identified how the target molecule interacts with the group. [104] The results of this study demonstrated that the hybridization process was more difficult to achieve because of crowding among the probe strands, but that once a probe-target duplex formed it was more stable than when only a single probe molecule was present. Similar crowding effects or increased binding strength could be possible for protein microarrays. Using the antibody- or fab-antigen models from this project, miniature arrays such as the $5 \times 5$ array shown in Figure 7.3 could be simulated. For example, it is likely that the binding behavior of hemagglutinin, with its three symmetric binding sites, would be greatly affected by multiple affinity molecules. Moreover, such studies could also investigate the impact that fab spacing has on binding performance, or if multiple scFvs show stronger binding than multiple fabs.

Figure 7.2: Comparison of a fab, scFv, and camelid antibody. Panel (a) shows the fab as translucent with the scFv portion on top as a ribbon diagram and Panel (b) shows the structure of a camelid antibody.
An ultimate goal of microarrays is to make them reliable enough to provide valid results even if un-purified (“dirty”) samples are added to them. This could be particularly important for medical diagnostics where minimum process time may be of necessity. In such cases the affinity molecules will be required to correctly recognize their targets out of the mixture. Future studies could look at how immobilized affinity molecules interact with different antigens that have similar epitopes. Additionally, such studies could be directed toward understanding the transport phenomenon that move antigens toward the array surface. In either case, it is likely that the coarse-grain model used for this dissertation would need to be modified to enable attractive, but non-specific, protein-protein interactions in addition to the defined native contacts within the antigen binding site, since non-specific interactions drive antibody cross-reactivity.

7.2.3 Simulating other protein complexes

Proteins immobilized to surfaces have the potential to influence numerous technologies. For example, using enzymes in industrial processes could reduce byproduct formation or enable cheaper production of specialty chemicals that are difficult to make with typical organic chemistry methods. Immobilizing the required enzymes to a surface would make them easy to remove from a reaction mixture so that they could be used again. However, some industrially-relevant enzymes, such as Xylose Isomerase (see Figure 7.4), are composed of many subunits that must work in unison. Like the fab-antigen complexes studied in this dissertation, these enzyme complexes have
Figure 7.4: Structure of Xylose Isomerase. This enzyme is composed of 4 symmetric subunits.

unique characteristics and will need to be studied in greater detail to determine the best way to stabilize the complex on a surface.

7.3 Conclusions

In summary, antibody microarrays represent a sensor technology that has the potential to revolutionize molecule detection in a variety of fields that are integral to modern life. Current iterations of these devices have not been able to find mainstream use because of reliability problems. Many variables need to be optimized to improve microarray performance, but details surrounding the protein-surface interactions remain limited. The work in this dissertation was designed to address this lack of knowledge by providing new insights into antibody stability on surfaces, the antibody-antigen binding process, and how antigen properties affect the binding process. The results represent initial steps toward a comprehensive view of the protein-surface interactions that dominate microarray function. Future work in this area will further improve our understanding of these interactions, and aid the development of devices that will be cheaper, more reliable, and produce more-complete molecular profile information.
REFERENCES


136


