

2015

Structural Determinants of Promiscuous and Specific Binding in Protein—Protein Complexes Using Component Analysis Techniques

Aiman Sherani
asherani@wellesley.edu

Follow this and additional works at: <https://repository.wellesley.edu/thesiscollection>

Recommended Citation

Sherani, Aiman, "Structural Determinants of Promiscuous and Specific Binding in Protein—Protein Complexes Using Component Analysis Techniques" (2015). *Honors Thesis Collection*. 312.
<https://repository.wellesley.edu/thesiscollection/312>

This Dissertation/Thesis is brought to you for free and open access by Wellesley College Digital Scholarship and Archive. It has been accepted for inclusion in Honors Thesis Collection by an authorized administrator of Wellesley College Digital Scholarship and Archive. For more information, please contact ir@wellesley.edu.

Structural Determinants of Promiscuous and Specific
Binding in Protein—Protein Complexes Using
Component Analysis Techniques

Aiman Sherani

Advisor: Mala L. Radhakrishnan

Submitted in Partial Fulfillment of the Prerequisite for Honors in Chemical
Physics

April 2015

© Wellesley College

Table of Contents

ABSTRACT	3
ACKNOWLEDGEMENTS	4
1. BACKGROUND AND INTRODUCTION.....	6
1.1 PROTEIN—PROTEIN INTERACTIONS	6
1.2 SPECIFICITY AND PROMISCUITY	10
1.3 GOALS.....	11
1.4 THEORY: BACKGROUND.....	13
1.5 THEORY: CONTINUUM ELECTROSTATICS.....	14
1.6 THEORY: COMPONENT ANALYSIS.....	20
1.7 THEORY: VAN DER WAALS ENERGY CALCULATION	22
1.8 THEORY: SOLVENT ACCESSIBLE SURFACE AREA CALCULATION	23
2. METHODS	24
2.1 WORKFLOW	24
2.2 PROTEIN COMPLEX SELECTION AND CRYSTAL STRUCTURE PREPARATION	25
2.3 CONTINUUM ELECTROSTATICS CALCULATIONS.....	28
2.4 VAN DER WAALS CALCULATION	29
2.5 SURFACE BURIAL CALCULATION.....	29
2.6 STATISTICAL ANALYSIS	30
3. RESULTS.....	32
3.1 DISTRIBUTION OF ELECTROSTATIC COMPONENT ANALYSIS.....	33
3.2 DISTRIBUTION OF VAN DER WAALS COMPONENT ANALYSIS.....	34
3.3 DISTRIBUTION OF SASA COMPONENT ANALYSIS.....	35
3.4 DO SPECIFIC PROTEINS HAVE SIDE CHAIN CONTRIBUTING MORE THAN BACKBONE WHEN COMPARED TO PROMISCUOUS PROTEINS?.....	37
3.5 DO SPECIFIC PROTEINS HAVE MORE SIDE CHAINS THAT CONTRIBUTE SUBSTANTIALLY, COMPARED TO PROMISCUOUS PROTEINS?.....	39
3.6 DO PROMISCUOUS PROTEINS HAVE MORE BACKBONES THAT CONTRIBUTE SUBSTANTIALLY, COMPARED TO SPECIFIC PROTEINS?.....	41
4. DISCUSSION	43
4.1 LIMITATIONS OF THE STUDY	44
4.2 FURTHER ANALYSIS OF RESULTS:	47
4.2.1 <i>Fluctuation of p-values due to random residue selection.</i>	48
4.3 ADDITIONAL FUTURE WORK	49
4.4 CONCLUSIONS.....	51
5. WORKS CITED	52

Abstract

Understanding the physical and chemical principles governing specificity and promiscuity in protein—protein binding is important both for understanding mechanisms of molecular recognition and for designing novel biomolecular systems. The goal of this project is to identify if the energetic contributions of structural moieties (e.g., side chains and backbones of individual residues) are different between promiscuous and specific protein—protein interactions. To achieve this goal, we are testing multiple hypotheses; for example, we hypothesize that specific proteins, which selectively bind to only one partner, preferentially utilize side chains to mediate binding when compared to promiscuous proteins, which may utilize the structurally consistent backbone moieties more preferentially. Electrostatic contributions of the structural moieties toward binding are quantified using component analysis techniques within a continuum electrostatic framework that takes solvent effects into account. Van der Waals and surface burial contributions are also evaluated. Two sets of protein complexes, identified by the literature whenever possible to be either promiscuous or specific, are being analyzed to test our hypotheses. Our preliminary results indicate our hypotheses may be partially verified in the case of promiscuous proteins. Interestingly, these predictions do not hold when considering the binding partners of the promiscuous proteins.

Acknowledgements

First and foremost, I would like to thank my amazing advisor, Professor Mala Radhakrishnan. She is simply the best teacher and mentor a student can have. Her support, guidance, patience, and honest feedback have helped me learn more than I could have hoped in my Wellesley career.

Second, I would like to thank the faculty on my thesis committee. Professors Christopher Arumainayagam, Glenn Stark, and Don Elmore. Thank you for your support and insight for this project. A special thank you to Professors Christopher Arumainayagam and Glenn Stark for supporting me in pursuing an interdepartmental major. I would also like to thank Professor Casssandra Pattanayak for her assistance with the statistical analysis.

I am also grateful to my fellow Radhakrishnan lab members for their encouragement and support. A special thank you to Ying Yi Zhang for completing a lot of the preliminary calculations for this project. I would also like to thank Jackie Blum'14, Connie Chen'15, and Helena Qi'14 for all of their help and support.

I would like to thank my family, especially my mom for always supporting and encouraging me to pursue my dreams. Thank you for all of your support, guidance, and unconditional love.

I would also like to thank my friends here at Wellesley. A special thank you to Tamara Biary'15, Harini Natrajan'15, Nidhi Saxena'15, Emily Rothkin'15 and Nour Azzouz'15. The past four years would have been starkly different without your constant support and love. Thank you for your encouragement and support in helping me to decide to pursue graduate school. Another special thank you to my fellow thesising buddies, Eunice Paik'15, Frieda Zhang'15, and Kirsten Blancato'15, for always being a source of support and encouragement. Thank you to my fellow Chem 335 classmates, Michelle Brann'15, Hannah Sim'16, Jane Zhu'16, and Hikari Murayama'16, for all of their support.

I would also like to acknowledge the Chemistry and Physics departments at Wellesley. Thank you to Professor Hu for her encouragement and advice. A special thank you to Professor Battat in helping me with my graduate school applications.

Finally, I would like to thank Wellesley College for giving me the opportunity to pursue this research.

1. Background and Introduction

1.1 Protein—Protein Interactions

Proteins are the molecular machines of cells. They are important for all cellular functions, including maintaining cell structure, facilitating cell transport, catalyzing reactions, enabling DNA replication, carrying out the immune response, aiding cellular viral entry, and synthesizing new molecules. Most protein functions are mediated by protein—protein interactions. For example, the protein erythropoietin binds with its receptor to control red blood cell production; perturbations in this interaction can result in anemia and other diseases [1]. As another

example, Hemagglutinin is a glycoprotein of the influenza virus and mediates viral entry into the cell via its interaction with cell membrane glycoproteins [2].

Calmodulin binds to multiple target proteins to regulate vital cell processes, including inflammation and smooth muscle contraction [3]. Investigating the

binding affinity of protein—protein interactions is important for understanding mechanisms of molecular recognition between proteins; the consequent interactions between these proteins drive biological systems. Furthermore, because disturbances in these interactions can bring about diseases such as cancer [4, 5] and amyloidosis [6, 7], these principles are important for understanding mechanisms of disease and for designing novel biomolecular systems that can regulate these interactions.

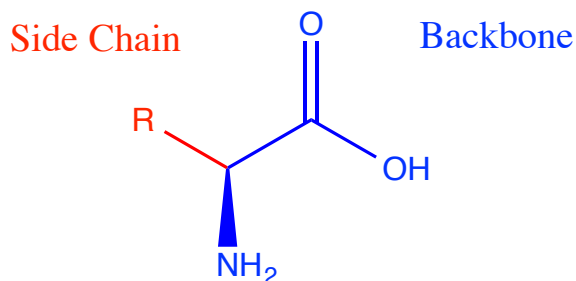


Figure 1.1.1: Amino Acid Structure

These are the components of an amino acid. The backbone atoms (blue) are conserved between each amino acid. The side chains (red) are different for each amino acid.

Proteins are polymers composed of multiple amino acids. There are 20 different amino acids found in nature. Each of them has an identical backbone but a different side chain. These side chains can be polar, non-polar, acidic, or basic. The amino acid sequence of a protein is called its primary structure. Each amino acid in this sequence is called a residue. Interactions between amino acids drive the protein to fold into a three dimensional structure. Residues on the surface of different proteins can interact with each

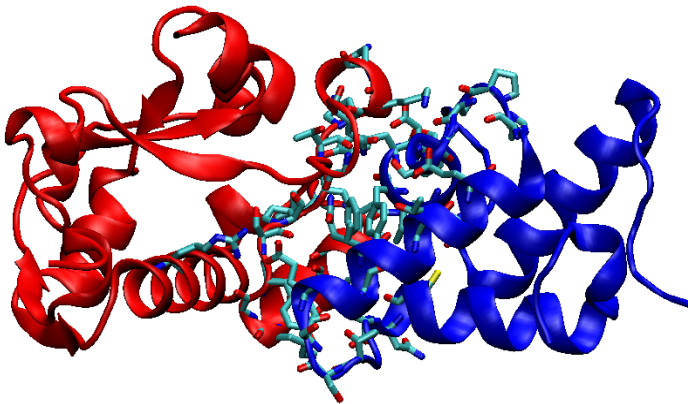


Figure 1.1.2: Interfacial Residues

Interfacial residues of a protein—protein complex are explicitly shown here. In this work, these residues are defined as the amino acids within a certain cutoff distance from the closest residue on the binding partner.

other. If the three-dimensional structure of two proteins is complementary, these proteins can bind to each other via an interface. The residues on a protein that are close to or in contact with the binding partner are called interfacial residues. Interactions between the

interfacial residues of the binding partners generally drive the formation of a protein—protein complex.

Protein association can be treated as a rapid and reversible process in equilibrium, where the binding affinity between the two partners can be obtained experimentally from the dissociation constant [8]. In aqueous solvent, water molecules are able to make hydrogen bonds with other polar molecules. When nonpolar molecules are present in aqueous solvent, the number of potential hydrogen bonds and the entropy of the solvent decreases [9]. Consequently, the change in surroundings of nonpolar molecules from solvent to other nonpolar molecules results in higher entropy for the solvent. This

phenomenon is termed the hydrophobic effect. Analogous to its prominent role in protein folding, the hydrophobic effect was believed to be the main driving force behind protein-protein association, because the entropy of solvent is increased upon association of macromolecules within it [10-12]. However, further work revealed that hydrophobic interactions were disproportionately more favorable for smaller interfaces than for larger interfaces and that the hydrophobic effect may not drive the association of complexes with larger interfaces [13]. Jones and Thornton found that for complexes comprised of proteins that also exist as monomers, the interfacial residues are less hydrophobic when compared to the interfacial residues of homodimers [14]. These data imply that hydrophobic interactions may not necessarily be the main driving forces of all protein—protein complexes. Additionally, certain interfacial residues, termed “hot spots”, were found to contribute significantly to the binding free energy using an experimental technique called alanine scanning, which evaluates the contribution of the side chain to the binding free energy by substituting a residue with alanine, thereby eliminating the side chain [15]. Hot spots are residues where substitution to alanine increases the binding free energy by at least 2.0 kcal/mol [16]. Prior analysis of hot spots revealed tryptophan, arginine, and tyrosine as major contributors to binding affinity while leucine, serine, threonine, and valine were not usually hot spot residues [17]. Because some of the hot spot residues had polar or even charged side chains, this finding implied that such residues may contribute to binding affinity due to specific physical mechanisms and that forces other than hydrophobicity are also involved in driving protein—protein interactions.

These forces include electrostatic [14] and van der Waals [18] interactions. The

idea of electrostatic complementarity between protein—protein interfaces was confirmed in a previous study [19]. A study that compared the interfaces of specific protein—protein complexes with nonspecific, crystal packing contacts, found van der Waals and electrostatic interactions had a greater role in the association of specific complexes, when compared to hydrophobic interactions [20]. Additionally, experimental [15] and computational [21] work has identified that most polar and charged interfacial residues are hot spots, implying electrostatic interactions are important in driving complex formation. Additionally, another study characterized the residues surrounding hot spot residues to be unimportant for binding, with the authors suggesting that these residues were present near the hot spot residues to shelter them from interacting with solvent [17]. This mechanism implies hot spot residues are stabilized upon binding because they are involved in more favorable interactions when the complex is bound. Polar and charged interfacial residues, when compared to the residues in the core of the protein, experience higher electrostatic stabilization upon the formation of the protein—protein complex [22, 23]. However, it should be noted that this net electrostatic stabilization is controversial [24]. Another study found that the pK_a shifts for acidic residues upon protein association were negative, suggesting that these residues often stabilize the complex when compared to a hypothetical complex in which that residue was mutated to an alanine [25]. The average pK_a shift is an indication of how much the electrostatic interactions compensate for the desolvation of interfacial residues. Other studies have shown that a higher amount of charged interfacial residues results in higher affinity [26, 27].

Van der Waals interactions are also thought to have a substantial energetic contribution in driving protein binding. This interaction results from the dispersion

attraction between non-polar molecules and the exchange repulsion between nuclei. These forces are of a shorter range than electrostatic interactions. However, their magnitude increases significantly as the distance between the binding partners decreases and are important in driving binding [18]. The van der Waals interactions between non-polar side chains are also thought to be favorable for binding [28].

Taken together, the findings of multiple studies confirm that electrostatics, van der Waals, and hydrophobic interactions are important driving forces in protein—protein association.

1.2 Specificity and Promiscuity

The specificity of protein—protein interactions can vary widely. Specific proteins bind tightly with only one partner and do not bind to other partners [29]. Other proteins, known as promiscuous proteins, can bind to multiple partners with ample affinity. A particular protein's tailored recognition of binding partners is due to its biological function. For example, calmodulin needs to bind to multiple binding partners to act as a “messenger” molecule and regulate multiple physiological processes such as inflammation and smooth muscle contraction [3, 30]. On the other hand, erythropoietin (Epo) binds specifically to its binding partner, erythropoietin receptor (EpoR), to control red blood cell production. Uncontrolled binding to other partners could lead to unwanted proliferation of other cell types, and disturbances in the Epo-EpoR interaction lead to anemia [1]. Understanding the structural determinants of promiscuity and specificity in protein—protein binding can have applications in biomolecular design because these findings can be applied to predict protein association [31, 32].

Previous studies have sought to investigate the mechanism of promiscuity and

specificity in protein—protein interactions. Specific complexes are thought to have interfaces that are complementary in terms of hydrogen bonds, steric interactions, and ionic interactions [8]. Electrostatics have been identified to be important in driving specific protein association, as charged molecules have been shown to be more specific binders when compared to hydrophobic molecules [33]. Other experiments have identified certain amino acids to be important in mediating specificity. Tyrosine and serine were shown to be important for the binding of protein—protein complexes with high affinity and specificity [34]. In a study done by Birtalan *et al.*, replacing the tyrosine with a tryptophan reduced specificity in different antibody-antigen complexes [35]. Promiscuity has been attributed to nonspecific, hydrophobic interactions between binding partners. However, James and Tawfik showed that for a particular promiscuous protein, each interaction with a partner makes a specific set of hydrogen bonds [36]. Nevertheless, promiscuous proteins have been identified to preferentially use common interface residues to bind to different binding partners [17, 37, 38]. While certain amino acids have been identified to be important in facilitating specificity, the energetic contributions of particular structural moieties in mediating specific or promiscuous interactions have not been studied to our knowledge.

1.3 Goals

The goal of this project is to understand if the energetic contributions of structural moieties are different between promiscuous and specific protein—protein interactions. In particular, this project will evaluate the following questions:

1. Are side chains more important in specific protein binding than in promiscuous protein binding?

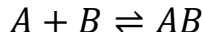
2. Are backbones more important in promiscuous protein binding than in specific protein binding?

A moiety is considered important if it contributes sufficiently favorably to the binding free energy. These questions are evaluated in terms of the electrostatic, van der Waals, and surface burial free energies. With respect to the first question, we hypothesize that specific proteins, which selectively bind to only one partner, may preferentially utilize side chains, which vary between different proteins, to recognize their binding partner; on the other hand, promiscuous proteins, which bind to multiple partners, are predicted to preferentially utilize the conserved backbone to recognize their binding partners. This study analyzes 5 promiscuous proteins, each complexed with one or more partners, and 8 specific protein—protein complexes. We employ a two sample test of proportions to evaluate the statistical significance of our findings.

The next section will describe the models we use to determine the important structural moieties in promiscuous and specific protein binding.

1.4 Theory: Background

In this work, we quantify structural determinants of promiscuity and specificity in protein—protein interactions by quantifying their contributions toward the binding free energy of the protein complexes. Hence, we now outline the theory behind the models we use to evaluate the electrostatic, van der Waals, and surface burial binding free energies. The binding of two proteins, A and B, associating to form a complex, AB, can be represented by the reversible reaction:



The change in the Gibbs free energy of binding is a measure of the favorability of the interaction between two proteins, i.e., the binding affinity. The binding free energy can be broken down into multiple terms, each due to a particular physical aspect:

$$\Delta G_{Total} = \Delta G_{Elec} + \Delta G_{SASA} + \Delta G_{vdW} + \Delta G_{int}$$

ΔG_{Elec} is the change in Gibbs free energy due to electrostatic interactions. In the classical molecular mechanics model we use in this work, the electrostatic energy is constituted of two components: the interactions between pairs of atoms with positive or negative partial charges and the interactions between atoms and the aqueous solvent. ΔG_{SASA} results from the change in solvent accessible surface area of each binding partner. It coarsely quantifies the hydrophobic effect. The van der Waals energy, ΔG_{vdW} , arises from interactions between temporary shifts of electron clouds that create dipoles, inducing dipoles in neighboring moieties that can favorably interact. ΔG_{int} is the change in the internal conformational energy and entropy of each protein upon binding. For all of our systems, we assume ΔG_{int} is zero because we assume rigid binding. In general, this assumption is commonly used [39] and is reasonable for protein complexes that do not

experience large conformational changes upon binding. This study focuses on the electrostatic, van der Waals, and surface burial interactions because, as summarized in the first section, these interactions are thought to largely drive protein—protein binding.

Below is a description of how we evaluate the electrostatic, van der Waals, and surface burial terms.

1.5 Theory: Continuum Electrostatics

Electrostatic interactions can be both nonspecific and specific because they have long-range and short-range effects [40]. Electrostatic properties depend on interactions between electronic wave functions and would ideally be modeled through quantum mechanics. However, quantum mechanical models are computationally expensive and their accurate application is limited to systems of small size, and so we use a classical molecular mechanics model in this work.

Assessing the electrostatic properties of a biomolecule involves quantifying its interaction with the solvent. Explicit solvent models depict interactions made to each water molecule and electrolyte ion. These models are used in conjunction with molecular dynamics simulations to characterize the changes in each solvent particle upon binding. Again, this method is computationally expensive and cannot be efficiently applied to a study with a scope as large as ours. For studies such as ours, implicit solvent models are often used to investigate electrostatic properties of biomolecules. Implicit models take solvent polarization and reorganization into account without explicitly depicting each water molecule and ion. They have been shown to give results comparable to explicit models in identifying low energy conformations of biomolecules [41].

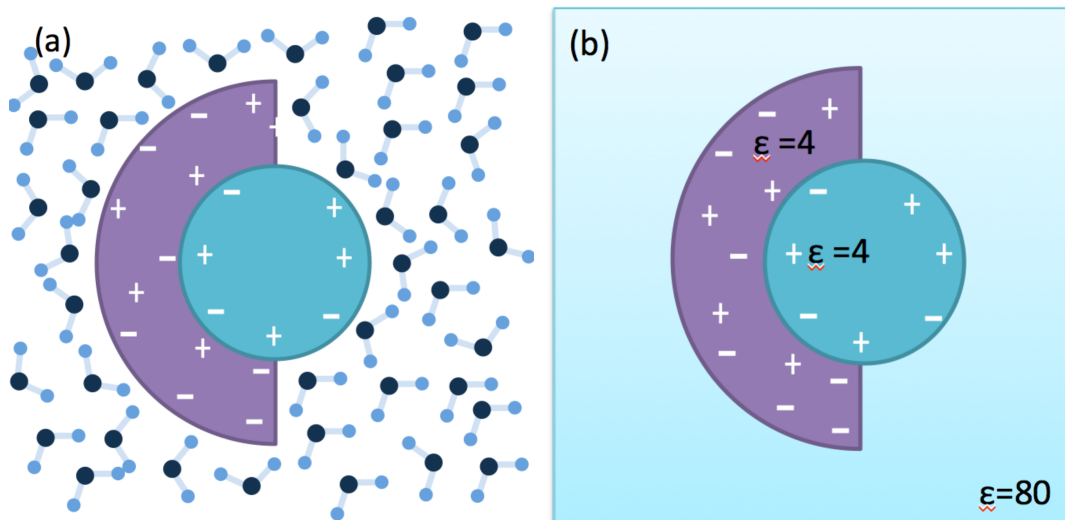


Figure 1.5.1: **Continuum Electrostatics**

(a) This figure is a schematic of an explicit solvent model of protein binding, where each water molecule is represented.

(b) This figure represents an implicit solvent model where the solvent is a highly polarizable medium and the proteins as a low dielectric medium with explicit point charges.

Continuum electrostatic models are a subset of implicit models. In such models, the solvent is represented as a highly polarizable bulk medium, with a high dielectric constant value of 80 [40]. The protein complex is represented as a low dielectric medium containing point charges [42]. The dielectric constant of the biomolecule(s), the protein—protein complex in our case, can range from 2 to 40, and would ideally depend on the physical properties of a particular biomolecule [23, 40]. Our model uses an inner dielectric value of 4, which is thought to account for the polarizability of the protein backbone [23, 43]. Our model also accounts for the presence of salt ions, and therefore employs the Poisson-Boltzmann Equation, which is described in further detail below, to find the electrostatic potentials of the protein complex and surrounding solvent.

The change in the electrostatic binding free energy during protein complex formation is the sum of three components. The ligand desolvation penalty is the energetic “cost” the protein must pay to interact with its receptor. The “cost” is associated with the

loss of favorable interactions of the interfacial residues with the surrounding water molecules. The receptor desolvation penalty is the unfavorable cost of the interfacial residues losing their interactions with solvent. The interaction energy is the solvent-screened interaction between charges on the binding partners. This energy is often favorable.

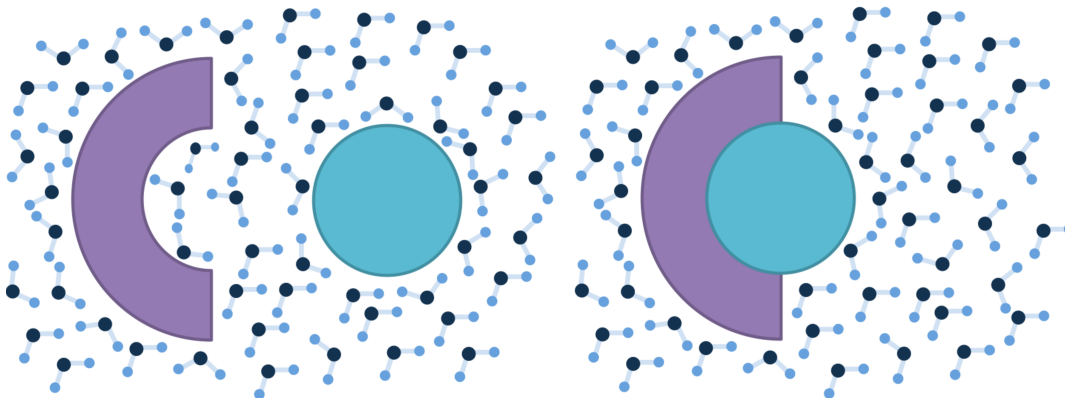


Figure 1.5.2: **Desolvation Penalty**

(a) This illustration shows both binding partners in solvent and able to interact with water.
 (b) This illustration shows the binding partners interacting with each other and unable to interact with solvent at the interface. The difference between panels (a) and (b) represents the desolvation “cost” both binding partners pay in order to interact with each other.

The following explanation of continuum electrostatics follows Michael Gilson’s approach and Emma Nechamkin’s (’12) thesis [44, 45]. The electrostatic potential in space, $\phi(\mathbf{r})$, can be found for a known charge density, $\rho(\mathbf{r})$, in a vacuum by solving the Poisson equation:

$$-\nabla \cdot \nabla \phi(\mathbf{r}) = \frac{\rho(\mathbf{r})}{\epsilon_0}$$

(Equation 1)

where ϵ_0 is the permittivity of free space constant.

If we assume a charge distribution of point charges, the solution to the Poisson Equation for two point charges in a vacuum is Coulomb’s law:

$$U(r_{12}) = \frac{q_1 q_2}{4\pi\epsilon_0 r_{12}}$$

(Equation 2)

where ϵ_0 is the permittivity of free space constant. This equation can allow us to calculate the electrostatic energy due to the two point charges. This value represents the amount of work required to move the two infinitely far apart point charges to a mutual distance of r_{12} .

Biomolecules are often surrounded by an aqueous, highly polarizable solvent. Therefore, in order to accurately evaluate the electrostatic energy of biomolecules, we cannot assume our charges are interacting in a vacuum. If we assumed the biological molecules were free-floating point charges in solvent, we could reasonably model them as surrounded by a medium with a uniform dielectric constant (ϵ). A dielectric constant accounts for the polarizability of the medium. The Poisson equation for a charge distribution, $\rho(\mathbf{r})$, in a medium with a constant dielectric constant of ϵ is:

$$-\nabla \cdot \nabla \phi(\mathbf{r}) = \frac{\rho(\mathbf{r})}{\epsilon_0 \epsilon}$$

(Equation 3)

The Poisson Equation assumes a linear response. The field created from solute interaction with the solvent is proportional to the charge of the solute. The solution to this equation for two point charges in a medium with a dielectric constant of ϵ is:

$$U(r_{12}) = \frac{q_1 q_2}{4\pi\epsilon_0 \epsilon r_{12}}$$

(Equation 4)

Intuitively, the presence of the dielectric medium weakens the interaction between the

two point charges by a factor of $1/\epsilon$. This occurrence is termed dielectric screening. The electric field due to the point charges is opposed by the induced field of the polarizable medium, resulting in an overall, weaker electric field.

Furthermore, the charges of interest in our model are surrounded by a non-uniform dielectric field. The degree of polarizability of surrounding charges in the protein—protein complex differs from the degree of polarizability of the solvent. Therefore, the dielectric constant can be modeled as a discontinuous function, $\epsilon(\mathbf{r})$, where the constant value would change instantaneously at the protein-solvent boundary. The Poisson Equation can now be represented as:

$$-\nabla \cdot [\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})] = \frac{\rho(\mathbf{r})}{\epsilon_0}$$

(Equation 5)

The discontinuous boundary alters the surrounding electric field by inducing surface charges at the protein-solvent boundary, which induce an electric field. Hence, this boundary significantly alters the electrostatic interactions in the system.

In addition to a highly polarizable medium, biomolecules are also surrounded by mobile salt ions. The Poisson equation does not account for the salt ions. This necessitates the use of the Poisson-Boltzmann equation, which arises from using Debye-Huckel theory to implicitly model the salt ions within the solvent. The concentration of salt ions, c_{bulk} , is approximated using the Boltzmann factor instead of explicitly accounting for each salt ion in the system. The Poisson-Boltzmann equation is:

$$-\epsilon_0 \nabla \cdot [\epsilon(\mathbf{r}) \nabla \phi(\mathbf{r})] = \rho^f(\mathbf{r}) + \sum_{i=1}^N q_i c_{i,bulk}(\mathbf{r}) e^{-\beta q_i \phi(\mathbf{r})}$$

(Equation 6)

where ρ^f are the source charges and $e^{-\beta q_i \phi(\mathbf{r})}$ is the Boltzmann factor of ions in electrostatic potential. The Linearized Poisson-Boltzmann equation arises when the first order term of ϕ of the Taylor expansion of the Boltzmann factor is used. This approximation is valid when the value of the exponent is close to zero, which is true when the potential is small (e.g., small magnitude charge distributions). The Linearized Poisson Boltzmann equation is:

$$-\epsilon_0 \nabla \cdot [\epsilon(\mathbf{r}) \nabla \phi(\mathbf{r})] = \rho^f(\mathbf{r}) - \epsilon_0 \epsilon(\mathbf{r}) \kappa^2(\mathbf{r}) \phi(\mathbf{r})$$

(Equation 7)

where

$$\kappa^2 \equiv \frac{\beta}{\epsilon \epsilon_0} \sum_1^N c_{i,bulk} q_i^2 = \frac{2}{\epsilon \epsilon_0 kT} I$$

Here, $I = \sum_i \frac{1}{2} c q_i^2$, which represents the ionic strength. The Linearized Poisson Boltzmann equation (LPBE) is easier to manipulate and solve than the Poisson Boltzmann Equation. The linearity allows us to use matrix representations to calculate the electrostatic energy of the entire system by taking the sum of the product of potentials and charges q_i at each point in space [40]:

$$G = \sum_i \frac{1}{2} q_i \phi_i$$

(Equation 8)

The factor of $\frac{1}{2}$ prevents the double counting of contributions for a pair of charges [46].

Additionally, the $\frac{1}{2}$ factor represents the entropic contribution of the solvent upon interacting with a given charge [40], and this is why the ultimate energies obtained are free energies as opposed to merely energies.

We use the Finite Difference Method to numerically solve the LPBE. The protein-protein complex is digitized onto a three-dimensional grid. The electrostatic potential, charge, and ionic strength are all defined at each grid point and the dielectric constant is defined on the grid lines [47]. Each atom in the protein complex is assigned a partial atomic charge, which is extrapolated onto the grid and therefore approximately represented.

The potential at each grid point is found by solving the LPBE. The partial second derivative along each dimension is found approximately from the finite differences between neighboring grid potentials. After the potentials are found, the total electrostatic free energy can now be solved for using equation 8. The potentials are calculated for both the bound and unbound states. Differences in potentials between the two states are taken to cancel out grid energy artifacts and to provide free energy differences.

1.6 Theory: Component Analysis

To quantify the importance of a structural characteristic in the electrostatic component of protein—protein binding, we use a technique called component analysis. Component analysis quantifies the contribution of a particular moiety to the electrostatic binding free energy. A moiety is considered important if there is a large difference between the original electrostatic binding free energy and the electrostatic binding free energy without a charge contribution from the certain moiety, i.e., with all partial atomic charges on that moiety set to zero. The moiety can be a particular side chain, backbone,

the entire residue, or entire regions of a protein. Component analysis can be represented as:

$$\Delta\Delta G = \Delta G_{\text{zeroed}} - \Delta G_{\text{original}}$$

(Equation 9)

A positive $\Delta\Delta G$ indicates that the moiety contributes favorably to the electrostatic binding of the protein—protein complex.

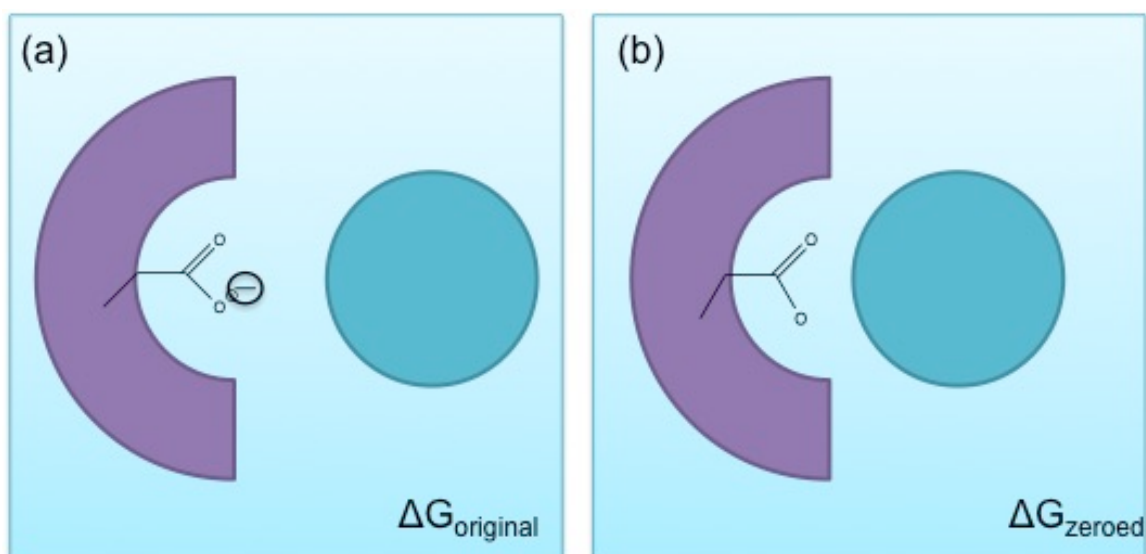


Figure 1.6.1: **Component Analysis**

This figure pictorially represents the component analysis technique.

(a) Depicts the original system. (b) Shows the system with the charge of a moiety set to zero. The change in the electrostatic binding free energy when all partial charges within a structural moiety of interest are set to zero quantifies the importance of that moiety in binding.

When considering a side chain moiety, component analysis is similar to alanine scanning, which is an experimental technique that evaluates the importance of the side chain to the binding free energy by substituting the residue of interest to alanine [15]. However, unlike alanine scanning, component analysis does not change the shape of the protein—protein interface and has no impact on the protein fold in the virtual model. Previous researchers have used component analysis to determine the electrostatic energetic

contributions of particular structures in protein complexes [26, 46, 48-52].

1.7 Theory: Van der Waals Energy Calculation

The CHARMM22 [53, 54] force field is used to evaluate the contribution of a structural moiety to the van der Waals binding free energy. The van der Waals interaction is modeled using the Lennard-Jones (LJ) potential energy function:

$$U(\mathbf{r}) = \varepsilon_{ij}^{min} \left[\left(\frac{R_{ij}^{min}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^{min}}{r_{ij}} \right)^6 \right]$$

(Equation 10)

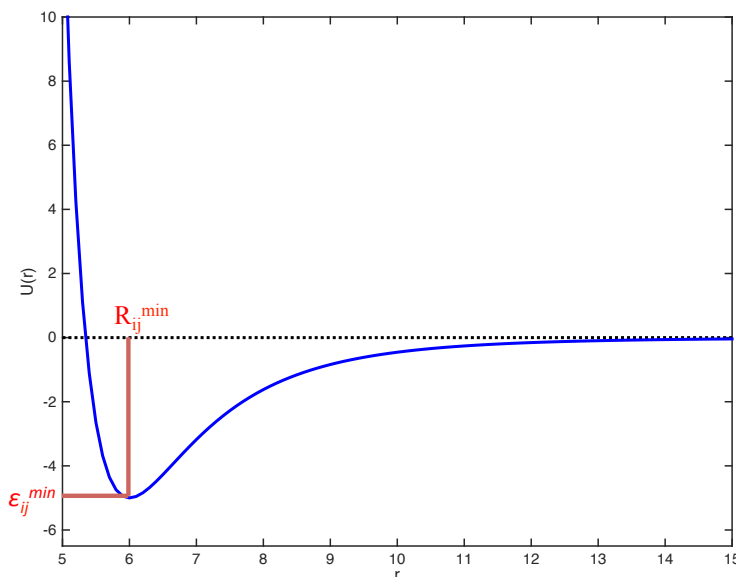


Figure 1.7.1: Lennard Jones Potential

This figure shows a graph of the Lennard Jones Potential. The well depth of the potential is represented by ε_{ij}^{min} and is a measure of the attraction between two molecules. The minimum of the potential corresponds to the R_{ij}^{min} value.

The LJ potential accounts for both the weak, short-ranged attractive dispersion interactions and the strong, shorter-ranged exchange repulsion between atomic centers, described below.

ε_{ij}^{min} is the well depth of the

LJ potential for interacting atoms i and j and is a

measure of the attraction

between two atoms. R_{ij}^{min} is

the corresponding distance of the LJ potential minimum and r_{ij} is the distance between

atoms i and j . ε_{ii}^{min} and R_{ii}^{min} are parameters that are determined for each atom type from

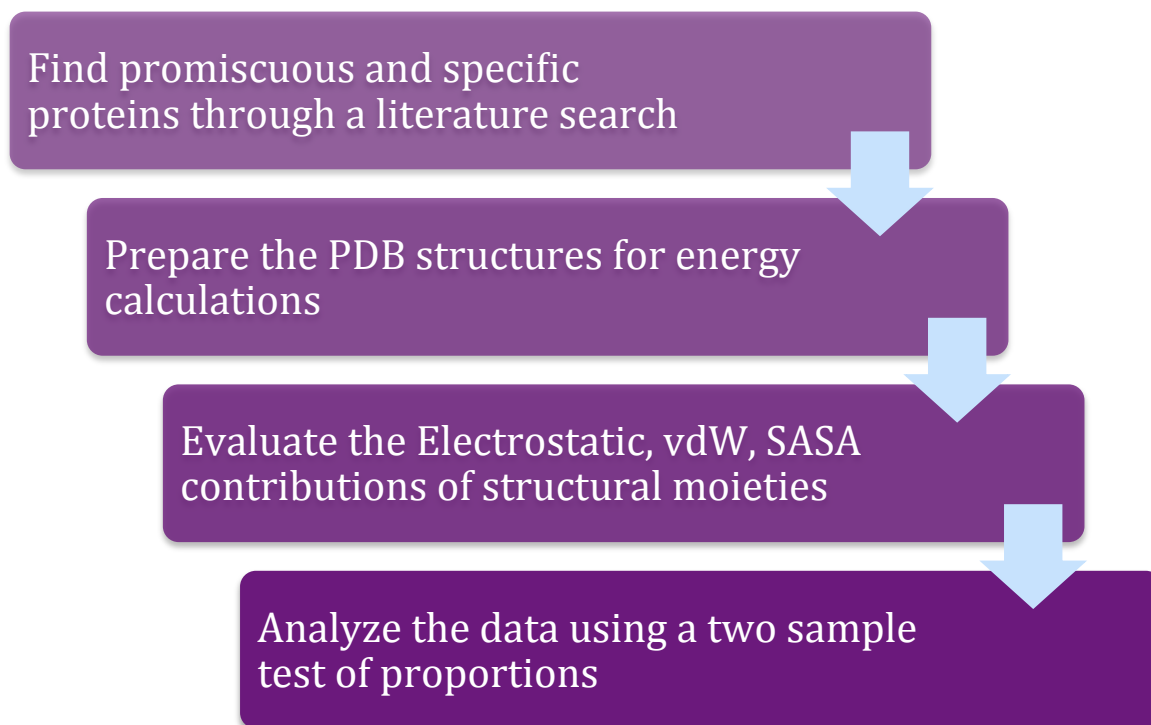
fitting to experimental and theoretical data. ϵ_{ij}^{min} is then found through a geometric mean of ϵ_{ii}^{min} and ϵ_{jj}^{min} [55]. Similarly, R_{ij}^{min} is found through an arithmetic mean of R_{ii}^{min} and R_{jj}^{min} [55]. The r^{-12} term represents the short-range repulsion between two molecules due to Pauli exchange repulsion. This functional form is chosen empirically, for computational convenience, and does not have a quantitative physical basis. The r^{-6} term represents the long-range attraction between molecules that results from the instantaneous dipole of the electron cloud of one molecule inducing a dipole of another molecule.

1.8 Theory: Solvent Accessible Surface Area Calculation

Kauzmann suggested long ago that the hydrophobic effect is a major contributor in protein association [12]. Other researchers discovered that the binding free energy of a complex was linearly proportional to the interfacial buried surface area [56]. However, further work showed this metric alone is not entirely predictive of the binding free energy and that accounting for other factors, such as the polar or nonpolar nature of each atom type, more accurately correlated buried surface area to the binding free energy [57]. Nevertheless, the surface area burial model is still commonly used for quantifying energetics due to the hydrophobic effect [14, 56, 58, 59], and it is the model used in this work as well.

2. Methods

2.1 Workflow



The workflow of the project is shown in the figure above. The rate-determining step of the thesis was finding and preparing the structures studied in this project.

2.2 Protein Complex Selection and Crystal Structure Preparation

The promiscuous and specific protein complexes studied in this project are shown in Tables 1 and 2, respectively.

Table 2.2.1: Promiscuous Protein Complexes			
Protein	Binding Partner	PDB ID	Resolution (Å)
BPTI [62]	Serine protease	3U1J [60]	1.80
	Trypsin	3BTK [61]	1.85
Trypsin [62]	Protease inhibitor	2XTT [63]	0.93
	Textilinin	3D65 [64]	1.64
	Amyloid beta-protein precursor	1TAW [65]	1.80
CaM [29]	CaM-dependent protein kinase II-alpha	1CM1 [66]	2.00
	Ryanodine receptor 1	2BCX [67]	2.00
	Voltage-dependent Ca ²⁺ channel	2F3Y [68]	1.45
	Glutamate NMDA receptor	2HQW [69]	1.90
	Neuronal nitric oxide synthase	2O60 [70]	1.55
	Alpha-II spectrin	2FOT [71]	2.45
CDK2 [72]	CksHs1 (regulatory protein)	1BUH [73]	2.60
TEM1-β lactamase [29]	β lactamase inhibitory protein	1JTG [74]	1.73
	β lactamase inhibitory protein II	1JTD [74]	2.30

Table 2.2.2: Specific Protein Complexes

Protein	Binding Partner	PDB ID	Resolution (Å)
Barnase [75]	Barstar	1BRS [76]	2.00
Epo [75]	Epo-R	1EER [77]	1.90
1F9	Apical membrane antigen 1	2Q8B [78]	2.30
Murin IgG	Interleukin-18	2VXT [79]	1.49
Lebrikizumab	Interleukin-13	4I77 [80]	1.90
Angiogenin	Ribonuclease Inhibitor	1A4Y [81]	2.00
Fab HyHEL-5	Hen egg-white lysozyme	1YQV [82]	1.70
Lysozyme	VH Single Domain antibody	1MEL [83]	2.50

The complexes were identified as promiscuous or specific through a literature search whenever possible (see column 1 of Tables 2.2.1 and 2.2.2). The crystal structures were required to contain a complex of two proteins. Because lower resolution structures can result in a higher uncertainty in interpretations of the electron density map [84], the resolution of the structures was restricted to 3.0 Å or better. The structures contained no DNA or RNA elements.

The crystal structures of the 22 protein complexes were obtained from the Protein Data Bank (PDB) [85]. A pairwise sequence alignment algorithm was used to compare the primary sequences of the promiscuous proteins' binding partners. Any binding partner that was more than 45% identical in sequence to another binding partner was not selected in this study. A subset of the structures were also aligned using the McLachlan algorithm [86] as implemented in the ProFit program (Martin, A.C.R.,

<http://www.bioinf.org.uk/software/profit/>). These requirements ensured that the binding partners considered were not identical or nearly identical (mutant variants, etc.) in sequence and structure.

The protein structures were prepared for computational analyses using the following steps. Crystallographically resolved water molecules that did not have 3 or more potential hydrogen bonding contacts with protein atoms were eliminated from the structures. Non-interfacial solvent ions were also removed from the structures, as salt was modeled implicitly. For asparagine (Asn) and glutamine (Gln) side chains, the symmetry of the electron density of the amide group makes determination of the likely rotameric state experimentally difficult, leading to a chance of selecting the “incorrect” rotamer during crystallization [87]. Additionally, the histidine (His) imidazole orientation and tautomerization states are also ambiguous in the crystal structures. For Asn, Gln, and His residues within 5 Å of each binding interface, the side chains were visually analyzed for the number of possible hydrogen bonds, according to distance and geometry, to determine the most likely rotameric and His tautomerization state. For residues farther from the interface, a C++ program written by Ying Yi Zhang '13, which also evaluates the number of possible hydrogen bonds, was used to determine the likely rotameric or His tautomerization state. The CHARMM22 [53, 54] force field was used to build hydrogen atoms into the structure, patch missing side chains, and cap blunt ends of residues near missing non-interfacial segments of the protein. Hydrogen atoms were built using the HBUILD facility [88]. The built-in patches and caps were minimized to decrease steric clashes. In one complex, the MODELLER [89] software package was also used to build in missing interfacial residues when backbone was also missing. Next, each

protein structure was rotated in space to maximize grid resolution. Finally, PARSE atomic radii and charges were used for the continuum electrostatics calculations [90]. The protein structures were visualized in VMD [91].

2.3 Continuum Electrostatics Calculations

The Linearized Poisson-Boltzmann Equation was solved numerically to find the electrostatic potential and ultimately, the electrostatic binding free energies. A locally built solver [92] that uses the Finite Difference Method (FDM) was used to solve the LPBE. Because the grid points at the edges have undefined neighboring potentials, a boundary condition is required for a solution. Initially, potentials were therefore solved for on a very low resolution grid in which the protein—protein complex constitutes 23% of the grid, such that the boundary potentials can be accurately assumed to obey limiting behavior, though the protein potentials obtained are very coarse. The solutions from this first iteration were used in the next iteration and the potentials were found for a system in which the complex occupies 92%, and ultimately 184% of the grid, resulting in more refined potentials for the protein. At the highest focusing, the grid was centered on an interfacial residue atom on one of the binding partners. Three translations of the grid were used and averaged. The dielectric constant of the solvent was set to 80. The inner dielectric constant of the protein complex was set to 4. The ionic strength of the solvent was set to 0.145M. The molecular surface was defined by rolling a water-sized (1.4 Å radius) sphere around the molecule(s), and the Stern layer probe radius was set to 2Å. The Stern layer is a layer around the molecule generated by rolling an ion-sized (here, 2 Å) sphere around the molecular surface to determine where the ionic strength is nonzero (outside the Stern Layer) and where it should still be zero, as those regions of solvent are

inaccessible to ions. A 257 X 257 X 257 cubic grid size was used for the FDM, yielding grid resolutions of 5.58-10.22 grids/Å at the highest focusing.

Component analysis calculations were done on all of the residues with at least one atom that was within 5 Å of an atom on the other binding partner. The contribution of a backbone or side chain moiety of each residue was determined by finding the change in the electrostatic binding free energy when the partial atomic charges on all atoms of that moiety were set to zero.

2.4 Van der Waals Calculation

We employed the Lennard Jones potential, using the CHARMM22 force field [53, 54], to evaluate the van der Waals energy contributions of each structural moiety. Specifically, the van der Waals interaction energy was found between an interfacial residue's backbone or side chain moiety, and the other protein molecule for the protein—protein complex as a sum of the pairwise LJ interactions between each atom on the moiety and each atom on the binding partner.

2.5 Surface Burial Calculation

The surface burial upon binding was calculated as the difference in solvent accessible surface area upon binding of the two proteins using CHARMM [54]. A 1.4 Å probe radius was used to generate the solvent-accessible surface area. This area can very roughly be related to the contribution of a residue, or its backbone or side chain, toward hydrophobic entropy-related solvent interactions [90].

2.6 Statistical Analysis

A one tailed, two sample test of proportions was used to evaluate if there was a statistically significant difference between the side chain and backbone contributions in specific and promiscuous protein complexes. The statistical analyses were performed using the R software package [93].

Additionally, the component analysis results for the promiscuous proteins were analyzed using three different perspectives. Promiscuous interactions can result from a characteristic present in the promiscuous protein itself, i.e., the protein that binds to multiple partners. Perspective #1 therefore considered all of the component analysis results from residues within the promiscuous proteins. Alternatively, promiscuity can result from a characteristic present on the binding partner. Perspective #2 considered the component analysis results from the binding partners of promiscuous proteins. Lastly, promiscuity can be a result of characteristics present in both the promiscuous protein and the binding partner. Perspective #3 considered the component analysis results from the entire promiscuous protein complex. Because we assume that this potential asymmetry does not exist for specific complexes, each perspective considered for the promiscuous data set was compared with all of the results from both partners of each specific protein binding complex.

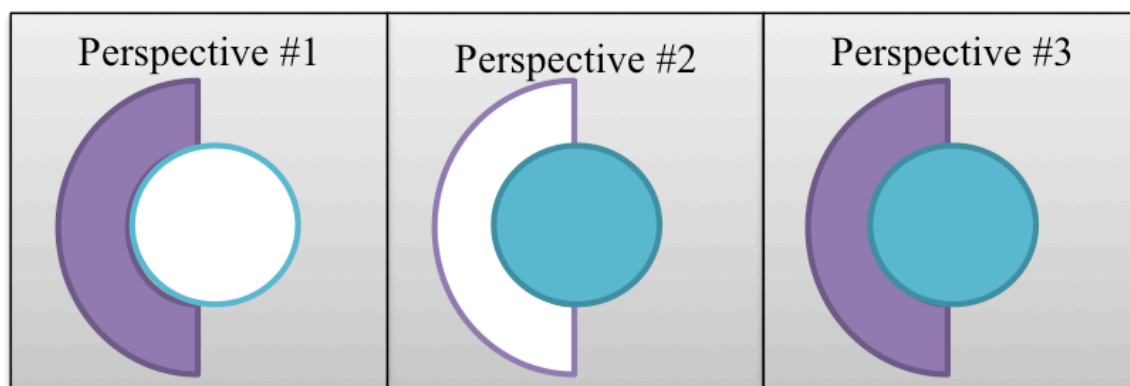


Figure 2.6.1: This figure represents how promiscuous proteins are considered in each perspective for the statistical analysis. Perspective #1 considered all of the promiscuous proteins. Perspective #2 considered all of the promiscuous protein binding partners. Perspective #3 considered the entire promiscuous protein complex.

For the promiscuous protein data set, the same promiscuous proteins were considered in complex with different binding partners. Except for cyclin dependent kinase 2, multiple structures of each promiscuous protein were examined. Consequently, the contributions of their residues were calculated multiple times, once for each binding partner. In order to avoid double counting the contributions of each residue, the contribution for each residue was chosen randomly from the multiple component analysis values that were obtained for that residue. This approach led to the discovery that the resulting p-values we observed fluctuated greatly depending on the random residues used. In order to mitigate the effect of these fluctuations and better understand the trends present in our data set, the random residues were generated a 1000 times for all three types of component analysis data, electrostatic, van der Waals, and SASA. The 1000 random trials were done separately for each perspective and the average proportions were used for the T-tests.

3. Results

Component analysis calculations were completed for the interfacial residues of all of the specific and promiscuous protein complexes in order to evaluate our hypotheses, i.e., if side chains are more important in specific protein binding than in promiscuous protein binding and if backbones are more important in promiscuous protein binding than in specific protein binding.

Sections 3.1-3.3 present the raw distributions of the data collected for the electrostatic, van der Waals, and SASA component analysis calculations. The plots on the left of each figure show the backbone contribution, while those on the right side of each figure display the side chain contribution. Three one tailed, two sample tests of proportions were done, each to evaluate one of the following hypotheses (termed “Test #1”, “Test #2”, and “Test #3” respectively):

1. Is the relative number of times the side chain contribution is higher than backbone contribution higher for specific proteins than for promiscuous proteins?
2. Is the proportion of substantial side chain contributions higher for specific proteins than for promiscuous proteins?
3. Is the proportion of substantial backbone contributions higher for promiscuous proteins than for specific proteins?

The results of the three statistical tests are discussed in sections 3.4-3.6, for each relevant contribution toward free energy in turn (electrostatic, vdW, and SASA). The raw proportions are shown in Tables 3.1-3.

3.1 Distribution of Electrostatic Component Analysis

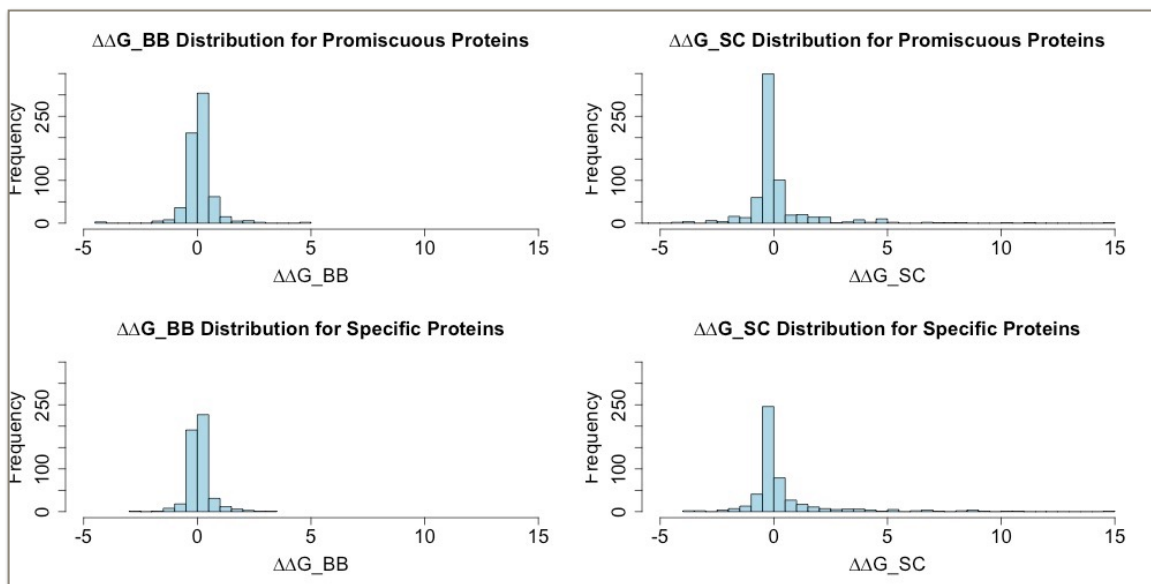


Figure 3.1.1: This figure shows the distribution of the electrostatic component analysis calculations for the interfacial residues of both promiscuous and specific protein complexes. 500 specific protein interfacial residues and 516 promiscuous protein interfacial residues were analyzed for this study. Note that a positive $\Delta\Delta G$ value means that the moiety contributes favorably to binding.

3.2 Distribution of van der Waals Component Analysis

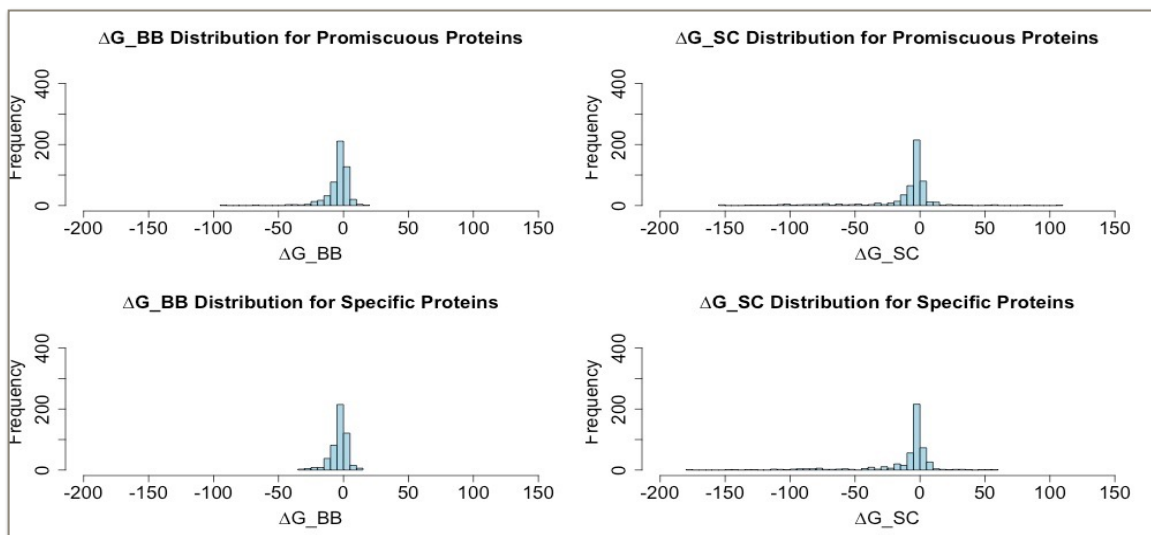


Figure 3.2.1: This figure shows the distribution of the van der Waals component analysis calculations for the interfacial residues of both promiscuous and specific protein complexes. 500 specific protein interfacial residues and 516 promiscuous protein interfacial residues were analyzed for this study. Note that a negative ΔG value means that the moiety contributes favorably to binding.

3.3 Distribution of SASA Component Analysis

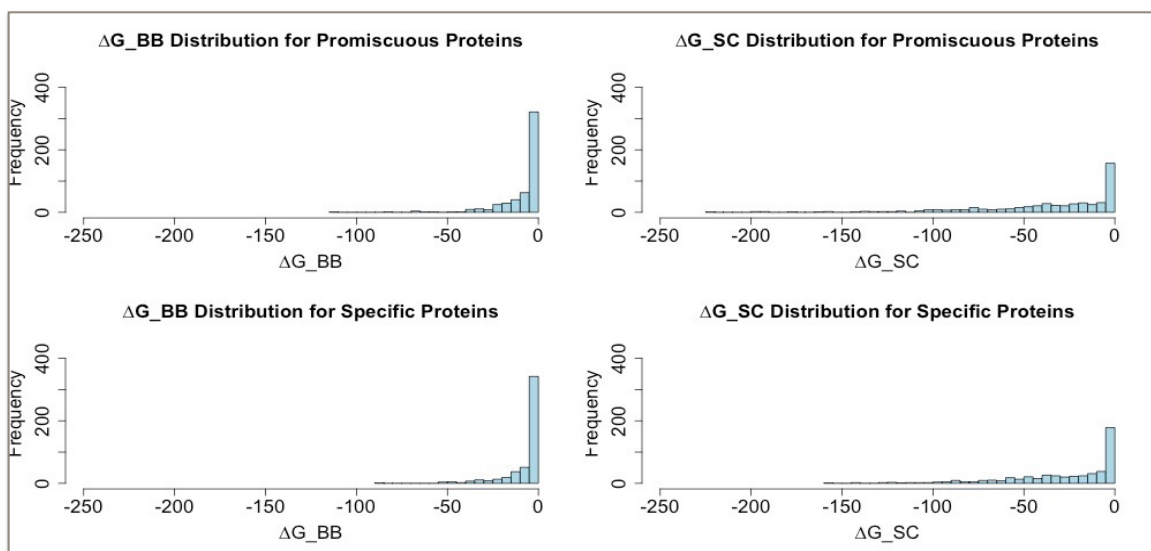


Figure 3.3.1: This figure shows the distribution of the SASA component analysis calculations for the interfacial residues of both promiscuous and specific protein complexes. 500 specific protein interfacial residues and 516 promiscuous protein interfacial residues were analyzed for this study. Note that a negative ΔG value means that the moiety contributes favorably to binding.

Electrostatics	Perspective #1			Perspective #2	Perspective #3			Specific
	Min	Max	Average		Min	Max	Average	
Test #1	0.373	0.510	0.439	0.443	0.418	0.465	0.441	0.480
Test #2	0.197	0.293	0.247	0.267	0.242	0.279	0.259	0.308
Test #3	0.242	0.354	0.293	0.321	0.286	0.329	0.310	0.22
Sample Size	198			318	516			500

Table 3.1: This table shows the raw proportions for each perspective for each statistical test for the electrostatics component analysis data. The minimum, maximum, and average proportions are shown for perspectives #1 and #3 because, as described in section 2.6, these proportions were observed to vary for different random residues.

van der Waals	Perspective #1			Perspective #2	Perspective #3			Specific
	Min	Max	Average		Min	Max	Average	
Test #1	0.5152	0.6364	0.5640	0.572	0.5446	0.5911	0.5692	0.556
Test #2	0.2071	0.2980	0.2531	0.405	0.3295	0.3643	0.3470	0.320
Test #3	0.1919	0.2828	0.2365	0.374	0.3043	0.3391	0.3215	0.290
Sample Size	198			318	516			500

Table 3.2: This table shows the raw proportions for each perspective for each statistical test for the van der Waals component analysis data. The minimum, maximum, and average proportions are shown for perspectives #1 and #3 because, as described in section 2.6, these proportions were observed to vary for different random residues.

SASA	Perspective #1			Perspective #2	Perspective #3			Specific
	Min	Max	Average		Min	Max	Average	
Test #1	0.5960	0.6768	0.6414	0.773	0.7093	0.7345	0.7228	0.686
Test #2	0.2020	0.2929	0.2504	0.437	0.3488	0.3818	0.3655	0.300
Test #3	0.1970	0.2677	0.2324	0.346	0.2888	0.3178	0.3026	0.238
Sample Size	198			318	516			500

Table 3.3: This table shows the raw proportions for each perspective for each statistical test for the SASA component analysis data. The minimum, maximum, and average proportions are shown for perspectives #1 and #3 because, as described in section 2.6, these proportions were observed to vary for different random residues.

3.4 Do specific proteins have side chain contributing more than backbone when compared to promiscuous proteins?

Do specific proteins have side chain contributing more than backbone, compared to promiscuous proteins? (Test #1)	Perspective #1	Perspective #2	Perspective #3
Electrostatics	0.186	0.168	0.118
Specific: 0.480	0.439	0.443	0.441
van der Waals	0.543	0.647	0.641
Specific: 0.556	0.5640	0.572	0.5692
SASA	0.149	0.996	0.890
Specific: 0.686	0.6414	0.773	0.723

Table 3.4.1: This table shows the p-values (bold) and proportions from the first t test. 500 interfacial residues were evaluated for the specific protein data set. 198, 318, and 516 promiscuous interfacial residues were evaluated for perspectives one, two, and three, respectively.

This table summarizes the results from the two sample test of proportions which examined the significance in the difference in the number of times the side chain contribution was greater than the backbone contribution between specific and promiscuous proteins. The null hypothesis is that the difference is less than or equal to 0, i.e., the specific proteins have a similar or lower side chain contribution when compared to promiscuous proteins. A low p value for the one tailed t test indicates specific proteins have a higher side chain contribution when compared to promiscuous proteins. A p value close to 1 suggests that the opposite may be true.

While most of the p-values are not considered statistically significant if we use the arbitrary 0.05 cutoff, the p-values for the electrostatics and SASA calculations from perspective #1 are small and may indicate a weak relationship between a higher side chain contribution, when compared to backbone contribution, in specific proteins when compared to promiscuous proteins. Likewise, the p-values for the SASA calculations for perspectives #2 and #3 are large, suggesting that the opposite of our hypothesis might be true in this case – that promiscuous proteins may have side chains contributing more than backbone, relative to specific proteins, although a two-tailed test of proportions should be done to verify this (and will be done as part of future work).

Note that the p values reported for perspectives #1 and #3 are the average p values over 1000 trials in each of which one instance of each residue within a promiscuous protein was randomly chosen from the multiple complexes that were modeled. There were large fluctuations in p values (for example 0.007*-0.7315 for the electrostatic component for Perspective #1), depending on the instances chosen, an issue that is discussed further in the Discussion.

3.5 Do specific proteins have more side chains that contribute substantially, compared to promiscuous proteins?

Do specific proteins have more side chains that contribute substantially when compared to promiscuous proteins? (Test #2)	Perspective #1	Perspective #2	Perspective #3
Electrostatics	0.066	0.119	0.048*
Specific: 0.308	0.247	0.267	0.259
van der Waals	0.049*	0.992	0.801
Specific: 0.320	0.2531	0.405	0.3470
SASA	0.112	1	0.984
Specific: 0.300	0.2504	0.437	0.3655

Table 3.4.2: This table shows the p-values(bold) and proportions from the second t-test. 500 interfacial residues were evaluated for the specific protein data set. 198, 318, and 516 promiscuous interfacial residues were evaluated for perspectives one, two, and three, respectively.

This table summarizes the results from the two sample test of proportions which examined whether the difference in the proportion of substantially favorable side chain contributions was significantly different between the specific and promiscuous protein data set. Here, the cutoff value was determined by assuming that the highest 30% of side chain contribution of the specific data quantified a substantial contribution. The electrostatic cutoff was determined to be 0.1142 kcal/mol. The van der Waals and SASA cutoffs were -5.802 and -37.1119 kcal/mol.

As can be observed in the table, the electrostatic p values, while not statistically significant except for Perspective #3, are small. Additionally, the van der Waals and SASA p values from perspective #1 are small as well, although interestingly, the p values for perspectives #2 and #3 are close to 1, suggesting that while the promiscuous proteins may have less significant side chain contributions than their specific counterparts, their binding partners may have *more* significant contributions than specific proteins (although this must be verified with a two-tailed statistical test).

3.6 Do promiscuous proteins have more backbones that contribute substantially, compared to specific proteins?

Do promiscuous proteins have more backbones that contribute substantially when compared to specific proteins? (Test #3)	Perspective #1	Perspective #2	Perspective #3
Electrostatics	0.026*	0.001*	0.001*
Specific: 0.22	0.293	0.321	0.310
van der Waals	0.909	0.008*	0.154
Specific: 0.290	0.2365	0.374	0.3215
SASA	0.523	0.001*	0.012*
Specific: 0.238	0.2324	0.346	0.3026

Table 3.4.3: This table shows the p-values (bold) and proportions from the third t-test. 500 interfacial residues were evaluated for the specific protein data set. 198, 318, and 516 promiscuous interfacial residues were evaluated for perspectives one, two, and three, respectively.

This table summarizes the results from the two sample test of proportions which examined whether the proportion of substantial backbone contributions was significantly different between the promiscuous and specific protein data set. Again, the cutoff value was determined by assuming that the highest 30% of backbone contribution of the promiscuous data quantified a substantial contribution. The electrostatic cutoff was determined to be 0.2 kcal/mol. The van der Waals and SASA cutoffs were -4.5 and -8.21 kcal/mol.

The p-values from the table above indicate that the promiscuous proteins have a substantial backbone contribution when compared to the specific proteins considering the electrostatics, van der Waals, and SASA energies, except in the case of perspective #1 vdW and SASA energies, again suggesting an interesting asymmetry in the energetics of recognition.

Nevertheless, taken together, these data might imply that in a certain subset of our tests, the side chain moieties make significantly more substantial contributions in specific proteins than in promiscuous proteins. Additionally, we find that in a larger subset of our tests, the promiscuous proteins have a significantly more substantial backbone contribution when compared to specific proteins, suggesting that the backbones may be important in promiscuous interactions. However, we do not find these trends through all perspectives and while considering all energies, and these data warrant careful scrutiny and further interpretation.

4. Discussion

For specific protein interactions, as predicted by our hypothesis, trends in the data reflect that electrostatic and SASA side chain contributions, through the first perspective, may be greater on average than the corresponding backbone contributions, although the data are not currently statistically significant. Additionally, the specific proteins appear to have substantially more side chain electrostatic and vdW contributions, but only when compared to certain perspectives of analyzing the promiscuous proteins, while from other perspectives, it is possible that the opposite may be true. We also found that from most, but not all, perspectives and energetic aspects (electrostatics, vdW, SASA), promiscuous proteins have a significantly substantial backbone contribution when compared to specific proteins. Taken together, this suggests that our original hypotheses may be partially supported, but there are some interesting subtleties that warrant further analysis.

For Tests #1 and #2, the first perspective was found to be most significant. This may imply the differences in molecular recognition between specific and promiscuous proteins may be due to interactions made by the promiscuous proteins. The high p values from the second and third perspectives imply the structural moieties of the promiscuous binding partners and specific proteins may have similar contributions, and in some cases, the promiscuous binding partners may be behaving more in line with our original characterization of specific proteins than the specific proteins themselves, although, as discussed below, additional statistical tests will need to be done to understand this better.

4.1 Limitations of the study

The goals of this study were to understand if structural characteristics of protein—protein complexes can preferentially contribute to specificity or promiscuity. The models used in this work to address these questions make both physical and situational assumptions, and understanding these approximations can help to contextualize our findings.

The protein complexes investigated in this study were identified to be either specific or promiscuous based on a literature search whenever possible. In this study, we define a specific protein as one that binds to only one partner. To our current knowledge, the specific complexes investigated in this study are thought to have one binding partner. However, in reality, it is difficult to prove that a protein will not bind to any other binding partners, so one must assume that the lack of evidence for binding other partners is sufficient to assume specificity, given that the proteins used have been widely studied. Furthermore, it is often difficult to quantify energetic contributions for promiscuous proteins with multiple partners. This is evidenced through the large fluctuations observed with the selection of random residues because the moiety contributions seem to be different for the moiety's interaction with each binding partner, which in itself is an interesting result. In this work, we generated selected one instance from each of the promiscuous protein residues 1000 times for each perspective and averaged the data to account for this shortcoming. In future studies, we hope to better understand whether the variations in residue contributions across partners depends on the nature of the interaction, and we also will keep the random set of residues for each trial constant across perspectives to allow for better interpretation of the results for each perspective. This change might allow us to better address whether the structural contributions from the

promiscuous protein or the binding partner (or both) are driving the promiscuity. Finally, this study was limited to studying protein—protein complexes for which an x-ray crystallography model exists. Consequently, the results may be biased toward proteins that can be crystallized, which may use a subset of all possible ways to mediate recognition, thus limiting the potential scope of the study.

The binding model used in this study assumes that proteins do not change conformation upon association. In reality, there is some change in the internal conformations of binding partners. The model implicitly accounts for this in a general sense by assuming the protein is polarizable (using a dielectric constant of 4), but large structural rearrangements may not be quantified accurately.

Finally, we employed a two sample test of proportions to analyze our data. One of the inherent assumptions made in this test is that the data are independent of each other. Consequently, we assumed that the component analysis values of neighboring residues were independent of each other. This assumption may be valid because neighboring residues in sequence are not necessarily interacting in structure. Accordingly, the interactions between neighboring residues may not be coupled. In future work, we can test this assumption by obtaining the correlation coefficients between the original dataset and the dataset when it is shifted by one residue (or in other ways to account for other potential types of coupling), and compare that with correlation coefficients of between randomly shuffled permutations of the dataset. Additionally, analysis from the second perspective utilized all of the data from the binding partners. Therefore, we assumed independence between corresponding residues on different partners in this case as well. However, different binding partners may make similar interactions with the promiscuous

protein, so relationships between their residues may exist as well. This question will be explored further in future work.

We chose to use a one-tailed t test for this study because we assumed there was a particular directionality in the data, i.e. the side chain contribution in specific proteins is higher than the side chain contribution in promiscuous proteins. The distributions of the data indicate that the trends present may oppose our original hypotheses. Future work will include a two-tailed t test to evaluate the significance of the observed opposing trends.

The results of Tests #2 and #3 are dependent on a cutoff that attempts to quantify a substantial contribution. We observe a lack of robustness of the results to the cutoff chosen: there is a range of p-values for different cutoffs. For example, for a substantial electrostatic contribution of 2 kcal/mol, the p-values for Test #2 are 0.02, 0.41, and 0.11 for perspectives #1, #2, and #3 (compare with table 3.4.2, 1st row). This test sought to evaluate how different the side chain contribution is between specific and promiscuous data sets. A better approach would be to do a continuous T test to evaluate the numerical difference between the side chain and backbone contribution, and this is a high priority in the near future. The lack of robustness also suggests that the distributions for promiscuous and specific energetic contributions may have qualitatively different shapes and tails. For example, considering the electrostatic, van der Waals, and SASA distributions, the variance in specific side chain contributions is higher than the variance in promiscuous side chain contributions. Additionally, the variance in the backbone contributions of promiscuous proteins appears to be greater when compared to specific data, in the van der Waals and SASA data set. Also from the van der Waals and SASA

data set, the promiscuous backbone contributions have a longer tail in the favorable contribution direction.

It should be noted that while the thesis was being written, bugs, an inevitable part of computational research, were found in the analysis scripts. Hence the p values and results presented are still to be considered preliminary. The next steps in this project will be to ensure the results found here can be independently validated by future researchers.

4.2 Further Analysis of Results:

Figure 3.2.1 shows the distributions of the van der Waals component analysis data. The distributions show that specific proteins appear to have a longer tail of ΔG_{SC} that are favorable when compared to promiscuous proteins. This finding supports our hypothesis. Similar to the electrostatics findings, the variance of ΔG_{BB} seems to be lower for specific proteins when compared to promiscuous proteins. ΔG_{BB} values appear to be more favorable for promiscuous proteins when compared to specific proteins. This finding also supports our hypothesis. Interestingly, the latter trend is statistically significant only for perspective #2 ($p=0.008$), with the opposite potentially being the case for perspective #1 ($p=0.909$). However, through the p-values for Test# 3 from the second and third perspectives, we find promiscuous proteins tend to have a higher substantial backbone contribution, when compared to specific proteins.

Figure 3.3.1 shows the distributions of the SASA component analysis data. From visually inspecting the distributions, it appears that promiscuous proteins appear to have a longer tail of ΔG_{SC} that are favorable when compared to specific proteins. This finding does not support our hypothesis. The corresponding p-values for Test #2 indeed seem to support the opposite hypothesis using perspectives #2 (p-value=1) and #3 (p-

value=0.984). ΔG_{BB} values appear to be more favorable for promiscuous proteins when compared to specific proteins. This finding supports our hypothesis. This visual trend is confirmed through the p-values for Test #3 from perspectives #2 and #3.

Some of the p-values from Test #1, including those in the electrostatic analyses, while not statistically significant, indicate a potentially differential side chain contribution from specific proteins when compared to promiscuous proteins that might warrant further study. For Test #1, the p-values appear to be close to 0.5 for the van der Waals data across all three perspectives ($p=0.543, 0.647, 0.641$). This may be due to the fact that the van der Waals calculations were done on the original crystal structure, which could have van der Waals clashes according to the CHARMM force field. We are currently performing the same calculations on the minimized structures to see if the van der Waals (and other) contributions change remarkably. Interestingly, from all three perspectives, the electrostatic p-values are lower than the van der Waals and SASA values (considering perspectives #2 and #3). This finding matches our expectations because electrostatic interactions are thought to drive specificity of protein—protein interactions [33].

4.2.1 Fluctuation of p-values due to random residue selection.

As discussed earlier, if there were multiple component analysis values for the same interfacial residue, one value was randomly picked for the statistical analysis in order to avoid double counting the contribution from the promiscuous proteins. Unexpectedly, we learned the p-values fluctuated greatly depending on the random residues used. These fluctuations may indicate that the contribution of the interfacial residues is dependent on the binding partner. A different subset of residues may

contribute more when binding to different partners. This finding is interesting because previous work suggests promiscuous proteins use common interfacial residues to bind to different binding partners [17, 37, 38]. An approach to address these fluctuations in future work could be to analyze more binding partners for each promiscuous protein and take the average residue contribution from multiple interactions.

4.3 Additional Future Work.

For the present study, we considered residues within 5 Å of the interface to be interfacial residues. It is possible that our cutoff was too high and our component analysis results may contain a large number of noncontributing residues, which may affect significance. However, the cutoff might be too low for including certain longer-range electrostatic interactions, which might play an important role in mediating interactions. Future researchers should consider how contribution of residues varies as a function of distance from the interface and pick a more appropriate cutoff, if applicable.

The component analysis results generated for the three types of binding free energy can be parsed through to find if certain trends correlate with specificity or promiscuity. For example, are polar amino acids or charged amino acids more significant in promiscuous proteins or in specific proteins? Additionally, the contribution of different structural moieties as a function of side chain length, molar mass, and pi stacking interaction can be considered. While it is believed that promiscuous proteins use the same set of amino acids to bind with their different partners, the promiscuous data set can also be analyzed to determine if there are particular amino acids in this set that are important in promiscuous proteins binding with their partners.

Another potential structural moiety of interest can be the entire backbone or side chain of the protein. Currently, we have started this analysis from the electrostatic perspective but future researchers should consider the contributions of the whole side chain or backbone of a protein from the van der Waals and electrostatic perspective.

Finally, in order to obtain a more robust data set, future research will consider more specific and promiscuous protein—protein interactions. Adding more data may also reduce the fluctuations observed due to picking random residues for promiscuous proteins.

4.4 Conclusions

Our findings suggest the specific proteins may have a higher side chain contribution when compared to promiscuous proteins, but only in certain ways. The low p-values observed for the electrostatics component analysis data for Test #1 may support our hypothesis that specific proteins preferentially use side chains, although the data are currently not statistically significant. The p-values from Test #1, perspective #2 suggest side chain contribution between specific proteins and promiscuous binding partners is similar. This may imply the intriguing result that the molecular recognition of promiscuous proteins is asymmetric between the protein itself and its binding partners, although further work is necessary to investigate this idea. We observed similar trends when considering a substantial side chain contribution (Test #2). Specific proteins appear to have more substantial side chain contributions when compared to the promiscuous proteins. Again, these trends do not hold for the binding partners. Finally, we find the promiscuous binding partners appear to have a substantial backbone contribution when compared to specific proteins. Interestingly, we do not find a difference in a substantial backbone contribution when we compare specific proteins and the promiscuous proteins themselves. This potential asymmetry may have implications regarding promiscuous protein molecular recognition.

5. Works Cited

1. Koury, M.J. and M.C. Bondurant, *The molecular mechanism of erythropoietin action*. European Journal of Biochemistry, 1992. **210**(3): p. 649-663.
2. Skehel, J.J. and D.C. Wiley, *Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin*. Annu Rev Biochem, 2000. **69**: p. 531-69.
3. Cheung, W.Y., *Calmodulin plays a pivotal role in cellular regulation*. Science, 1980. **207**(4426): p. 19-27.
4. Schreiber, R.D., L.J. Old, and M.J. Smyth, *Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion*. Science, 2011. **331**(6024): p. 1565-70.
5. Lapenna, S. and A. Giordano, *Cell cycle kinases as therapeutic targets for cancer*. Nat Rev Drug Discov, 2009. **8**(7): p. 547-66.
6. Hard, T. and C. Lendel, *Inhibition of amyloid formation*. J Mol Biol, 2012. **421**(4-5): p. 441-65.
7. Eisenberg, D. and M. Jucker, *The amyloid state of proteins in human diseases*. Cell, 2012. **148**(6): p. 1188-203.
8. Kastiris, P.L. and A.M.J.J. Bonvin, *On the binding affinity of macromolecular interactions: daring to ask why proteins interact*. Vol. 10. 2013.
9. Silverstein, T.P., *The Real Reason Why Oil and Water Don't Mix*. Journal of Chemical Education, 1998. **75**(1): p. 116.
10. Tsai, C.-J., D. Xu, and R. Nussinov, *Structural motifs at protein-protein interfaces: Protein cores versus two-state and three-state model complexes*. Protein Science, 1997. **6**(9): p. 1793-1805.
11. Tsai, C.-J., et al., *Studies of protein-protein interfaces: A statistical analysis of the hydrophobic effect*. Protein Science, 1997. **6**(1): p. 53-64.
12. Kauzmann, W., *Some factors in the interpretation of protein denaturation*. Adv Protein Chem, 1959. **14**: p. 1-63.
13. Chen, J., N. Sawyer, and L. Regan, *Protein-protein interactions: General trends in the relationship between binding affinity and interfacial buried surface area*. Protein Science, 2013. **22**(4): p. 510-515.
14. Jones, S. and J.M. Thornton, *Principles of protein-protein interactions*. Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(1): p. 13-20.
15. Clackson, T. and J.A. Wells, *A hot spot of binding energy in a hormone-receptor interface*. Science, 1995. **267**(5196): p. 383-6.
16. Thorn, K.S. and A.A. Bogan, *ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions*. Bioinformatics, 2001. **17**(3): p. 284-5.
17. Bogan, A.A. and K.S. Thorn, *Anatomy of hot spots in protein interfaces*. J Mol Biol, 1998. **280**(1): p. 1-9.
18. Roth, C.M., B.L. Neal, and A.M. Lenhoff, *Van der Waals interactions involving proteins*. Biophysical Journal, 1996. **70**(2): p. 977-987.

19. McCoy, A.J., V. Chandana Epa, and P.M. Colman, *Electrostatic complementarity at protein/protein interfaces*. J Mol Biol, 1997. **268**(2): p. 570-84.
20. Prasad Bahadur, R., et al., *A Dissection of Specific and Non-specific Protein-Protein Interfaces*. Journal of Molecular Biology, 2004. **336**(4): p. 943-955.
21. Massova, I. and P.A. Kollman, *Computational Alanine Scanning To Probe Protein-Protein Interactions: A Novel Approach To Evaluate Binding Free Energies*. Journal of the American Chemical Society, 1999. **121**(36): p. 8133-8143.
22. Xu, D., S.L. Lin, and R. Nussinov, *Protein binding versus protein folding: the role of hydrophilic bridges in protein associations*. J Mol Biol, 1997. **265**(1): p. 68-84.
23. Sheinerman, F.B., R. Norel, and B. Honig, *Electrostatic aspects of protein-protein interactions*. Curr Opin Struct Biol, 2000. **10**(2): p. 153-9.
24. Hendsch, Z.S. and B. Tidor, *Do salt bridges stabilize proteins? A continuum electrostatic analysis*. Protein Sci, 1994. **3**(2): p. 211-26.
25. Kundrotas, P.J. and E. Alexov, *Electrostatic Properties of Protein-Protein Complexes*. Biophysical Journal, 2006. **91**(5): p. 1724-1736.
26. Green, D.F. and B. Tidor, *Design of improved protein inhibitors of HIV-1 cell entry: Optimization of electrostatic interactions at the binding interface*. Proteins, 2005. **60**(4): p. 644-57.
27. Kangas, E. and B. Tidor, *Charge optimization leads to favorable electrostatic binding free energy*. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics, 1999. **59**(5 Pt B): p. 5958-61.
28. Frieden, E., *Non-covalent interactions: Key to biological flexibility and specificity*. Journal of Chemical Education, 1975. **52**(12): p. 754.
29. Schreiber, G. and A.E. Keating, *Protein binding specificity versus promiscuity*. Curr Opin Struct Biol, 2011. **21**(1): p. 50-61.
30. Stevens, F.C., *Calmodulin: an introduction*. Canadian Journal of Biochemistry and Cell Biology, 1983. **61**(8): p. 906-910.
31. Halperin, I., et al., *Principles of docking: An overview of search algorithms and a guide to scoring functions*. Proteins, 2002. **47**(4): p. 409-43.
32. Camacho, C.J., H. Ma, and P.C. Champ, *Scoring a diverse set of high-quality docked conformations: a metascore based on electrostatic and desolvation interactions*. Proteins, 2006. **63**(4): p. 868-77.
33. Radhakrishnan, M.L. and B. Tidor, *Specificity in molecular design: a physical framework for probing the determinants of binding specificity and promiscuity in a biological environment*. J Phys Chem B, 2007. **111**(47): p. 13419-35.
34. Fellouse, F.A., et al., *Molecular recognition by a binary code*. J Mol Biol, 2005. **348**(5): p. 1153-62.
35. Birtalan, S., R.D. Fisher, and S.S. Sidhu, *The functional capacity of the natural amino acids for molecular recognition*. Mol Biosyst, 2010. **6**(7): p. 1186-94.
36. James, L.C. and D.S. Tawfik, *The specificity of cross-reactivity: promiscuous antibody binding involves specific hydrogen bonds rather than nonspecific hydrophobic stickiness*. Protein Sci, 2003. **12**(10): p. 2183-93.

37. Martin, J., *Beauty Is in the Eye of the Beholder: Proteins Can Recognize Binding Sites of Homologous Proteins in More than One Way*. PLoS Computational Biology, 2010. **6**(6): p. e1000821.
38. Zhang, Y.Y., *Physical Determinants of Promiscuity and Specificity*. 2013.
39. Kieslich, C.A., R.D. Gorham Jr, and D. Morikis, *Is the rigid-body assumption reasonable?: Insights into the effects of dynamics on the electrostatic analysis of barnase–barstar*. Journal of Non-Crystalline Solids, 2011. **357**(2): p. 707-716.
40. Radhakrishnan, M., *Designing electrostatic interactions in biological systems via charge optimization or combinatorial approaches: insights and challenges with a continuum electrostatic framework*. Theoretical Chemistry Accounts, 2012. **131**(8): p. 1-24.
41. Huang, A. and C.M. Stultz, *Conformational Sampling with Implicit Solvent Models: Application to the PHF6 Peptide in Tau Protein*. Biophysical Journal, 2007. **92**(1): p. 34-45.
42. Morozov, A.V., T. Kortemme, and D. Baker, *Evaluation of Models of Electrostatic Interactions in Proteins*. The Journal of Physical Chemistry B, 2003. **107**(9): p. 2075-2090.
43. Gilson, M.K. and B.H. Honig, *The dielectric constant of a folded protein*. Biopolymers, 1986. **25**(11): p. 2097-2119.
44. Gilson, M.K., *Introduction to Continuum Electrostatics, with Molecular Applications*, in *Self-Published at* <http://www.biophysics.org/Portals/1/PDFs/Education/gilson.pdf%3E>. 2000.
45. Nechamkin, E., *Investigating the Structural Determinants of Electrostatic Binding among Protein-protein Complexes: A Systematic, Large-scale Computational Study*, in *Chemistry*. 2012, Wellesley College: Wellesley, MA. p. 80.
46. Carrascal, N. and D.F. Green, *Energetic Decomposition with the Generalized-Born and Poisson–Boltzmann Solvent Models: Lessons from Association of G-Protein Components*. The Journal of Physical Chemistry B, 2010. **114**(15): p. 5096-5116.
47. Sharp, K.A. and B. Honig, *Electrostatic Interactions in Macromolecules: Theory and Applications*. Annual Review of Biophysics and Biophysical Chemistry, 1990. **19**(1): p. 301-332.
48. Qi, H.W., et al., *The Effect of Macromolecular Crowding on the Electrostatic Component of Barnase–Barstar Binding: A Computational, Implicit Solvent-Based Study*. PLoS ONE, 2014. **9**(6): p. e98618.
49. lee, L.-P. and B. Tidor, *Optimization of binding electrostatics: Charge complementarity in the barnase-barstar protein complex*. Protein Science : A Publication of the Protein Society, 2001. **10**(2): p. 362-377.
50. Minkara, M.S., P.H. Davis, and M.L. Radhakrishnan, *Multiple drugs and multiple targets: an analysis of the electrostatic determinants of binding between non-nucleoside HIV-1 reverse transcriptase inhibitors and variants of HIV-1 RT*. Proteins, 2012. **80**(2): p. 573-90.

51. Midelfort, K.S., et al., *Substantial energetic improvement with minimal structural perturbation in a high affinity mutant antibody*. J Mol Biol, 2004. **343**(3): p. 685-701.
52. Hendsch, Z.S. and B. Tidor, *Electrostatic interactions in the GCN4 leucine zipper: substantial contributions arise from intramolecular interactions enhanced on binding*. Protein Science : A Publication of the Protein Society, 1999. **8**(7): p. 1381-1392.
53. MacKerell, A.D., et al., *All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins*. The Journal of Physical Chemistry B, 1998. **102**(18): p. 3586-3616.
54. Brooks, B.R., et al., *CHARMM: A program for macromolecular energy, minimization, and dynamics calculations*. Journal of Computational Chemistry, 1983. **4**(2): p. 187-217.
55. Brooks, B.R., et al., *CHARMM: The biomolecular simulation program*. Journal of Computational Chemistry, 2009. **30**(10): p. 1545-1614.
56. Chothia, C. and J. Janin, *Principles of protein-protein recognition*.
57. Eisenberg, D. and A.D. McLachlan, *Solvation energy in protein folding and binding*.
58. Horton, N. and M. Lewis, *Calculation of the free energy of association for protein complexes*. Protein science : a publication of the Protein Society, 1992. **1**(1): p. 169-181.
59. Privalov, P.L., *Stability of proteins: small globular proteins*. Adv Protein Chem, 1979. **33**: p. 167-241.
60. Noble, C.G., et al., *Ligand-bound structures of the dengue virus protease reveal the active conformation*. J Virol, 2012. **86**(1): p. 438-46.
61. Helland, R., et al., *The crystal structures of the complexes between bovine beta-trypsin and ten P1 variants of BPTI*. J Mol Biol, 1999. **287**(5): p. 923-42.
62. Zhang, Q.C., et al., *PrePPI: a structure-informed database of protein-protein interactions*. Nucleic Acids Res, 2013. **41**(Database issue): p. D828-33.
63. Wahlgren, W.Y., et al., *The catalytic aspartate is protonated in the Michaelis complex formed between trypsin and an in vitro evolved substrate-like inhibitor: a refined mechanism of serine protease action*. J Biol Chem, 2011. **286**(5): p. 3587-96.
64. Millers, E.-K.I., et al., *Crystal structure of textilinin-1, a Kunitz-type serine protease inhibitor from the venom of the Australian common brown snake (Pseudonaja textilis)*. FEBS Journal, 2009. **276**(11): p. 3163-3175.
65. Scheidig, A.J., et al., *Crystal structures of bovine chymotrypsin and trypsin complexed to the inhibitor domain of Alzheimer's amyloid beta-protein precursor (APPI) and basic pancreatic trypsin inhibitor (BPTI): engineering of inhibitors with altered specificities*. Protein Sci, 1997. **6**(9): p. 1806-24.
66. Wall, M.E., J.B. Clarage, and G.N. Phillips, *Motions of calmodulin characterized using both Bragg and diffuse X-ray scattering*. Structure, 1997. **5**(12): p. 1599-612.
67. Maximciuc, A.A., et al., *Complex of calmodulin with a ryanodine receptor target reveals a novel, flexible binding mode*. Structure, 2006. **14**(10): p. 1547-56.

68. Fallon, J.L., et al., *Structure of calmodulin bound to the hydrophobic IQ domain of the cardiac Ca(v)1.2 calcium channel*. *Structure*, 2005. **13**(12): p. 1881-6.
69. Ataman, Z.A., et al., *The NMDA receptor NR1 C1 region bound to calmodulin: structural insights into functional differences between homologous domains*. *Structure*, 2007. **15**(12): p. 1603-17.
70. Valentine, K.G., et al., *Crystal structure of calmodulin-neuronal nitric oxide synthase complex*. To be Published.
71. Simonovic, M., et al., *Structure of the calmodulin alphaII-spectrin complex provides insight into the regulation of cell plasticity*. *J Biol Chem*, 2006. **281**(45): p. 34333-40.
72. Pavletich, N.P., *Mechanisms of cyclin-dependent kinase regulation: structures of cdk, their cyclin activators, and cip and INK4 inhibitors 1,2*. *Journal of Molecular Biology*, 1999. **287**(5): p. 821-828.
73. Bourne, Y., et al., *Crystal structure and mutational analysis of the human CDK2 kinase complex with cell cycle-regulatory protein CksHs1*. *Cell*, 1996. **84**(6): p. 863-74.
74. Lim, D., et al., *Crystal structure and kinetic analysis of beta-lactamase inhibitor protein-II in complex with TEM-1 beta-lactamase*. *Nat Struct Biol*, 2001. **8**(10): p. 848-52.
75. Orchard, S., et al., *The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases*. *Nucleic Acids Res*, 2014. **42**(Database issue): p. D358-63.
76. Buckle, A.M., G. Schreiber, and A.R. Fersht, *Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-A resolution*. *Biochemistry*, 1994. **33**(30): p. 8878-89.
77. Syed, R.S., et al., *Efficiency of signalling through cytokine receptors depends critically on receptor orientation*. *Nature*, 1998. **395**(6701): p. 511-6.
78. Coley, A.M., et al., *Structure of the malaria antigen AMA1 in complex with a growth-inhibitory antibody*. *PLoS Pathog*, 2007. **3**(9): p. 1308-19.
79. Argiriadi, M.A., et al., *Unusual water-mediated antigenic recognition of the proinflammatory cytokine interleukin-18*. *J Biol Chem*, 2009. **284**(36): p. 24478-89.
80. Ultsch, M., et al., *Structural basis of signaling blockade by anti-IL-13 antibody Lebrikizumab*. *J Mol Biol*, 2013. **425**(8): p. 1330-9.
81. Papageorgiou, A.C., R. Shapiro, and K.R. Acharya, *Molecular recognition of human angiogenin by placental ribonuclease inhibitor--an X-ray crystallographic study at 2.0 A resolution*. *Embo j*, 1997. **16**(17): p. 5162-77.
82. Cohen, G.H., et al., *Water molecules in the antibody-antigen interface of the structure of the Fab HyHEL-5-lysozyme complex at 1.7 A resolution: comparison with results from isothermal titration calorimetry*. *Acta Crystallogr D Biol Crystallogr*, 2005. **61**(Pt 5): p. 628-33.
83. Desmyter, A., et al., *Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme*. *Nat Struct Biol*, 1996. **3**(9): p. 803-11.
84. DePristo, M.A., P.I.W. de Bakker, and T.L. Blundell, *Heterogeneity and Inaccuracy in Protein Structures Solved by X-Ray Crystallography*. *Structure*, 2004. **12**(5): p. 831-838.

85. Berman, H.M., et al., *The Protein Data Bank*. Nucleic Acids Research, 2000. **28**(1): p. 235-242.
86. McLachlan, A., *Rapid comparison of protein structures*. Acta Crystallographica Section A, 1982. **38**(6): p. 871-873.
87. Weichenberger, C.X. and M.J. Sippl, *NQ-Flipper: validation and correction of asparagine/glutamine amide rotamers in protein crystal structures*. Bioinformatics, 2006. **22**(11): p. 1397-1398.
88. Brünger, A.T. and M. Karplus, *Polar hydrogen positions in proteins: Empirical energy placement and neutron diffraction comparison*. Proteins: Structure, Function, and Bioinformatics, 1988. **4**(2): p. 148-156.
89. Fiser, A. and A. Šali, *Modeller: Generation and Refinement of Homology-Based Protein Structure Models*, in *Methods in Enzymology*, Charles W. Carter, Jr. and M.S. Robert, Editors. 2003, Academic Press. p. 461-491.
90. Sitkoff, D., K.A. Sharp, and B. Honig, *Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models*. The Journal of Physical Chemistry, 1994. **98**(7): p. 1978-1988.
91. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics*. J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.
92. Altman, M.D., *Computational ligand design and analysis in protein complexes using inverse methods, combinatorial search, and accurate solvation modeling*. 2006, Massachusetts Institute of Technology.: Cambridge.
93. Team, R.D.C., *R: A language and environment for statistical computing* 2010, R Foundation for Statistical Computing: Vienna, Austria.