UC Riverside
UC Riverside Electronic Theses and Dissertations

Title
Mechanistic Exploration of Complement System Interactions and Design of Complement-Targeted Therapeutics

Permalink
https://escholarship.org/uc/item/6cw475n3

Author
Gorham Jr, Ronald Dennis

Publication Date
2013

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Mechanistic Exploration of Complement System Interactions and Design of Complement-Targeted Therapeutics

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Bioengineering

by

Ronald Dennis Gorham Jr.

August 2013

Dissertation Committee
  Dr. Dimitrios Morikis, Chairperson
  Dr. David D. Lo
  Dr. Jiayu Liao
  Dr. Chia-en A. Chang

Copyright by
Ronald Dennis Gorham Jr.
2013

The Dissertation of Ronald Dennis Gorham Jr. is approved:

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the academic and personal support of many individuals. I first would like to thank my advisor, Dr. Dimitrios Morikis, for his unending support during my PhD. I could not have asked for a better advisor. You have always made yourself available to discuss research and brainstorm, any time of the day, but yet allowed me pursue my own interests and become independent. I also would like to thank my committee members (Dr. David Lo, Dr. Jiayu Liao, and Dr. Chia-en Chang) for their ideas and support during my PhD. Also, thank you Dr. Lo and Dr. Liao for giving me the opportunity to get my hands wet in the lab here at UCR.

I am exceedingly grateful for all the opportunities I have had during my PhD, many of which most others never have the luck to experience. I would like to acknowledge the UCR Graduate Research Mentorship Program, UCR Dissertation Year Fellowship, NSF East Asia and Pacific Summer Institute Fellowship, Utrecht University Short-Stay Fellowship, and the Whitaker International Fellowship for providing funding during my dissertation research, and affording the opportunity for scientific, professional, personal, and cultural development, both in Riverside and abroad. Thank you to Dr. Jung-Hsin Lin at Academia Sinica in Taipei for the chance to work with you during Summer 2011. I really enjoyed my experience there, and learned a great deal about virtual screening (in only 2 months) that really helped in my research. A big thank you to Drs. Suzan Rooijakkers and Jos van Strijp at UMC Utrecht for adopting me as a PhD student for an entire year. Although research didn’t completely work out as planned, I ended up learning so many new things, in terms of both skills and scientific knowledge, and I am confident that my stay in Holland will help me during my future endeavors.

On a personal level, there are countless people that have helped me during my PhD, and have made the journey very enjoyable! First of all, my fellow BioMoDeL members helped me through the ups and downs of my PhD. Drs. Chris Kieslich and Aliana López De Victoria, you made the whole process much more fun, and I am grateful for having labmates that could also be close friends. I hope for many BioMoDeL reunions to come! I also thank the newest BioMoDeLer, Zied Gaieb, for helping me get through my last year. When I returned to Riverside last August, we didn’t know each other, but throughout the last year, we restored the cohesiveness of our group and brainstormed many ideas together. Thanks to everyone else who has been a part of the lab over the years, and especially to Dr. Phanos Tamamis, for your tremendous help in many of my projects.

Thanks to everyone in the Bioengineering Department at UCR, especially my fellow classmates. We helped each other through classes and qualifiers in the beginning, and now five years later, we are ready to move on and change the world! One of the reasons I chose to come to UCR Bioengineering in the first place was the close-knit community, and I enjoyed all of the fun times we shared together. I would like especially thank Dr. Victor Rodgers for his support throughout my PhD. Even though he wasn't my own advisor, he believed in me and always pushed me to accomplish things I didn't otherwise think I was capable of. Thanks to Denise Sanders for helping me with so many things throughout my PhD, including anything from handling simple questions and paperwork, to giving me heads up with regards to countless professional opportunities, and just being there when I needed someone to talk to.

I also thank the entire MMB Department at UMC Utrecht for making me feel at home while in Holland. Thanks to my fellow labmates Evelien Berends, Daphne Stapels, Annemarie Kuipers, Maartje Ruyken, and everyone else for helping me learn countless new techniques in the lab. Thanks to all my other friends and colleagues in Utrecht (especially Rutger Luteijn, Sue Aitken, and Michiel van Gent), for being nice to "Ron the American" and making me feel welcome. We had so many good times and I will never forget you guys!

I would like to thank my best friend Sandeep Dhall for all the good times and memories. We made it through together, and I am glad that we have become great friends along the way. Thanks for everything sire! Thank you so much to my girlfriend Kaila Bennett. I honestly can't imagine my life without you, and I am so happy we met in BIEN 120. Thank you for always being there for me, and calming me during my most stressful times, and helping me progress on a scientific level as well. Also, thanks to my adopted family, The Bennett's, for treating me like one of your own and all the things you have done for me. Finally, I want to thank my family, especially my parents, for all of your hard work and support to get me to where I am today.

The text of this dissertation, in part, is a reprint of the material as it appears in:

1. Gorham, R.D., Jr, C.A. Kieslich, A. Nichols, N.U. Sausman, M. Foronda, et al. 2011. An evaluation of poisson-boltzmann electrostatic free energy calculations through comparison with experimental mutagenesis data. Biopolymers. 95: 746–754.
2. Gorham, R., Jr, C. Kieslich, and D. Morikis. 2011. Complement Inhibition by Staphylococcus aureus: Electrostatics of C3d–EfbC and C3d–Ehp Association. Cel. Mol. Bioeng. 5: 32–43.

The co-author Dimitrios Morikis directed and supervised the research which forms the basis for this dissertation. Other co-authors listed provided methodological and technical expertise.

DEDICATION

I dedicate this dissertation to

my loving girlfriend Kaila Bennett,

and my parents,

for always believing in me,

and pushing me to reach for the stars.

ABSTRACT OF THE DISSERTATION

Mechanistic Exploration of Complement System Interactions and Design of Complement-Targeted Therapeutics

by

Ronald Dennis Gorham Jr.

Doctor of Philosophy, Graduate Program in Bioengineering
University of California, Riverside, August 2013
Dr. Dimitrios Morikis, Chairperson

Protein-protein interactions play a crucial role in most biological functions. Structural biology provides a glimpse into the atomic world, helping to further understand the relation between protein structure and function. Computational methods hold promise for understanding the dynamics of proteins and protein-protein interactions, but require careful selection of tools and parameters that are applicable for probing a phenomenon of interest. We describe the utility of Poisson-Boltzmann electrostatic calculations and their applicability in analysis of protein-protein interactions and protein design. We discuss a benchmark of these calculations against experimental mutagenesis data, in order to choose parameters appropriate for calculating free energies of protein association. Poisson-Boltzmann electrostatic calculations, in conjunction with other computational methods, were used to better understand the molecular mechanism by which *Staphylococcus aureus* evades the complement system. Recent crystallographic structures and experimental work provided insight into the molecular interactions between three secreted staphylococcal virulence factors (Efb, Ecb, and Sbi) and complement protein C3d. Our work elucidated specific residues of Efb and Ecb crucial for C3d binding, and suggested templates for the design of C3d- and Efb/Ecb-derived peptides for therapeutic design. We also used electrostatic calculations and molecular and Brownian dynamics simulations to investigate two distinct binding modes of Sbi (domain IV) to C3d, and provide insight into the physiological contexts in which each binding site may play a role.

Finally, we used these data as a basis for therapeutic design. We used molecular dynamics simulations to assess rationally designed peptides aimed at competitively inhibiting interactions between complement C3d and its host and pathogenic ligands. We also describe a comprehensive approach for virtual high-throughput screening of small molecules that can bind to C3d. Our data provide frameworks for the analysis of host-pathogen interactions and drug design, and identify several potential C3d-binding molecules that serve as a foundation for the design of complement-targeted and anti-infective therapeutics.

TABLE OF CONTENTS

Chapter 1: Introduction **1**
- 1.1 The complement system 1
- 1.2 Complement and autoimmunity 4
- 1.3 Pathogenic immune evasion 4
- 1.4 Complement-targeted and anti-infective therapeutics 5
- 1.5 Structural biology of the complement system 6
- 1.6 Electrostatic analysis of protein-protein interactions 7
- 1.7 Structure-based drug design 8
- 1.8 Overview 10
- 1.9 References 11

Chapter 2: An Evaluation of Poisson-Boltzmann Electrostatic Free Energy Calculations Through Comparison with Experimental Mutagenesis Data **18**
- 2.1 Introduction 18
- 2.2 Methods 20
- 2.3 Results 25
- 2.4 Discussion 27
- 2.5 References 33

Chapter 3: Complement Inhibition By *Staphylococcus aureus*: Electrostatics of C3d-Efb-C and C3d-Ecb Association **38**
- 3.1 Introduction 38
- 3.2 Methods 39
- 3.3 Results 43
- 3.4 Discussion 50
- 3.5 Concluding Remarks 53
- 3.6 References 54

Chapter 4: Molecular Analysis of the Interaction Between Staphylococcal Virulence Factor Sbi-IV and Complement C3d **58**
- 4.1 Introduction 58
- 4.2 Methods 59
- 4.3 Results 62
- 4.4 Discussion 66
- 4.5 References 69

Chapter 5: Identification of Low Molecular Mass Inhibitors of C3d-Ligand Interactions: A Step Toward New Complement-Targeted Therapeutics **72**
- 5.1 Introduction 72
- 5.2 Methods 73
- 5.3 Results and Discussion 79
- 5.4 References 95

Chapter 6: Perspectives and Future Directions **98**

Appendices

- Appendix A: Theoretical Basis for Computational Methods 105
- Appendix B: Supporting Data for Chapter 2 112
- Appendix C: Supporting Data for Chapter 3 126
- Appendix D: Supporting Data for Chapter 4 137
- Appendix E: Supporting Data for Chapter 5 144
- Appendix F: Experimental Testing of a New Generation of Compstatin-Family Peptides 152
- Appendix G: Protocols for Functional Assays of Complement Inhibition 160
- Appendix H: Curriculum Vitae 166

LIST OF FIGURES

Chapter 1: Introduction

1.1 Overview of complement activation and regulation 2
1.2 Amplification loop of the alternative pathway 3

Chapter 2: An Evaluation of Poisson-Boltzmann Electrostatic Free Energy Calculations Through Comparison with Experimental Mutagenesis Data

2.1 Thermodynamic cycle for association and solvation of protein complexes 23
2.2 Molecular graphics representations of selected protein complexes 25

Chapter 3: Complement Inhibition By *Staphylococcus aureus*: Electrostatics of C3d-Efb-C and C3d-Ecb Association

3.1 Electrostatically-driven association of C3d-Efb-C and C3d-Ecb complexes 39
3.2 Molecular graphics representations of C3d-Efb-C and C3d-Ecb 45
3.3 Clustering and free energy results for C3d-Efb-C 46
3.4 Clustering and free energy results for C3d-Ecb 47

Chapter 4: Molecular Analysis of the Interaction Between Staphylococcal Virulence Factor Sbi-IV and Complement C3d

4.1 Molecular graphics representations of C3d-Sbi-IV complexes 62
4.2 Change in interfacial SASA during MD trajectories of C3d-Sbi-IV complexes 63
4.3 Interaction occupancy during MD trajectories of C3d-Sbi-IV complexes 64
4.4 Comparison of association rates of C3d-Sbi-IV complexes 65
4.5 Unbinding forces for C3d-Sbi-IV complexes 65

Chapter 5: Identification of Low Molecular Mass Inhibitors of C3d-Ligand Interactions: A Step Toward New Complement-Targeted Therapeutics

5.1 Molecular graphics representations of the C3d-derived peptide (C3dp) template 78
5.2 Quantitative comparison of C3dpv4 structural stability and binding 80
5.3 Quantitative comparison of C3dpv8 binding 81
5.4 Molecular graphics representations of the Efb- and Ecb-derived peptides 82
5.5 Experimental data for Efb- and Ecb-derived peptides 83
5.6 Molecular graphics and details of Virtual Screen A (pharmacophores and docking) 84
5.7 Microscale thermophoresis experiments for direct molecule binding to C3d 85
5.8 Overview of improved computational protocol for Virtual Screen B 86
5.9 Electrostatically-driven association of C3d and CR2 87
5.10 Analysis of interactions between C3d and CR2 88
5.11 Development of a dynamic pharmacophore model for C3d-CR2 89
5.12 C3d receptor ensemble from MD simulation and RMSD-based clustering 92

Appendix B: Supporting Data for Chapter 2

B.1 Scatter plots of calculated and experimental data for Barnase-Barstar (0 mM) 116
B.2 Scatter plots of calculated and experimental data for Barnase-Barstar (150 mM) 117
B.3 Scatter plots of calculated and experimental data for Im9-E9 DNase (0 mM) 118
B.4 Scatter plots of calculated and experimental data for Im9-E9 DNase (150 mM) 119
B.5 Scatter plots of calculated and experimental data for RNase Inhibitor-Angiogenin (0 mM) 120
B.6 Scatter plots of calculated and experimental data for RNase Inhibitor-Angiogenin (150 mM) 121
B.7 Scatter plots of calculated and experimental data for hGH-hGHbp (0 mM) 122
B.8 Scatter plots of calculated and experimental data for hGH-hGHbp (150 mM) 123
B.9 Scatter plots of calculated and experimental data for gp120-CD4 (0 mM) 124
B.10 Scatter plots of calculated and experimental data for gp120-CD4 (150 mM) 125

Appendix C: Supporting Data for Chapter 3

C.1 Clustering of C3d mutants (from C3d-Efb-C complex) at 0 mM ionic strength 129
C.2 $\Delta G^{Coulombic}$ values for C3d mutants (from C3d-Efb-C complex) at 0 mM ionic strength 129
C.3 $\Delta G^{solution}$ values for C3d mutants (from C3d-Efb-C complex) at 0 mM ionic strength 130
C.4 $\Delta\Delta G^{solvation}$ values for C3d mutants (from C3d-Efb-C complex) at 0 mM ionic strength 130
C.5 Clustering of C3d mutants (from C3d-Efb-C complex) at 150 mM ionic strength 131
C.6 $\Delta G^{Coulombic}$ values for C3d mutants (from C3d-Efb-C complex) at 150 mM ionic strength 131
C.7 $\Delta G^{solution}$ values for C3d mutants (from C3d-Efb-C complex) at 150 mM ionic strength 132
C.8 $\Delta\Delta G^{solvation}$ values for C3d mutants (from C3d-Efb-C complex) at 150 mM ionic strength 132
C.9 Clustering of C3d mutants (from C3d-Ecb complex) at 0 mM ionic strength 133
C.10 $\Delta G^{Coulombic}$ values for C3d mutants (from C3d-Ecb complex) at 0 mM ionic strength 133
C.11 $\Delta G^{solution}$ values for C3d mutants (from C3d-Ecb complex) at 0 mM ionic strength 134
C.12 $\Delta\Delta G^{solvation}$ values for C3d mutants (from C3d-Ecb complex) at 0 mM ionic strength 134
C.13 Clustering of C3d mutants (from C3d-Ecb complex) at 150 mM ionic strength 135
C.14 $\Delta G^{Coulombic}$ values for C3d mutants (from C3d-Ecb complex) at 150 mM ionic strength 135
C.15 $\Delta G^{solution}$ values for C3d mutants (from C3d-Ecb complex) at 150 mM ionic strength 136
C.16 $\Delta\Delta G^{solvation}$ values for C3d mutants (from C3d-Ecb complex) at 150 mM ionic strength 136

Appendix D: Supporting Data for Chapter 4

D.1 Structures of known ligands in complex with C3d 137
D.2 Electrostatic free energies of C3d-Sbi-IV (Complex 1) 138
D.3 Electrostatic free energies of C3d-Sbi-IV (Complex 2) 139
D.4 Number of nonpolar atomic contacts during MD trajectories 140
D.5 Number of hydrogen bonds during MD trajectories 140
D.6 Number of salt bridges during MD trajectories 140
D.7 Number of nonpolar atomic contacts during SMD trajectories 141
D.8 Number of hydrogen bonds during SMD trajectories 141
D.9 Number of salt bridges during SMD trajectories 141
D.10 Sbi binding to C3 and C3b 142

Appendix E: Supporting Data for Chapter 5

E.1 Sturctures of molecules from Virtual Screen A selected for experimental testing 144
E.2 Competitive binding ELISAs for small molecules 145
E.3 Molecular graphics representation of C3d bound to CR2, Efb-C, Ecb, and Sbi-IV 146
E.4 Sturctures of molecules from Virtual Screen B selected for experimental testing 147

Appendix F: Experimental Testing of a New Generation of Compstatin-Family Peptides

F.1 ELISA and hemolytic assay data for compstatin peptides 155

LIST OF TABLES

Chapter 2: An Evaluation of Poisson-Boltzmann Electrostatic Free Energy Calculations Through Comparison with Experimental Mutagenesis Data

2.1 Properties of selected protein complexes from the Alanine Scanning Energetics Database 24
2.2 Correlation between calculated and experimental data for protein complexes 27
2.3 Intermolecular interactions within selected protein complexes 31

Chapter 3: Complement Inhibition By *Staphylococcus aureus*: Electrostatics of C3d-Efb-C and C3d-Ecb Association

3.1 Intermolecular Coulombic interactions in C3d-Efb-C complex 44
3.2 Intermolecular Coulombic interactions in C3d-Ecb complex 44

Chapter 5: Identification of Low Molecular Mass Inhibitors of C3d-Ligand Interactions: A Step Toward New Complement-Targeted Therapeutics

5.1 C3d-derived peptide sequences 78
5.2 Efb- and Ecb-derived peptide sequences 78
5.3 Molecules selected from Virtual Screen A for experimental testing 85
5.4 Properties of dynamic pharmacophore features for C3d-CR2 90
5.5 Molecules selected from Virtual Screen B for experimental testing 93

Appendix B: Supporting Data for Chapter 2

B.1 Coarse/fine grid dimension and resolution of selected protein complexes 112
B.2 Correlation between Calculated and Experimental Free Energy (unabridged) 113
B.3 Statistical ANOVA p-values for calculated and experimental free energy data 114
B.4 Dependence of free energy correlation on grid resolution (Barnase-Barstar) 114
B.5 Dependence of free energy correlation on grid resolution (Im9-E9 DNase) 115
B.6 Dependence of free energy correlation on grid resolution (RNase Inhibitor-Angiogenin) 115

Appendix C: Supporting Data for Chapter 3

C.1 Grid dimensions and lengths for electrostatics calculations 126
C.2 Intramolecular interactions within C3d (from C3d-Efb-C complex) 126
C.3 Intramolecular interactions within Efb-C (from C3d-Efb-C complex) 127
C.4 Intramolecular interactions within C3d (from C3d-Ecb complex) 127
C.5 Intramolecular interactions within Ecb (from C3d-Ecb complex) 128

Appendix F: Experimental Testing of a New Generation of Compstatin-Family Peptides

F.1 Compstatin peptide sequences 153

CHAPTER 1: INTRODUCTION

1.1 The complement system

The complement system is a branch of innate immunity that provides a first line of defense against pathogens in the bloodstream. During infection, complement proteins recognize pathogen-associated molecular patterns, which initiate complement activation via three independent pathways (classical, lectin, and alternative). Enzymatic cleavage steps generate convertase enzymes, which cleave C3 into C3a and C3b. C3a binds to the C3a receptor (C3aR) on host cells, inducing chemotaxis and inflammatory responses. C3b opsonizes pathogenic cell surfaces and recruit phagocytes to the site of infection. Alternatively, C3b form additional alternative pathway C3 convertases, or can bind to existing C3 convertases, forming distinct convertases capable of cleaving C5. Cleavage of C5 yields C5a and C5b, which are structural homologs of C3a and C3b, respectively. While C5a plays a similar role to C3a, C5b is functionally distinct; it forms a complex with C6 (and subsequently C7, C8, and polymeric C9) to generate the membrane attack complex (C5b-9), forming a pore in the cell membrane and leading to lysis and cell death (1-4) (see Figure 1.1 for an overview of complement activation).

Complement also acts as a link between innate and adaptive immunity. Adaptive immunity is "complemented" when previously encountered antigens are recognized by IgG and IgM, leading to activation of the classical pathway and initiation of innate immune response. Conversely, complement can elicit an adaptive immune response. C3b on pathogen surfaces can be cleaved to iC3b (and subsequently to C3d), both of which can bind complement receptor 2 (CR2) on B cells (Figure 1.1). This interaction leads to augmentation of antibody production, reducing the threshold for B cell activation by several orders of magnitude (5-7).

Specific host-pathogen interactions are necessary to activate the classical pathway (CP) and lectin pathway (LP) of complement. In contrast, the alternative pathway (AP) is continually activated at low levels in serum, and thus is of great interest in understanding broad and rapid initiation of innate immune responses to pathogen exposure. C3 is hydrolyzed at a slow rate to form C3(H_2O), which adopts a C3b-like conformation. Although C3(H_2O) is not cleaved (i.e. C3a is still present), its altered conformation

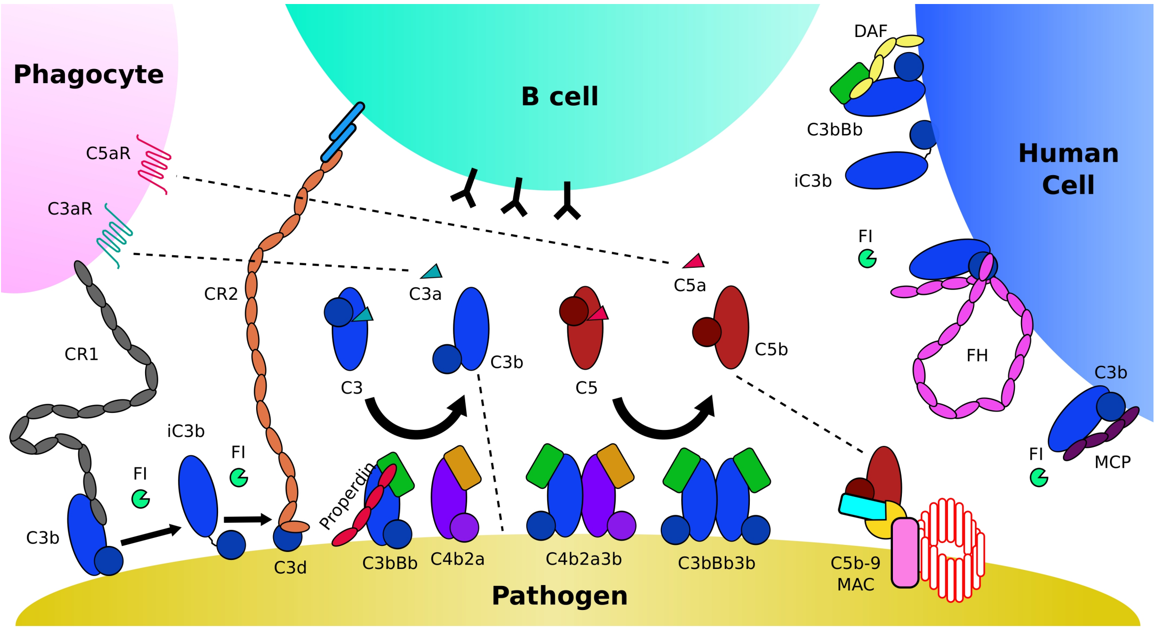


Figure 1.1 Overview of complement activation and regulation. Activation of the classical, lectin, and alternative pathways lead to formation of C3 convertase enzymes (C4b2a and C3bBb) on the surface of pathogenic cells. These convertases cleave C3 into C3a and C3b. C3a binds to C3aR on neutrophils, and C3b covalently attaches to the pathogen surface. Attached C3b can form additional C3bBb convertases (which can bind properdin for stabilization), or join existing convertases to form C3bBb3b or C4b2a3b. These molecules can in turn cleave C5. C5a binds to C5aR, initiating chemotaxis and inflammatory responses, while C5b binds to C6, and subsequently C7, C8, and polymeric C9 to form the lytic membrane attack complex (MAC). CR1 (on the phagocyte surface) binds pathogen-associated C3b, leading to targeted phagocytosis. CR1 also acts as a cofactor for FI, which cleaves C3b into iC3b and then to C3d. C3d binds to CR2 on B cells, which can enhance production of antibodies near the site of infection. On host cells, membrane-bound complement regulators MCP and DAF lead to C3b inactivation and C3bBb dissociation, respectively, protecting the cells from complement attack. Additionally, FH can simultaneously recognizes C3b and polyanion binding sites, thus it is recruited to host cell surfaces and prevents formation of C3bBb.

facilitates binding of factor B (FB), a prerequisite for C3 convertase formation. Factor D (FD) then cleaves FB into two fragments, Ba and Bb. The latter remains bound to C3(H_2O), forming an active convertase (C3(H_2O)Bb) that can bind and cleave C3 into C3a and C3b. Newly-formed C3b molecules can then form additional convertase enzymes (C3bBb) in the vicinity, initiating a positive feedback loop in complement activation (Figure 1.2).

Although rapid activation and amplification of complement AP activity is beneficial during infection, inadequate regulation can lead to damage to host cells and tissues. A number of regulatory mechanisms play a critical role in distinguishing between host and pathogenic cells (Figure 1.1). As mentioned earlier, C3bBb amplifies complement activation by continually cleaving C3, resulting in increased opsonization and downstream complement activation. This amplification is limited by the intrinsic instability of C3bBb, which has a half-life of ~60 seconds. On pathogenic cells, binding of the soluble glycoprotein properdin stabilizes this complex, increasing its half-life by an order of magnitude and further enhancing AP activation. On the other hand, Factor H (FH), another soluble complement regulator, binds to C3b on host cells and diminishes AP activation. FH contains polyanion-binding sites, which facilitate association with host cells, while the N- and C-terminal domains can bind to C3b and C3d. Interaction with C3b competitively inhibits C3b-FB binding, and in turn, prevents formation of C3bBb. Additionally, FH acts as a cofactor for factor I (FI), a serine protease that can cleave and inactivate C3b (to iC3b). Other complement receptors and regulators provide a direct role in host cell protection. Complement receptor 1 (CR1) is expressed on the surface of most blood cells, and is primarily responsible for clearance of complement-opsonized antigens, via interaction with C3b. CR1 can also act as a cofactor for FI, facilitating inactivation of C3b. Membrane cofactor protein (MCP) also facilitates FI-mediated cleavage of C3b. Decay-accelerating factor (DAF) decreases the half-life of the already unstable C3bBb complex, reducing levels of AP activation (8, 9).

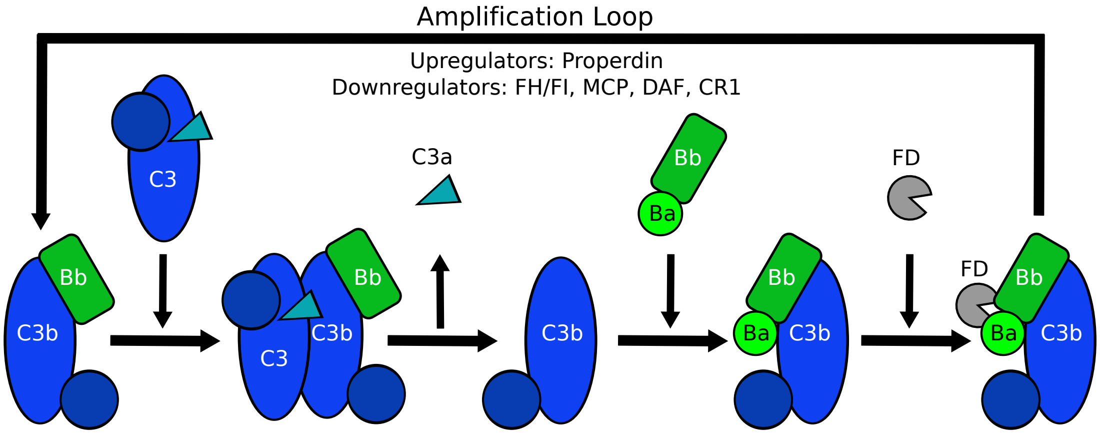


Figure 1.2 Amplification loop of the alternative pathway. Once C3b is formed (either via initial C3(H_2O)-mediated alternative pathway activation or via the classical/lectin pathways), C3b can bind full-length factor B. Factor D can then bind and cleave C3bB, yielding an additional convertase molecule (C3bBb). Newly-formed convertases can bind C3 molecules, and cleave them to C3b, and so on. Binding of properdin to C3bBb stabilizes the complex and upregulates amplification, while complement regulators FH, MCP, DAF, and CR1 all reduce amplification.

1.2 Complement and autoimmunity

Under normal circumstances, complement plays key roles in detection and elimination of pathogens, clearance of immune complexes and apoptotic cells, and development of adaptive immune response and B cell tolerance to self antigens. In many autoimmune diseases, complement fails to perform one of more of these duties. Complement is implicated in many autoimmune diseases, including systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), and age-related macular degeneration (AMD), among others. These diseases typically involve one or more of the following; excessive complement activation, defective regulation, or complement deficiency (10). In AMD, polymorphisms resulting in single amino acid substitutions in C3, FB, and/or FH impair regulation of the AP, leading to complement activation on the retina and gradual loss of central vision (11, 12). The role of complement in SLE and RA is more complex. Onset is characterized by deficiencies of specific complement factors, which can impair proper waste disposal and induce production of autoantibodies. Subsequently, over-activation of complement facilitates disease progression and can lead to inflammation and tissue damage (13, 14). Thus, autoimmunity can intriguingly involve both impaired and enhanced complement activity.

1.3 Pathogenic immune evasion

Pathogens employ a variety of strategies to evade detection by the immune system. To promote survival in the bloodstream, many pathogens express surface-associated and secreted virulence factors that mediate complement activity. A prime example is *Staphylococcus aureus* (*S. aureus*). *S. aureus* expresses several virulence factors that interact with complement proteins and complexes, many of which have been well-studied and characterized in recent years (15-20). Although *S. aureus* is not susceptible to direct killing by complement via the membrane attack complex, complement activation promotes rapid phagocytosis of the bacterium. Owing to its major role in pathogen detection and clearance, AP is a primary target for staphylococcal immune evasion. Chemotaxis inhibitory protein of *S. aureus* (CHIPS) and staphylococcal superantigen-like protein 7 (SSL7) block C5a-mediated chemoattraction via direct binding to C5aR and C5, respectively (21-23). Staphylococcal complement inhibitor (SCIN) inhibits the cleavage activity of C3bBb, in turn reduces bacterial opsonization and C3a/C5a-mediated chemoattraction of neutrophils (24, 25). Finally, three additional C3bBb inhibitors, extracellular fibrinogen-binding protein (Efb) (26), extracellular complement-binding protein (Ecb) (27), and staphylococcal binder of immunoglobulins (Sbi) (28), have diverse structural and functional properties, but share a C3d-binding domain and similar complement regulatory activity (27, 29-33). Each exhibits multiple modes of complement inhibition, which are still not fully understood.

1.4 Complement-targeted and anti-infective therapeutics

The role of complement in autoimmune diseases makes the system an attractive target for therapeutics. Over the past 20 years, many complement inhibitors and regulators have been developed, but only two complement-targeted therapeutic are currently on the market (34). Complement is not highly amenable to traditional drug development. The only traditional drug targets in complement are chemotactic receptors and serine proteases, the latter of which are notoriously difficult to target due to conservation of serine protease structure and lack of specificity (35). Most activation and regulation steps in complement involve multimeric complexes with large protein-protein interfaces, which are challenging to inhibit with small drug-like molecules (36). Consequently, nearly all therapeutics in preclinical/clinical development are biopharmaceuticals. Aside from target specificity, another important consideration is level of complement inhibition. While in certain pathological situations it is beneficial to reduce complement activation to prevent damage to host cells and tissues, blocking complement activity entirely increases susceptibility to infection and may promote development of new autoimmune conditions due to accumulation of immune complexes and apoptotic/necrotic cells (35). Therefore, AP is an attractive target, since complement can still be activated via the CP and LP (37, 38). Currently, several complement therapeutics are in the pipeline. These therapeutics include soluble forms of complement regulators, antibodies against specific complement proteins, and peptides/peptidomimetics that bind receptors or block interactions (38). Latest efforts are aimed specifically at blocking C3 convertase activity via fused complement regulators that target multiple sites simultaneously (i.e. TT30, mini-FH) (39). An alternative approach involves learning from microbes. Some success has come from efforts to inhibit complement using the staphylococcal protein CHIPS and vaccinia viral protein VCP, however immunogenicity and proteolytic degradation pose critical barriers to development of native microbial proteins as therapeutics (39). Using knowledge from microbial virulence factors, including complement inhibition mechanisms and detailed analysis of complement proteins, it may be possible to derive novel complement therapeutics (40).

Interactions between microbial and complement proteins also give rise to the design of new anti-infective strategies. *S. aureus* is a prime target, as increasing resistance to traditional antibiotics and incidence of nosocomial infections pose imminent public health risks (41). While S. aureus typically causes mild skin infections, invasive strains can lead to bacteremia, osteomyelitis, infective endocarditis, and death (15, 41, 42). The wealth of available knowledge regarding interactions between staphylococcal virulence factors and complement provides a solid foundation for the design of highly specific bacterial inhibitors.

1.5 Structural biology of the complement system

Activation and regulation of complement relies on the interaction between complement proteins, receptors and regulators. Ten years ago, only 56 complement structures were deposited in the protein data bank, many of which were not unique and consisted of protein fragments, mutants, or proteins crystallized under differing conditions. Furthermore, only one structure of a complement protein complex was structurally resolved, the original structure of the C3d-CR2 complex (43). The need for better structural understanding of complement was summarized soon thereafter, motivating efforts to elucidate complement mechanisms and interactions using structural biology (3). Beginning in 2005, Piet Gros and coworkers determined crystallographic structures of C3 and C3b, the central proteins of complement activation. The technical and scientific expertise derived from this work formed the basis for the determination of many large proteins and protein complexes in the complement system (44). We now understand how the AP convertase is formed, step-by-step, to form the active C3bBb molecule (25, 45-47). We understand how AP activation is regulated by FH and FI (48, 49). Models have been proposed for binding of C3 to C3bBb and its cleavage mechanism (25). To some extent, C5 convertase formation is now understood (50, 51), and our knowledge of membrane attack complex (C5b-9) formation is greatly enhanced (52). Finally, controversy over the structure of the C3d-CR2 complex, the link between innate and adaptive immunity, has been resolved (53, 54). Structural biology has also made progress in elucidating mechanisms of complement evasion by *S. aureus*. Several ligands have been cocrystallized with complement proteins and complexes (25, 27, 29, 33, 55), even helping to gain insight into the elusive mechanism of the inherently unstable C3bBb complex (25). Indeed, structural biology advances have revealed molecular details and mechanisms that have improved our understanding of complement activation, regulation, and evasion.

One of the most noteworthy observations from nearly all complement interactions is the ubiquity of electrostatic interactions. Electrostatic interactions seem invariably important in driving the association of complement proteins, in a high ionic strength environment (56-66). There is little doubt that electrostatics accelerates complement immune responses during infection. Likewise, it is evident that electrostatics plays an equally important role in driving association of bacterial virulence factors with complement proteins (67-72). Not only is electrostatics important for complex formation and stability, but it also often plays a pivotal role in driving conformational changes and contributing to allosteric activation and inhibition mechanisms (31, 73). This property of the complement system makes it an attractive system to study using computational methods.

1.6 Electrostatic analysis of protein-protein interactions

In recent years, computational methods have rapidly evolved and become an integral part of the analysis of protein-protein interactions. Depending on the accuracy of the underlying physics, algorithm efficiency, and parameterization, computational techniques can provide an excellent complement to experimental techniques and methods. Current methods can accurately reproduce biophysical binding and energetic data, and thus are often employed in a predictive manner. Furthermore, some computational methods can lend insight into phenomena unseen in experiment.

A wealth of new structural information makes complement more amenable than ever to computational biophysics (74). Since electrostatics is implicated in complement interactions, analysis of electrostatics can shed light on the association and stability of complement interactions, and how protein-protein interactions may give rise to important conformational changes (66, 75, 76). Poisson-Boltzmann (PB) calculations represent the forefront of electrostatic analysis using implicit solvation in proteins (77-79). Since solvent molecules and ions are not explicitly modeled, PB calculations can efficiently determine electrostatic potential distributions and electrostatic free energies of proteins. In combination with computational alanine-scanning mutagenesis, PB free energy calculations can predict relative contributions of specific charged amino acids to protein complex stability (a more detailed theoretical explanation of PB calculations can be found in Appendix A.1). The role of electrostatics in protein-protein association can be probed further with Brownian dynamics (BD) simulations. BD simulations have been used by several groups to analyze the diffusional and electrostatically-driven association of proteins (80-86), and to quantify their relative association (k_{on}) rates (see Appendix A.2). Molecular dynamics (MD) provide the most detailed view of protein-protein interactions, and are widely used to computationally model the dynamics of protein structure at atomic resolution (87, 88). MD simulations can provide insight into conformational changes that may be difficult to elucidate using experimental methods. Furthermore, simulations can provide detailed analysis on the stability of protein-protein interactions, including side-chain fluctuations, persistence of nonbonded interactions, and changes in buried interfacial surface area (see Appendix A.3). Integration of the aforementioned methods, in conjunction with available biophysical and immunological data, provides a foundation for design of peptidic and small molecule inhibitors targeting complement interactions.

1.7 Structure-based drug design

Structural biology enables an atomistic view of proteins and provides insight into their function. Rapid progression of structural biology has paved the way for sophisticated computational drug design. Development and improvement of algorithms and frameworks has led to numerous successful drug design endeavors (89-92), which may eventually revolutionize pharmaceutical development. Computation is used in a wide variety of applications, including design of new protein variants, peptide/peptidomimetic design, and virtual high-throughput screening (vHTS) for small drug-like molecules. While protein design can lead to highly-specific therapeutics, they are typically expensive, difficult to deliver, and prone to proteolytic degradation. Small drug-like molecules have ideal pharmacological properties, but are typically suited for targeting the cavities of GPCRs and enzyme active sites, where buried environments provide maximal interaction between the protein and small molecule. In addition, they often lack specificity, resulting in undesired off-target effects. Peptides represent a middle-ground, achieving a balance between specificity, cost, and stability.

Knowledge of the atomic-level structure of a particular protein-protein interaction of interest provides a foundation for peptide design and vHTS. In the design of peptides, the binding partner of the drug target can be used as an initial template. Secondary structural elements (i.e. helices) can be isolated and simulated using BD and MD simulations as described above. BD simulations can provide insight into association of peptides compared to wild-type proteins, while MD can be used to assess secondary structure, structural stability, and interactions with its binding partner. Iterative rounds of sequence optimization and simulation can yield peptides with optimized structure and target affinity. vHTS can be used to rapidly screen large databases of small drug-like molecules for predicted binding affinity for a specific region on a target of interest. Two widely-used methods in vHTS include pharmacophore screening and molecular docking. Pharmacophore screening begins with a model, comprised of spheres, which mimics the spatial and geometrical arrangement of physicochemical properties of an active ligand, with respect to its receptor. Molecules are aligned with the pharmacophore model, and retained as hits if its chemical moieties (ionizable groups, hydrogen bond donors and acceptors, hydrophobic/aromatic groups) match a specified number of features in the model. In molecular docking, molecules explore varied conformations within a predefined region of the target protein, and empirical and force-field based scoring functions are used to rank molecules according to predicted binding affinity.

While protein therapeutics are moderately successful in mediating complement activity, there has been limited success with small molecule therapeutics (93). Peptides and peptidomimetics, including compstatin (94) and PMX-53 (95), hold promise as potent complement-targeted therapeutics. Both molecules have evolved via integration of experimental and computational methods, and continue to be improved (3, 93, 96-98). Following their success, design of peptides targeting other complement interactions has been proposed, including initial reports of success in blocking AP convertase formation (40). Furthermore, few efforts exist to screen for small-drug like complement inhibitors. Recent structural advances in complement afford the opportunity to pursue structure-based drug design.

1.8 Overview

The work described in this thesis aims to investigate the interactions of host and staphylococcal ligands with complement C3d to gain insight into molecular mechanisms of alternative pathway complement inhibition by *Staphylococcus aureus*. Furthermore, this information is used as a basis for the design of peptidic and small drug-like molecules with potential applications as complement-targeted or anti-infective therapeutics. In Chapter 2, we describe the implementation of AESOP, a computational framework developed by our group for analysis of electrostatic similarities of proteins, toward predicting electrostatic free energies of association of protein complexes. We compare calculated Poisson-Boltzmann electrostatic free energies of association with experimental binding data for a diverse selection protein complexes and mutants, and provide insight into the parameter selection, applicability, and limitations of such types of calculations. In Chapter 3, we examine the structural basis for alternative pathway complement inhibition by *Staphylococcus aureus* virulence factors Efb-C and Ecb. We utilize free energy calculations and electrostatic clustering to elucidate the role of electrostatics in these interactions, and to identify residues/regions required for interaction between Efb-C and Ecb with C3d. In Chapter 4, we investigate the dual binding paradigm of anther staphylococcal virulence factor Sbi with C3d. We describe the integration of several dynamic tools to provide insight into the physiological binding mode of the C3d-Sbi complex. Finally, in Chapter 5, we describe the development and application of peptide design and virtual high-throughput screening protocols toward identifying molecules aimed disrupting C3d-ligand interactions.

1.9 References

1. Walport, M.J. 2001. Complement. First of two parts. N. Engl. J. Med. 344: 1058–1066.

2. Walport, M.J. 2001. Complement. Second of two parts. N. Engl. J. Med. 344: 1140–1144.

3. Morikis, D., and J.D. Lambris. 2005. Structural Biology of the Complement System. Boca Raton, FL, USA: CRC Press.

4. Ricklin, D., G. Hajishengallis, K. Yang, and J.D. Lambris. 2010. Complement: a key system for immune surveillance and homeostasis. Nat. Immunol. 11: 785–797.

5. Fearon, D. 1998. The complement system and adaptive immunity. Semin. Immunol. 10: 355–361.

6. Carroll, M.C. 2004. The complement system in regulation of adaptive immunity. Nat. Immunol.. 5: 981–986.

7. Carroll, M.C., and D.E. Isenman. 2012. Regulation of humoral immunity by complement. Immunity. 37: 199–207.

8. Zipfel, P.F., and C. Skerka. 2009. Complement regulators and inhibitory proteins. Nat. Rev. Immunol. 9: 729–740.

9. Meri, S., and H. Jarva. 2011. Complement Regulation. Vox Sanguinis. 74: 291–302.

10. Wagner, E., and M.M. Frank. 2009. Therapeutic potential of complement modulation. Nat. Rev. Drug. Discov. 9: 43–56.

11. Holers, V.M. 2008. The spectrum of complement alternative pathway-mediated diseases. Immunol. Rev. 223: 300–316.

12. Sarma, J.V., and P.A. Ward. 2010. The complement system. Cell Tissue Res. 343: 227–235.

13. Walport, M.J. 2002. Complement and systemic lupus erythematosus. Arthritis Res. 4: S279-S293.

14. Ballanti, E., C. Perricone, E. Greco, M. Ballanti, G. Di Muzio, et al. 2013. Complement and autoimmunity. Immunol. Res. 56: 477–491.

15. Foster, T.J. 2005. Immune evasion by staphylococci. Nat. Rev. Micro. 3: 948–958.

16. Rooijakkers, S.H.M., K.P.M. van Kessel, and J.A.G. van Strijp. 2005. Staphylococcal innate immune evasion. Trends Microbiol. 13: 596–601.

17. Chavakis, T., K.T. Preissner, and M. Herrmann. 2007. The anti-inflammatory activities of Staphylococcus aureus. Trends Immunol. 28: 408–418.

18. Geisbrecht, B. 2009. Staphylococcal complement inhibitors: biological functions, recognition of complement components, and potential therapeutic implications. Current Topics in Complement II. New York, NY: Springer US. pp. 221–236.

19. Serruto, D., R. Rappuoli, M. Scarselli, P. Gros, and J.A.G. van Strijp. 2010. Molecular mechanisms of complement evasion: learning from staphylococci and meningococci. Nat. Rev. Micro. 8: 393–399.

20. Laarman, A., F. Milder, J. Strijp, and S. Rooijakkers. 2010. Complement inhibition by gram-positive pathogens: molecular mechanisms and therapeutic implications. J. Mol. Med. 88: 115–120.

21. de Haas, C.J.C., K.E. Veldkamp, A. Peschel, F. Weerkamp, W.J.B. van Wamel, et al. 2004. Chemotaxis inhibitory protein of Staphylococcus aureus, a bacterial antiinflammatory agent. J. Exp. Med. 199: 687–695.

22. Langley, R., B. Wines, N. Willoughby, I. Basu, T. Proft, et al. 2005. The Staphylococcal superantigen-like protein 7 binds IgA and complement C5 and inhibits IgA-FcαRI binding and serum killing of bacteria. J. Immunol. 174: 2926-2933.

23. Laursen, N.S., N. Gordon, S. Hermans, N. Lorenz, N. Jackson, et al. 2010. Structural basis for inhibition of complement C5 by the SSL7 protein from Staphylococcus aureus. Proc. Natl. Acad. Sci. U.S.A. 107: 3681–3686.

24. Rooijakkers, S.H.M., M. Ruyken, A. Roos, M.R. Daha, J.S. Presanis, et al. 2005. Immune evasion by a staphylococcal complement inhibitor that acts on C3 convertases. Nat. Immunol. 6: 920–927.

25. Rooijakkers, S.H.M., J. Wu, M. Ruyken, R. van Domselaar, K.L. Planken, et al. 2009. Structural and functional implications of the alternative complement pathway C3 convertase stabilized by a staphylococcal inhibitor. Nat. Immunol. 10: 722–728.

26. Lee, L.Y.L., M. Höök, D. Haviland, R.A. Wetsel, E.O. Yonter, et al. 2004. Inhibition of complement activation by a secreted Staphylococcus aureus protein. J. Infect. Dis. 190: 571–579.

27. Hammel, M., G. Sfyroera, S. Pyrpassopoulos, D. Ricklin, K.X. Ramyar, et al. 2007. Characterization of Ehp, a secreted complement inhibitory protein from Staphylococcus aureus. J. Biol. Chem. 282: 30051–30061.

28. Zhang, L., K. Jacobsson, J. Vasi, M. Lindberg, and L. Frykberg. 1998. A second IgG-binding protein in Staphylococcus aureus. Microbiology. 144: 985–991.

29. Hammel, M., G. Sfyroera, D. Ricklin, P. Magotti, J.D. Lambris, et al. 2007. A structural basis for complement inhibition by Staphylococcus aureus. Nat. Immunol. 8: 430–437.

30. Jongerius, I., B.L. Garcia, B.V. Geisbrecht, J.A.G. van Strijp, and S.H.M. Rooijakkers. 2010. Convertase inhibitory properties of Staphylococcal extracellular complement-binding protein. J. Biol. Chem. 285: 14973–14979.

31. Ricklin, D., H. Chen, M. Hammel, B.L. Garcia, W.J. McWhorter, et al. 2010. Allosteric inhibition of complement function by the Staphylococcal immune evasion protein Efb. Mol. Immunol. 47: 2244–2245.

32. Burman, J.D., E. Leung, K.L. Atkins, M.N. O'Seaghdha, L. Lango, et al. 2008. Interaction of human complement with Sbi, a staphylococcal immunoglobulin-binding protein - Indications of a novel mechanism of complement evasion by Staphylococcus aureus. J. Biol. Chem. 283: 17579–17593.

33. Clark, E.A., S. Crennell, A. Upadhyay, A.V. Zozulya, J.D. Mackay, et al. 2010. A structural basis for Staphylococcal complement subversion: X-ray structure of the complement-binding domain of Staphylococcus aureus protein Sbi in complex with ligand C3d. Mol. Immunol. 48: 452–462.

34. Qu, H., D. Ricklin, H. Bai, H. Chen, E.S. Reis, et al. 2013. New analogs of the clinical complement inhibitor compstatin with subnanomolar affinity and enhanced pharmacokinetic properties. Immunobiology. 218: 496–505.

35. Ricklin, D., and J.D. Lambris. 2007. Complement-targeted therapeutics. Nat. Biotechnol. 25: 1265–1275.

36. Wells, J.A., and C.L. McClendon. 2007. Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. Nature. 450: 1001–1009.

37. Holers, V. 2004. The alternative pathway of complement in disease: opportunities for therapeutic targeting. Mol. Immunol. 41: 147–152.

38. Mollnes, T.E., and M. Kirschfink. 2006. Strategies of therapeutic complement inhibition. Mol. Immunol. 43: 107–121.

39. Ricklin, D. 2012. Manipulating the mediator: modulation of the alternative complement pathway C3 convertase in health, disease and therapy. Immunobiology. 217: 1057–1066.

40. Gao, Y.-P., J. Dong, X. Zhang, Y. Liu, Q. Lu, et al. 2013. A novel Peptide can mimic extracellular fibrinogen-binding protein to block the activation of complement system. Cell Biochem. Biophys. 66: 753–757.

41. Kobayashi, S.D., and F.R. DeLeo. 2009. An update on community-associated MRSA virulence. Curr. Opin. Pharmacol. 9: 545–551.

42. Sibbald, M.J.J.B., A.K. Ziebandt, S. Engelmann, M. Hecker, A. de Jong, et al. 2006. Mapping the Pathways to Staphylococcal Pathogenesis by Comparative Secretomics. Microbiol. Mol. Biol. Rev. 70: 755–788.

43. Szakonyi, G., J.M. Guthridge, D. Li, K. Young, V.M. Holers, and X.S. Chen. 2001. Structure of Complement Receptor 2 in Complex with Its C3d Ligand. Science. 292: 1725–1728.

44. Lea, S.M., and S. Johnson. 2012. Putting the structure into complement. Immunobiology. 217: 1117–1121.

45. Janssen, B.J.C., E.G. Huizinga, H.C.A. Raaijmakers, A. Roos, M.R. Daha, et al. 2005. Structures of complement component C3 provide insights into the function and evolution of immunity. Nature. 437: 505–511.

46. Janssen, B.J.C., A. Christodoulidou, A. McCarthy, J.D. Lambris, and P. Gros. 2006. Structure of C3b reveals conformational changes that underlie complement activity. Nature. 444: 213–216.

47. Forneris, F., D. Ricklin, J. Wu, A. Tzekou, R.S. Wallace, et al. 2010. A safety catch on immune response. Science. 330: 1816–1820.

48. Wu, J., Y.-Q. Wu, D. Ricklin, B.J.C. Janssen, J.D. Lambris, et al. 2009. Structure of complement fragment C3b–factor H and implications for host protection by complement regulators. Nat. Immunol. 10: 729–734.

49. Roversi, P., S. Johnson, J.J.E. Caesar, F. McLean, K.J. Leath, et al. 2011. Structural basis for complement factor I control and its disease-associated sequence polymorphisms. Proc. Natl. Acad. Sci. U.S.A. 108: 12839–12844.

50. Fredslund, F., N.S. Laursen, P. Roversi, L. Jenner, C.L.P. Oliveira, et al. 2008. Structure of and influence of a tick complement inhibitor on human complement component 5. Nat. Immunol. 9: 753–760.

51. Laursen, N.S., K.R. Andersen, I. Braren, E. Spillner, L. Sottrup-Jensen, et al. 2011. Substrate recognition by complement convertases revealed in the C5-cobra venom factor complex. Embo J. 30: 606–616.

52. Hadders, M.A., D. Bubeck, P. Roversi, S. Hakobyan, F. Forneris, et al. 2012. Assembly and regulation of the membrane attack complex based on structures of C5b6 and sC5b9. Cell Rep. 1: 200–207.

53. van den Elsen, J.M.H., and D.E. Isenman. 2011. Complement receptor complex structure. Science. 332: 608–611.

54. Morikis, D. F1000.com/10371956.

55. Ippel, J.H., C.J.C. de Haas, A. Bunschoten, J.A.G. van Strijp, J.A.W. Kruijtzer, et al. 2009. Structure of the tyrosine-sulfated C5a receptor N terminus in complex with chemotaxis inhibitory protein of Staphylococcus aureus. J. Biol. Chem. 284: 12363–12372.

56. Blom, A.M., J. Webb, B.O. Villoutreix, and B. Dahlbäck. 1999. A cluster of positively charged amino acids in the C4BP alpha-chain is crucial for C4b binding and factor I cofactor function. J. Biol. Chem. 274: 19237–19245.

57. Morikis, D., and J.D. Lambris. 2004. The electrostatic nature of C3d-complement receptor 2 association. J. Immunol. 172: 7537–7547.

58. Kieslich, C.A., H. Vazquez, G.N. Goodman, A.L. de Victoria, and D. Morikis. 2011. The effect of electrostatics on factor H function and related pathologies. J. Mol. Graph. Model. 29: 1047–1055.

59. Kieslich, C.A., and D. Morikis. 2012. The two sides of complement C3d: Evolution of electrostatics in a link between innate and adaptive immunity. PLoS Comput. Biol. 8: e1002840.

60. Morikis, D., and L. Zhang. 2006. An immunophysical study of the complement system: Examples for the pH dependence of protein binding and stability. J. Non-Cryst. Sol. 352: 4445–4450.

61. Zhang, L., B. Mallik, and D. Morikis. 2007. Immunophysical Exploration of C3d–CR2(CCP1-2) Interaction Using Molecular Dynamics and Electrostatics. J. Mol. Biol. 369: 567–583.

62. Cheung, A.S., C.A. Kieslich, J. Yang, and D. Morikis. 2010. Solvation effects in calculated electrostatic association free energies for the C3d-CR2 complex and comparison with experimental data. Biopolymers. 93: 509–519.

63. Kieslich, C.A., D. Morikis, J. Yang, and D. Gunopulos. 2011. Automated computational framework for the analysis of electrostatic similarities of proteins. Biotechnol. Prog. 27: 316–325.

64. Hakkoymaz, H., C.A. Kieslich, R.D. Gorham Jr, D. Gunopulos, and D. Morikis. 2011. Electrostatic similarity determination using multiresolution analysis. Mol. Inf. 30: 733-746.

65. El-Assaad, A.M., C.A. Kieslich, R.D. Gorham Jr, and D. Morikis. 2011. Electrostatic exploration of the C3d-FH4 interaction using a computational alanine scan. Mol. Immunol. 48: 1844–1850.

66. Kieslich, C.A., P. Tamamis, R.D. Gorham Jr, A. López de Victoria, N.U. Sausman, et al. 2012. Exploring protein-protein and protein-ligand interactions in the immune system using molecular dynamics and continuum electrostatics. Curr. Phys. Chem. 2: 324-343.

67. Zhang, L., and D. Morikis. 2006. Immunophysical properties and prediction of activities for vaccinia virus complement control protein and smallpox inhibitor of complement enzymes using molecular dynamics and electrostatics. Biophys. J. 90: 3106–3119.

68. Haspel, N., D. Ricklin, B.V. Geisbrecht, L.E. Kavraki, and J.D. Lambris. 2008. Electrostatic contributions drive the interaction between Staphylococcus aureus protein Efb-C and its complement target C3d. Protein Sci. 17: 1894–1906.

69. Ricklin, D., S.K. Ricklin-Lichtsteiner, M.M. Markiewski, B.V. Geisbrecht, and J.D. Lambris. 2008. Cutting Edge: Members of the Staphylococcus aureus extracellular fibrinogen-binding protein family inhibit the interaction of C3d with complement receptor 2. J. Immunol. 181: 7463–7467.

70. Pyaram, K., C.A. Kieslich, V.N. Yadav, D. Morikis, and A. Sahu. 2010. Influence of electrostatics on the complement regulatory functions of Kaposica, the complement inhibitor of Kaposi's Sarcoma-associated herpesvirus. J. Immunol. 184: 1956–1967.

71. Gorham Jr, R.D., C.A. Kieslich, and D. Morikis. 2011. Complement inhibition by Staphylococcus aureus: Electrostatics of C3d–EfbC and C3d–Ehp association. Cel. Mol. Bioeng. 5: 32–43.

72. Sfyroera, G., M. Katragadda, D. Morikis, S.N. Isaacs, and J.D. Lambris. 2005. Electrostatic modeling predicts the activities of orthopoxvirus complement control proteins. J. Immunol. 174: 2143–2151.

73. Gros, P., F.J. Milder, and B.J.C. Janssen. 2008. Complement driven by conformational changes. Nat. Rev. Immunol. 8: 48–58.

74. Morikis, D., and J. Lambris. 2004. Physical methods for structure, dynamics and binding in immunological research. Trends Immunol. 25: 700–707.

75. Gorham, R.D., C.A. Kieslich, and D. Morikis. 2010. Electrostatic clustering and free energy calculations provide a foundation for protein design and optimization. Ann. Biomed. Eng. 39: 1252–1263.

76. Wu, J., and D. Morikis. 2006. Molecular thermodynamics for charged biomacromolecules. Fluid Phase Equilibria. 241: 317–333.

77. Baker, N.A. 2004. Poisson–Boltzmann Methods for Biomolecular Electrostatics. In: Methods in Enzymology. Elsevier. pp. 94–118.

78. Honig, B., and A. Nicholls. 1995. Classical electrostatics in biology and chemistry. Science. 268: 1144–1149.

79. Lu, B.Z., Y.C. Zhou, M.J. Holst, and J.A. McCammon. 2008. Recent progress in numerical methods for the Poisson-Boltzmann equation in biophysical applications. Commun. Comput. Phys. 3: 973-1009.

80. Elcock, A., R. Gabdoulline, R. Wade, and J. McCammon. 1999. Computer simulation of protein-protein association kinetics: Acetylcholinesterase-fasciculin. J. Mol. Biol. 291: 149–162.

81. Ermak, D.L., and J.A. McCammon. 1978. Brownian dynamics with hydrodynamic interactions. J. Chem. Phys. 69: 1352.

82. Gabdoulline, R.R., and R.C. Wade. 1997. Simulation of the diffusional association of barnase and barstar. Biophys. J. 72: 1917–1929.

83. Wade, R.C., R.R. Gabdoulline, S.K. Lüdemann, and V. Lounnas. 1998. Electrostatic steering and ionic tethering in enzyme-ligand binding: insights from simulations. Proc. Natl. Acad. Sci. U.S.A. 95: 5942–5949.

84. Spaar, A., C. Dammer, R.R. Gabdoulline, R.C. Wade, and V. Helms. 2006. Diffusional encounter of barnase and barstar. Biophys. J. 90: 1913–1924.

85. Mereghetti, P., R.R. Gabdoulline, and R.C. Wade. 2010. Brownian dynamics simulation of protein solutions: Structural and dynamical properties. Biophys. J. 99: 3782–3791.

86. Grant, B.J., D.M. Gheorghe, W. Zheng, M. Alonso, G. Huber, et al. 2011. Electrostatically biased binding of kinesin to microtubules. PLoS Biol. 9: e1001207.

87. Elcock, A.H., D. Sept, and J.A. McCammon. 2001. Computer simulation of protein−protein interactions. J. Phys. Chem. B. 105: 1504–1518.

88. Adcock, S.A., and J.A. McCammon. 2006. Molecular dynamics: Survey of methods for simulating the activity of proteins. Chem. Rev. 106: 1589–1615.

89. Clark, D.E. 2008. What has virtual screening ever done for drug discovery? Expert Opin. Drug Discov. 3: 841–851.

90. Villoutreix, B.O., R. Eudes, and M.A. Miteva. 2009. Structure-based virtual ligand screening: recent success stories. Comb. Chem. High Throughput Screen. 12: 1000–1016.

91. Amaro, R.E., A. Schnaufer, H. Interthal, W. Hol, K.D. Stuart, et al. 2008. Discovery of drug-like inhibitors of an essential RNA-editing ligase in Trypanosoma brucei. Proc. Natl. Acad. Sci. U.S.A. 105: 17278–17283.

92. Altman, M.D., A. Ali, G.S.K.K. Reddy, M.N.L. Nalam, S.G. Anjum, et al. 2008. HIV-1 protease inhibitors from inverse design in the substrate envelope exhibit subnanomolar binding to drug-resistant variants. J. Am. Chem. Soc. 130: 6099–6113.

93. Qu, H., D. Ricklin, and J.D. Lambris. 2009. Recent developments in low molecular weight complement inhibitors. Mol. Immunol. 47: 185–195.

94. Sahu, A., B. Kay, and J.D. Lambris. 1996. Inhibition of human complement by a C3-binding peptide isolated from a phage-displayed random peptide library. J. Immunol. 157: 884–891.

95. Wong, A.K., A.M. Finch, G.K. Pierens, D.J. Craik, S.M. Taylor, et al. 1998. Small molecular probes for G-protein-coupled C5a receptors: Conformationally constrained antagonists derived from the C terminus of the human plasma protein C5a. J. Med. Chem. 41: 3417–3425.

96. Ricklin, D., and J.D. Lambris. 2008. Compstatin: A complement inhibitor on its way to clinical application. In: Current Topics in Complement II. New York, NY: Springer US. pp. 262–281.

97. Monk, P.N., A.-M. Scola, P. Madala, and D.P. Fairlie. 2007. Function, structure and therapeutic potential of complement C5a receptors. British J. Pharmacol. 152: 429–448.

98. Woodruff, T.M., K.S. Nandakumar, and F. Tedesco. 2011. Inhibiting the C5-C5a receptor axis. Mol. Immunol. 48: 1631–1642.

CHAPTER 2: AN EVALUATION OF POISSON-BOLTZMANN ELECTROSTATIC FREE ENERGY CALCULATIONS THROUGH COMPARISON WITH EXPERIMENTAL MUTAGENESIS DATA

2.1 Introduction

Many biological processes rely on the interactions between specific biomolecules and can be explained on the basis of the dynamic movements and noncovalent interactions within a biomolecular complex. In particular, electrostatic forces are known to play a crucial role in protein function (1-8). In highly and oppositely charged protein systems, electrostatic steering is purported to facilitate the formation of a bound complex (5, 9-15). The association of proteins is thought to be comprised of two main steps; recognition and binding. According to the two-step model proposed by McCammon et al. (16), proteins undergo accelerated diffusion to form a weak nonspecific encounter complex, which is driven primarily by long-range electrostatic interactions. In this initial recognition step, the global electrostatic potential distribution of each protein aids in accelerated diffusion, in which proteins are guided toward each other by complementary charge distribution (3, 16-19). Subsequently, short-range van der Waals and electrostatic interactions are formed to stabilize the final bound complex.

In the analysis of electrostatics, the fundamental governing equation is Coulomb's Law. Since proteins function in solvated environments, which have different dielectric properties than the protein interior, it becomes necessary to represent the effects of solvent molecules on the electrostatic potential of a protein. Molecular dynamics simulations are often used to describe protein-protein interactions in detail. Simulations typically include both covalent and noncovalent interactions, and provide a time-dependent analysis of a complete atomistic model of the system in a physiological environment, including explicit solvent interactions. Consequently, simulations are computationally expensive and simplifying assumptions are often made to save time and resources.

As an alternative to Coulombic calculations and molecular dynamics simulations, Poisson-Boltzmann electrostatics calculations have been widely applied in the analysis of biomolecular interactions and represent the forefront of accurate and efficient calculation of electrostatic potential for biomolecules (3, 8, 15, 20-24). Poisson-Boltzmann free energy calculations account for both long- and short-range electrostatic interactions, but only implicitly account for desolvation and solvent shielding upon complex formation. In addition, water molecules and ions are typically not explicitly modeled in the simulation, but their presence is implicitly accounted for using a variable dielectric environment and continuum solvation. Since water molecules are not explicitly modeled, the effects of solvent polarizability on the electrostatic potential of the protein are included via a position-dependent dielectric coefficient. The effect of ionic screening is accounted for by an ion accessibility term, which in turn incorporates ionic strength. By defining an appropriate boundary for dielectric coefficient and ionic accessibility, the Poisson-Boltzmann equation can be solved numerically for electrostatic potential.

While Poisson-Boltzmann methods provide a fast approximation of protein-protein electrostatic interactions, simplifying assumptions are known to compromise the predictive capabilities of these calculations (3, 7, 21, 25-28). The selection of appropriate dielectric coefficients and dielectric boundaries has been the subject of debate in recent literature, and is perhaps one of the largest sources of variability in continuum electrostatics calculations (5-7, 27, 29-36). It is well known from experimental data that water has a dielectric coefficient of approximately 78, resulting from the highly polar nature of water molecules. The protein interior, especially for large globular proteins, is completely shielded from solvent and thus should have a significantly lower coefficient. In contrast, it is reported that dynamics and structural relaxation of the protein interior significantly increases the effective dielectric coefficient (6, 26, 30, 37-40). Researchers have reported the application of dielectric coefficients between 2 and 40 to represent the protein interior, although values are consistently much lower than the surrounding solvent (20, 22, 37). Other groups have investigated the use of distance-dependent dielectric functions, in which a lower dielectric coefficient is used to model residues buried deeply within the protein interior compared to those proximal to the surface (6, 7, 29, 39, 41). It is of interest to determine an appropriate dielectric coefficient based on correlations between calculated and experimental data.

Numerous recent studies have applied Poisson-Boltzmann free energy calculations in examining the association of protein complexes (12, 22, 25, 28, 32, 34, 41-48). In combination with computational alanine-scanning mutagenesis, it is possible to use continuum electrostatics to predict protein mutants that enhance or inhibit binding. Our lab has recently developed a high-throughput computational protocol to elucidate the role of electrostatics in protein association, stability, and function. This approach, known as

Analysis of Electrostatic Similarities Of Proteins (AESOP), involves a series of steps which ultimately yield a set of free energy values for single-alanine mutants of a protein complex from Poisson-Boltzmann electrostatics calculations (47). The idea behind this method is that mutation of ionizable residues will introduce a perturbation in the electrostatic potential of the protein, and the effect of this mutation on protein complex association (and in turn, protein function) can be quantitatively assessed in an efficient manner. AESOP is useful in engineering protein mutants with tailored physicochemical properties and functions, which can aid in the development of therapeutics for a variety of infections, diseases, and conditions.

Herein, we report a comparison of electrostatic free energy values with free energies based on experimental binding and kinetic data. We show correlations between calculated and experimental data for alanine mutants of five protein complexes, representing a diverse dataset for evaluation of our methods. Additionally, we examine the effects of dielectric and ionic screening on protein association and suggest appropriate values for the dielectric coefficient of the protein interior. We are interested in determining whether electrostatics alone can predict the relative binding of wild-type and mutated protein complexes, as well as elucidate which conditions provide the most accurate calculations.

2.2 Methods

In order to evaluate computational free energy calculations with experimental data, we utilized the Alanine Scanning Energetics Database (ASEdb) (49) to select several systems for rigorous comparison. Criteria for selection included the availability of a co-crystal structure of the protein complex and binding/kinetic data for 10 or more experimental alanine mutants of charged amino acids. The resulting five systems (with corresponding experimental data) included Barnase-Barstar (50, 51), Im9-E9 DNase (52), RNase Inhibitor-Angiogenin (53-55), hGH-hGHbp (56, 57), and gp120-CD4 (58). A remaining system that satisfied our selection criteria, the Factor VII-Tissue Factor protein complex, was not used because it contained carboxyglutamic acid residues, which arise as a result of posttranslational modification. Complications with parametrization of these residues led to the decision to exclude this protein complex from our analysis, despite the abundance of experimental mutagenesis data.

The three-dimensional co-crystal structure for each protein complex was obtained from the Protein Data Bank (PDB) (59). The accession codes used were 1BRS (Barnase-Barstar) (60), 1BXI (Im9-E9 DNase) (61), 1A4Y (RNase Inhibitor-Angiogenin) (62), 3HHR (hGH-hGHbp) (63), and 1GC1 (gp120-CD4) (64). The Barnase-Barstar complex consisted of three nearly identical dimers, however chains A and C (Barnase) and D and E (Barstar) were missing N-terminal residues. Chain F was the only complete Barstar molecule (which included two basic residues missing in the other Barstar molecules), thus the C:F complex was chosen, despite the two missing N-terminal uncharged residues in Chain C. The Im9-E9 DNase structure contained only one complex structure with no missing residues, thus the original PDB file was used. The RNase Inhibitor-Angiogenin structure contained two identical complex structures, thus the first was chosen for our calculations. The gp120-CD4 structure contained the eponymous protein complex, in addition to a monoclonal antibody. The antibody chains were removed from the PDB file for our analysis. The PDB structure of the hGH-hGHbp protein complex contained several gaps in the protein sequence, in addition to missing residues at the C-terminal end of hGHbp. The amino acid sequence from 3HHR was obtained in Chimera (65) and aligned with the native full sequence of the hGH-hGHbp complex (obtained from RCSB) using ClustalW to identify specific locations of missing residues (66). A homology model was built using Modeller (67), with the structure from 3HHR was used as a template for our homology model, in order to model missing loops. The entire structure was subjected to optimization using Modeller, and the alpha-carbon RMSD between the template and model structures was 0.483 Å. For each refined PDB file, incomplete side chains were replaced using the Dock Prep tool and Dunbrack rotamer library in Chimera.

The template protein complex structures (obtained as described above) were subjected to hydrogen bond optimization and removal of van der Waals clashes using WHATIF. WHATIF (68) was also used to generate mutations of charged amino acids (arginine, lysine, glutamic acid, and aspartic acid) in the optimized protein complex to alanine. We did not subject structures to any energy minimization or dynamic relaxation. Mutant protein complexes were subsequently split into their respective component proteins, yielding three sets of mutant PDB files (one for each component plus the complex) for each system. Structural changes following mutation were not considered, as these changes were found not to have a significant effect on electrostatics calculations (69). Atomic radii, partial charges, and hydrogen atoms were added to all PDB files using PDB2PQR (70) and the PARSE force field (71). PQR files were generated without structural optimization or rearrangement, using the --nodebump and --noopt flags in PDB2PQR. Each protein and protein complex was embedded in a three-dimensional discretized grid containing 129 × 129 × 129 points, and the Adaptive Poisson-Boltzmann Solver (APBS) (72) was used to calculate electrostatic potential at each grid point. Grid spacing for each protein complex were determined using VMD (73). The grid resolution for each protein complex ranged between 0.5 and 1.5 Å in each dimension, depending on the size of the protein, magnitude of electrostatic potential and its distribution surrounding the protein, and calculation conditions (Table B1). Free energy for each protein and protein complex was calculated according to

$$G_{elec} = \frac{1}{2}\sum_{i} q_i \varphi_i \tag{1}$$

In Eq. 1, q_i and φ_i represent charge and electrostatic potential at each grid point, respectively. Charges, dielectric coefficients, and ion accessibilities were assigned at each point, in order to solve the Poisson-Boltzmann equation for electrostatic potential. Dielectric and ion accessibility were assigned based on the solvent-accessible and ion-accessible protein surfaces, determined using a probe of radii equal to 1.4 and 2.0 Å, respectively. Free energies of complex association and solvation were calculated according to the thermodynamic cycle in Figure 2.1. Free energy was calculated for a given protein complex and its components in both a reference and solvated state, and reported free energy values were calculated according to:

$$\Delta G^{ref} = G_{AB}^{ref} - G_{A}^{ref} - G_{B}^{ref} \quad (2),$$

$$\Delta G^{solu} = G_{AB}^{solu} - G_{A}^{solu} - G_{B}^{solu} \quad (3),$$

$$\Delta\Delta G^{solvation} = \Delta G_{AB}^{solvation} - \Delta G_{A}^{solvation} - \Delta G_{B}^{solvation} = \Delta G^{solu} - \Delta G^{ref} \quad (4).$$

Since charges are distributed on a discretized grid in APBS, erroneous self energies can arise due to the distribution of a single charge across multiple grid points, which lead to propagation of error in electrostatic

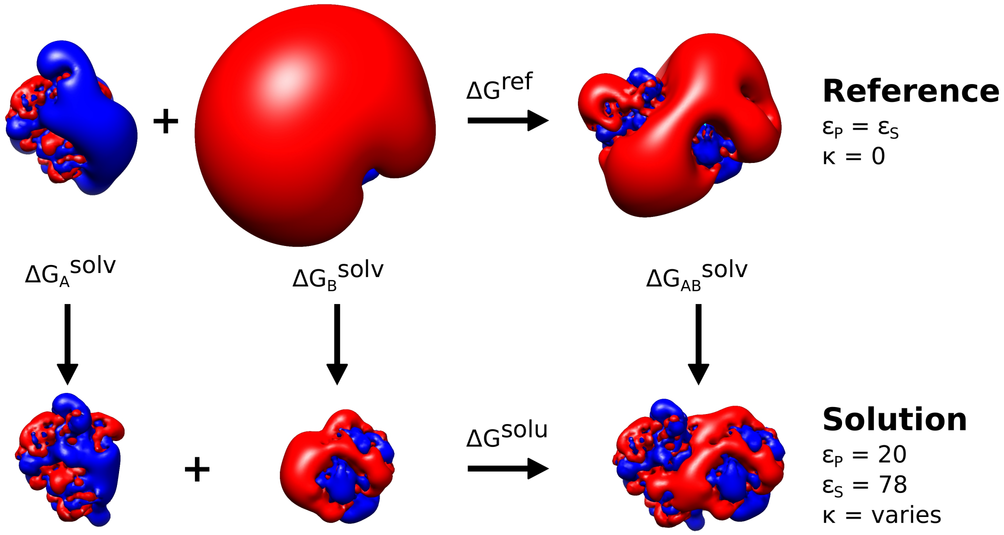


Figure 2.1 Thermodynamic cycle for association and solvation of protein complexes. The top horizontal process represents protein association in a reference environment, with no water molecules or ions present. Dielectric constants within and surrounding the protein are equivalent. The bottom horizontal process represents protein association in ionic solution, with dielectric coefficients varying between the protein interior and the solvent. The ion accessibility term (κ, inverse Debye length) is no longer zero and depends on ion concentration. Vertical processes represent the free energy change associated with solvation of each protein or complex relative to the reference state. The images show isopotential contours for Barnase (*A*), Barstar (*B*), and Barnase-Barstar (*AB*) calculated using protein dielectric coefficient of 2, ionic strengths of 0 mM (top) and 150 mM (bottom), and isovalues of ±1 k_BT/e.

calculations (21, 25, 74). In PB free energy calculations, a reference state must be implemented and each atom must be discretized to the same grid points in all individual calculations in order to eliminate these grid artifacts. Also, in order to alleviate this issue, protein structures must remain static after initial optimization, such that they retain the same conformation in both free and bound states. $\Delta\Delta G^{solvation}$ values calculated in this manner ensure cancellation of all grid artifacts. In order to determine accurate values for ΔG^{ref} and ΔG^{solu}, we used the APBS supplementary module Coulomb to calculate Coulombic potential of each protein and protein complex, with $\Delta G^{Coulombic}$ calculated according to:

$$\Delta G^{Coulombic} = G_{AB}^{Coulombic} - G_{A}^{Coulombic} - G_{B}^{Coulombic} \qquad (5).$$

Using this value in place of ΔG^{ref}, we were able to calculate $\Delta G^{solution}$ according to:

$$\Delta G^{solution} = \Delta\Delta G^{solvation} + \Delta G^{Coulombic} \qquad (6).$$

From Eqs. 5 and 6, we can obtain accurate representations of the free energy of association in a reference and solvated state, without the influence of grid artifacts.

Experimental binding and kinetic data were used to calculate binding free energies for protein complexes and their respective mutants. Data was collected under various conditions, including buffer concentration (10-100 mM), pH (6-8), and ionic strength of solution (0-200 mM). All measurements were performed at 25°C. Experimental binding energies were calculated according to:

$$\Delta G^{exp} = -RT\,ln\left(K_D\right) \qquad (7).$$

Calculated and experimental free energy values for mutant proteins were normalized based on the values for the wild-type protein complex ($\Delta G^{mut}/\Delta G^{wt}$) in order to compare multiple datasets of kinetic data for a given protein complex. Pearson correlation coefficients and ANOVA tests were calculated for $\Delta\Delta G^{solvation}$, $\Delta G^{Coulombic}$, and $\Delta G^{solution}$ with ΔG^{exp} using the R software package.

Visualization of protein structures and electrostatic potential distributions was performed using UCSF Chimera, which in turn was used to generate molecular graphics presented here. Calculation of per-residue solvent-accessible surface area (SASA) values for protein complexes was also performed in UCSF Chimera, using a probe of 1.4 Å and connecting the center points of the probe as it is rolled along the protein surface (without hydrogen atoms).

Table 2.1. Properties of selected protein complexes from the Alanine Scanning Energetics Database.

Protein	Charge	Length[a]	% Buried[b]	% Charged[c]	Buried SASA
Barnase	+2e	108	14	24	1558
Barstar	-6e	89	17	27	
Im9	-9e	85	13	25	1565
E9 Dnase	+5e	131	13	36	
RNase Inhibitior	-22e	460	6	22	2681
Angiogenin	+10e	123	18	24	
hGH	-5e	190	15	24	2673
hGHbp	-6e	205	14	25	
gp120	+4e	297	6	18	1963
CD4	+5e	181	11	24	

[a] Total number of residues in each protein.

[b] Percentage of total SASA of free protein that is buried upon complex formation ($SASA^{free} - \Sigma[SASA^{free,res[i]} - SASA^{complex,res[i]}]$).

[c] Percentage of residues that are charged at physiological pH in each protein (i.e. Asp, Glu, Arg, Lys).

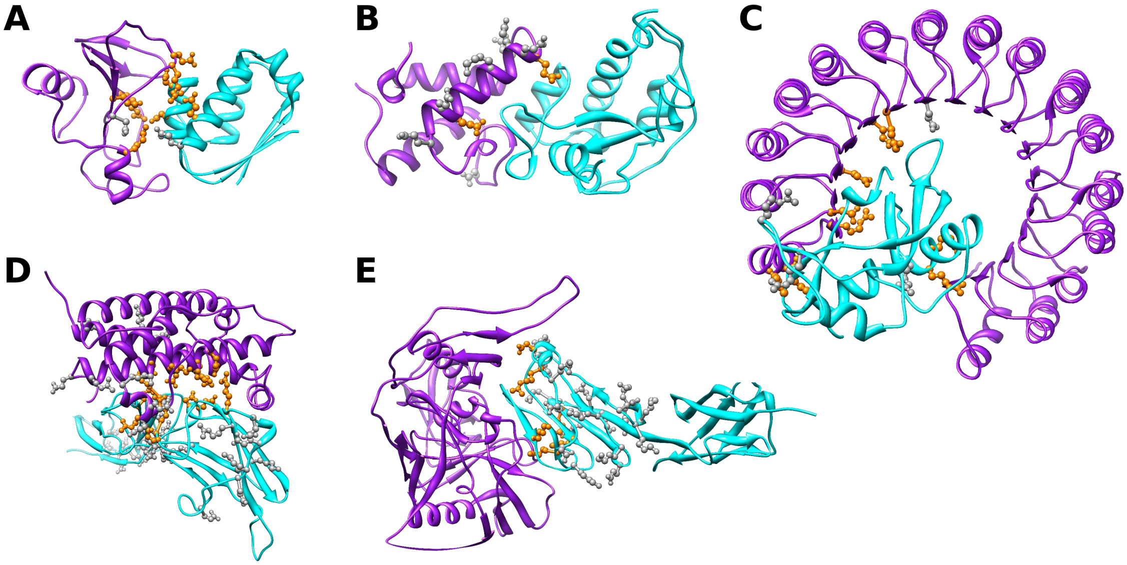


Figure 2.2 Molecular graphics representations of selected protein complexes. Proteins are shown in ribbon representation, with experimentally mutated residues displayed as ball and stick. The first and second proteins listed in the protein complex name are colored in purple and cyan, respectively. Residues within the interface are colored in orange, and those outside the interface are colored in gray. The protein complexes shown are (a) Barnase-Barstar, (b) Im9-E9 DNase, (c) RNase Inhibitor-Angiogenin, (d) hGH-hGHbp, and (e) gp120-CD4.

2.3 Results

We selected five systems for the evaluation of electrostatic free energy calculations and the evaluation of specific calculation criteria and parameters. Each protein complex is described briefly in Table 2.1 and depicted in Figure 2.2. As discussed above, the protein complexes were selected based on the availability of a co-crystal structure of the complex and sufficient experimental data to perform statistical correlations with our calculated free energy values. Coincidentally, we found that the dataset was indeed a diverse representation of charged protein systems. Barnase-Barstar (Figure 2.2A) and Im9-E9 DNase (Figure 2.2B) both represent highly and oppositely charged systems, which are relatively small in overall size. RNase Inhibitor-Angiogenin (Figure 2.2C) is a much larger protein complex, although the interfacial area is only ~60% larger than those of Barnase-Barstar and Im9/E9 DNase (Table 2.1). Each component is highly charged, which is thought to contribute to the extremely tight binding observed experimentally. Finally, the last two systems (hGH-hGHbp and gp120-CD4, Figures 2.2D and 2.2E) possess charged components of similar charge, thus overall electrostatic potential is hypothesized to have a diminished role in the long-range association of these protein complexes.

Using alanine-scanning mutagenesis and Poisson-Boltzmann electrostatics calculations, we calculated free energies of association and solvation of protein mutants for which experimental data were available. Based on the thermodynamic cycle in Figure 2.1, we calculated association free energy in both a reference ($\Delta G^{Coulombic}$) and solvated ($\Delta G^{solution}$) state. Additionally, we calculated $\Delta\Delta G^{solvation}$ for each mutant based on Eq. 4, which incorporates both association and solvation effects by closing the thermodynamic cycle of Figure 2.1. We used experimental binding free energies (determined from dissociation rate constants and Eq. 7) as a benchmark to evaluate each of the three free energy values calculated under various conditions.

Scatter plots illustrate the degree of correlation between calculated and experimental data (Figures B.1-B.10). It is evident that for Barnase-Barstar (Figures B.1-B.2) and Im9-E9 DNase (Figures B.3-B.4), the trend of calculated data closely parallels that of experimental data. It is of interest to note that the magnitudes of calculated electrostatic free energy changes upon mutation are significantly larger than those observed experimentally, particularly at low protein dielectric coefficients. This disparity can be attributed to the fact that only electrostatic interactions are accounted for, and the effects of van der Waals interactions and entropic effects are not directly included in the calculations. Despite large differences in magnitude, relative calculated and experimental free energies exhibit a high degree of correlation. RNase Inhibitor-Angiogenin shows slight correlation between calculated and experiemental data, however there is noticeably more scatter in the data (Figures B.5-B.6). The remaining two systems (hGH-hGHbp and gp120-CD4) show no discernable correlation (Figures B.7-B.10).

We performed electrostatics calculations under a wide variety of conditions, in order to evaluate the influence of protein dielectric, ionic strength, grid resolution, and linearization of the Poisson-Boltzmann equation. In order to quantify these effects and efficiently compare our results with experimental data, we determined statistical correlations between calculated and experimental free energy values under each condition. Pearson correlation coefficients are reported in Table 2.2 for each of the five mutated protein complex, under conditions including two different ionic strengths and four values of protein dielectric coefficient. ANOVA tests were used to test for statistical significance of correlations

(Table B3). The full analysis also incorporates the combinatorial effects of the above variables along with two grid dimensions and Poisson-Boltzmann linearization (Table B2).

Table 2.2. Correlation between calculated and experimental data for protein complexes.[a]

Ionic Strength		**ΔΔG (Solvation)**				**ΔG (Solution)**				**ΔG (Coulombic)**			
		ε_P=2	**ε_P=10**	**ε_P=20**	**ε_P=40**	**ε_P=2**	**ε_P=10**	**ε_P=20**	**ε_P=40**	**ε_P=2**	**ε_P=10**	**ε_P=20**	**ε_P=40**
0 mM	1	-0.83	-0.83	-0.83	-0.83	0.19	0.71	0.79	0.82	0.83	0.83	0.83	0.83
	2	-0.74	-0.71	-0.69	-0.66	0.17	0.81	0.81	0.78	0.73	0.73	0.73	0.73
	3	-0.49	-0.49	-0.48	-0.48	0.22	0.42	0.46	0.47	0.48	0.48	0.48	0.48
	4	-0.16	-0.17	-0.17	-0.17	-0.17	-0.07	0.01	0.09	0.13	0.13	0.13	0.13
	5	0.15	0.15	0.15	0.15	0.16	0.00	-0.07	-0.11	-0.13	-0.13	-0.13	-0.13
150 mM	1	-0.83	-0.83	-0.83	-0.83	0.06	0.59	0.74	0.81	--	--	--	--
	2	-0.73	-0.70	-0.66	-0.59	-0.50	0.92	0.94	0.95	--	--	--	--
	3	-0.49	-0.48	-0.48	-0.46	0.06	0.29	0.38	0.46	--	--	--	--
	4	-0.16	-0.15	-0.14	-0.11	-0.18	-0.10	-0.02	0.11	--	--	--	--
	5	0.15	0.16	0.17	0.18	0.24	0.27	0.25	0.18	--	--	--	--

[a] $\Delta\Delta G^{solvation}$, $\Delta G^{Coulombic}$, and $\Delta G^{solution}$ values were calculated according to Eqs. 4, 5, and 6. All free energy values were calculated using four different dielectric coefficients and two ionic strengths. Calculations were performed using coarse grid dimensions that equal protein dimensions multiplied by 1.5 and fine grid fit to protein dimensions. The numbers on the left refer to the following complexes: 1: Barnase-Barstar, 2: Im9-E9 DNase, 3: RNase Inhibitor-Angiogenin, 4: hGH-hGHbp, 5: gp120-CD4.

2.4 Discussion

In this study, we looked to evaluate the accuracy of Poisson-Boltzmann electrostatic calculations in predicting the association of proteins. In our initial work, we compared our calculated free energy values to published kinetic data for the Barnase-Barstar protein complex. Often considered a gold-standard system for electrostatics calculations, Barnase-Barstar mutational data has served as a benchmark for numerous computational studies (9-12, 14, 32, 34, 42-44, 50, 51, 69). Our initial analysis involved calculations in which the protein interior was represented by a low dielectric coefficient ($\varepsilon_P = 2$) and ionic strength of solution modeled physiological salt concentrations (I = 150 mM). Free energy calculations based on the thermodynamic cycle in Figure 2.1 yielded $\Delta G^{Coulombic}$, $\Delta G^{solution}$, and $\Delta\Delta G^{solvation}$ values for each experimental mutant. We observed strong correlation between $\Delta G^{Coulombic}$ and $\Delta\Delta G^{solvation}$ values with experimental data, but noted a lack of correlation with $\Delta G^{solution}$ (Table 2.2). This result was intriguing, since the bottom horizontal process of the thermodynamic cycle should most accurately represent protein-protein association in a physiological environment. We noted that the differences between the top and bottom of the thermodynamic cycle included dielectric and ionic screening. We have thoroughly examined both effects, as well as other parameters, in order to address the question of why $\Delta G^{solution}$ does not satisfactorily predict protein association.

We performed a series of calculations for Barnase-Barstar, in which we examined the effects of protein dielectric, ionic strength, grid resolution, and Poisson-Boltzmann linearization on the correlation between calculated and experimental free energy values. Our results showed that ionic screening had only a small effect on free energy values and correlations, whereas protein dielectric had a more dramatic effect (Table 2.2). The remaining two factors had negligible effects on the correlation values (Table B2). We found that both $\Delta\Delta G^{solvation}$ and $\Delta G^{Coulombic}$ correlate well with experimental free energy under all conditions with correlation magnitudes in agreement with data from Wang et al (44). The correlation of $\Delta G^{solution}$ values with experimental data yielded an interesting result; as protein dielectric was increased from 2 to 40, the correlation increased drastically (with a large increase in going from 2 to 10). Using a dielectric coefficient of 2, as discussed above, yielded no significant correlation in the data; however increasing the dielectric coefficient to 40 gave correlations similar to those obtained for $\Delta\Delta G^{solvation}$ and $\Delta G^{Coulombic}$. This result suggests the application of a higher dielectric coefficient to represent the protein interior.

In order to further explore the results from our Barnase-Barstar analysis, we examined four additional protein complexes from the Alanine Scanning Energetics Database. Similar correlations were observed for the Im9-E9 DNase protein complex (Table 2.2). Correlations of 0.7 or greater for both $\Delta\Delta G^{solvation}$ and $\Delta G^{Coulombic}$ indicated strong predictive capabilities of both Poisson-Boltzmann and Coulombic free energy calculations for the association of highly and oppositely charged systems, however $\Delta G^{solution}$ values yielded the strongest correlations overall. We observed a distinct trend in which $\Delta\Delta G^{solvation}$ correlations decrease slightly as protein dielectric increases. It is interesting to note that unlike Barnase-Barstar, $\Delta G^{solution}$ of Im9-E9 DNase (at dielectric coefficients of 10, 20, and 40) predicts experimental free energies with near perfect degrees of correlation, but shows very weak to moderate correlation at low values of protein dielectric. This result provides further evidence for the application of a higher dielectric coefficient ($\varepsilon_P > 2$) in the protein interior.

The RNase Inhibitor-Angiogenin complex yielded similar correlation trends as Barnase-Barstar, but with significantly weaker correlation coefficients (Table 2.2). Solvation and Coulombic free energies provided similar predictive capability for experimental results, whereas $\Delta G^{solution}$ always performed worse. A similar dielectric-dependent trend was observed for $\Delta G^{solution}$ correlation values, with a significantly weaker ionic strength-dependent effect. The correlations of RNase Inhibitor-Angiogenin are noticeably weaker than for Barnase-Barstar and Im9-E9 DNase. This is likely due to an inherent problem we have noticed in PB calculations. For systems involving two proteins which greatly differ in size and/or magnitude of charge, we noticed that a single alanine mutation of a large and highly charged protein will have a greatly reduced influence on electrostatic energy compared to a similar mutation on a less highly charged or smaller protein, even if the mutations are located at the protein-protein interface. In the case of RNase Inhibitor-Angiogenin, RNase Inhibitor is large and highly charged (460 residues, -22e) and Angiogenin is small and less highly charged (123 residues, +10e). Looking at interfacial residues alone, we can examine one residue from Angiogenin and one from RNase Inhibitor that form strong intermolecular salt bridges, which become disrupted upon mutation. D435A shows an increase in $\Delta G^{solution}$ ($\varepsilon_P = 40$) of 11.7 kJ/mol, whereas R492A shows an increase of 24.3 kJ/mol. Experimental data shows an opposite trend, in which R492A increases ΔG by 3.7 kJ/mol and D435A increases ΔG by 14.6 kJ/mol. We have observed this overpredicting behavior for mutants of small proteins in the past, and the reason for this discrepancy is unclear, however this phenomenon explains the relatively weak correlations observed for RNase Inhibitor-Angiogenin. Additionally, the data may be skewed by the low pH at which experimental data were collected (pH 6.0). Despite weaker correlations, the overall calculated and experimental free energy trends were similar to those observed for Barnase-Barstar, indicating that our free energy calculations exhibit consistent behavior for highly and oppositely charged systems.

We noted above that a discrepancy exists in the dielectric-dependent trends and magnitude of correlation between $\Delta G^{solution}$ and ΔG^{exp} for Im9-E9 DNase. We propose that this result can be explained by the proximity of experimental mutations to the protein-protein interface (Figure 2.2A-2.2C). Unlike both Barnase-Barstar and RNase Inhibitor-Angiogenin, most of the mutated residues in the Im9-E9 DNase complex are solvent exposed. Since these residues are located on the protein surface in both the non-complexed and complexed form of Im9, they will not experience a change in dielectric environment based on our calculation methods (Figure 2.1), and thus free energy values are not highly sensitive to variations in the dielectric environment of the protein interior. Also, since most mutations are located away from the interface, their primary contribution to protein complex association would arise from long-range interactions and contribution to the global electrostatic potential distribution of the protein. In contrast, Barnase-Barstar mutants are concentrated near the binding interface, and thus are more susceptible to environmental changes upon binding since they experience a change in dielectric environment between the uncomplexed and complexed form of the protein (Figure 2.1). In this case, we observe strongest correlations where the difference in dielectric screening upon binding is minimal.

We further tested Barnase-Barstar, Im9-E9 DNase, and RNase Inhibitor-Angiogenin, in order investigate whether mutation influences free energy correlation. In addition to calculating correlations for entire datasets, we looked at correlations of only interfacial residues for Barnase-Barstar (16 of 21 mutants) and RNase Inhibitor-Angiogenin (9 of 13 mutants), and only non-interfacial residues for Im9-E9 DNase (8 of 10 mutants). When non-interfacial residues were removed from free energy correlations, correlations and trends of all three free energy values for Barnase-Barstar and RNase Inhibitor-Angiogenin remained largely unchanged. However, when the two interfacial residues were removed from Im9-E9 DNase calculations, the correlations equalized for each of the three free energy measures (Tables B4-B6). The dielectric-dependent trend disappeared for $\Delta G^{solution}$ correlations, suggesting that residues at the protein-protein interface are sensitive to protein dielectric coefficient selection, whereas non-interfacial residues are not affected.

The final two systems tested were both comprised of highly charged proteins of similar net charge type. Correlation coefficients for hGH-hGHbp and gp120-CD4 are shown in Table 2.2. There are no significant correlations between calculated and experimental data, with all correlation coefficients below 0.3 and $p > 0.05$ for all cases (Table B3). Lack of correlation is also evinced in Figures B.7-B.10, which illustrate that no corresponding trend exists between the two sets of data. This result is not surprising, since gp120-CD4 and hGH-hGHbp are not oppositely charged systems. Electrostatic interactions are posited to play a dominant role in highly and oppositely charged systems, contributing significantly to the formation of a weak encounter complex. For similarly charged systems, electrostatic interactions often play an important role in complex binding and specificity, however the long-range electrostatic steering effect is lost. For example, in the hGH-hGHbp complex, there are as many or more salt bridges and hydrogen bonds stabilizing the interface as in the three oppositely charged systems (Table 2.3). While electrostatics is pivotal for binding and stability of this protein complex, it does not provide the long-range driving force as expected for Barnase-Barstar and like systems.

During our examination of free energy correlations, we noticed that free energy calculated from Coulombic potential performed nearly as well as both $\Delta\Delta G^{solvation}$ and $\Delta G^{solution}$ values calculated from Poisson-Boltzmann electrostatics. This finding holds true for nearly all cases, with the exception of Im9-E9 DNase, where $\Delta G^{solution}$ values under high protein dielectric conditions showed correlations of nearly unity. Based on this result, we propose that Coulombic free energy ($\Delta G^{Coulombic}$) can yield a computationally efficient approximation of relative binding affinities of protein complexes and corresponding mutants. Although this measure is not necessarily expected to reproduce experimental data, it would provide a rapid screen to narrow the search for mutants with desired binding activity.

Table 2.3. Intermolecular interactions within selected protein complexes.

Protein	Hydrogen Bond	Coulombic Interactions (< 5 Å)		Coulombic Interactions (< 8 Å)	
		Favorable	**Unfavorable**	**Favorable**	**Unfavorable**
Barnase Barstar	6	3	0	6	3
Im9 E9 Dnase	10	2	0	5	1
RNase Inhibitior Angiogenin	8	5	2	12	6
hGH hGHbp	11	5	0	9	4
gp120 CD4	6	2	0	7	5

Our results show strong correlations between electrostatic free energy calculations and experimental kinetic data for systems characterized by electrostatically-driven association. The strongest overall correlations were shown for the Barnase-Barstar and Im9-E9 DNase complexes, while RNase Inhibitor-Angiogenin showed moderate correlation as well. As expected, we observed no correlation between calculated and experimental free energy values for each of the two similarly charged systems. We observed that correlations with $\Delta G^{solution}$ are sensitive to the dielectric environment of the protein and whether or not the mutated residue was located in the interface. Finally, we observed that $\Delta G^{Coulombic}$ values closely predicted relative experimental free energies, despite the simplicity of calculation. Overall, these results suggest the use of electrostatic free energy calculations for an initial prediction of relative binding affinities of protein complexes and their mutants. While Poisson-Boltzmann electrostatics may provide an accurate representation of association events, these calculations are sensitive to the choice of protein dielectric coefficient and proximity to the interface. We suggest the use of a high dielectric coefficient ($\varepsilon_P \approx$ 20-40) in order to account for fluctuations in protein structure and solvent accessibility, as well as the polar character found at the interface of highly charged systems. In addition, we propose the use of Coulombic free energy calculations as a simple first approximation of relative protein association, as these calculations are shown to provide an accurate picture of association in highly and oppositely charged systems.

2.5 References

1. Perutz, M. 1978. Electrostatic effects in proteins. Science. 207: 1187–1191.
2. Warshel, A. 1981. Electrostatic basis of structure-function correlation in proteins. Acc. Chem. Res. 14: 284–290.
3. Davis, M.E., and J.A. McCammon. 1990. Calculating electrostatic forces from grid-calculated potentials. J. Comput. Chem. 11: 401–409.
4. Kangas, E., and B. Tidor. 1999. Charge optimization leads to favorable electrostatic binding free energy. Phys. Rev. E. 59: 5958–5961.
5. Sheinerman, F., and R. Norel. 2000. Electrostatic aspects of protein-protein interactions. Curr. Opin. Struct Biol. 10: 153–159.
6. Morozov, A.V., T. Kortemme, and D. Baker. 2003. Evaluation of models of electrostatic interactions in proteins. J. Phys. Chem. B. 107: 2075–2090.
7. Warshel, A., P.K. Sharma, M. Kato, and W.W. Parson. 2006. Modeling electrostatic effects in proteins. Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics. 1764: 1647–1676.
8. Lu, B.Z., Y.C. Zhou, M.J. Holst, and J.A. McCammon. 2008. Recent progress in numerical methods for the Poisson-Boltzmann equation in biophysical applications. Commun. Comput. Phys. 3: 973–1009.
9. Gabdoulline, R.R., and R.C. Wade. 1997. Simulation of the diffusional association of barnase and barstar. Biophys. J. 72: 1917–1929.
10. Gabdoulline, R.R., and R.C. Wade. 1998. Brownian dynamics simulation of protein–protein diffusional encounter. Methods. 14: 329–341.
11. Camacho, C.J., Z. Weng, S. Vajda, and C. DeLisi. 1999. Free energy landscapes of encounter complexes in protein-protein association. Biophys J. 76: 1166–1178.
12. Elcock, A., R. Gabdoulline, R. Wade, and J. McCammon. 1999. Computer simulation of protein-protein association kinetics: Acetylcholinesterase-fasciculin. J. Mol. Biol. 291: 149–162.
13. Adcock, S.A., and J.A. McCammon. 2006. Molecular dynamics: Survey of methods for simulating the activity of proteins. Chem. Rev. 106: 1589–1615.
14. Spaar, A., C. Dammer, R.R. Gabdoulline, R.C. Wade, and V. Helms. 2006. Diffusional encounter of barnase and barstar. Biophys. J. 90: 1913–1924.
15. Dong, F., and B. Olsen. 2008. Computational methods for biomolecular electrostatics. Methods in cell biology. 84: 843–870.
16. McCammon, J.A., S.H. Northrup, and S.A. Allison. 1986. Diffusional dynamics of ligand-receptor association. J. Phys. Chem. 90: 3901–3905.
17. Janin, J. 1997. The kinetics of protein-protein recognition. Proteins. 28: 153–161.

18. Selzer, T., and G. Schreiber. 2001. New insights into the mechanism of protein-protein association. Proteins. 45: 190–198.

19. Zhang, L., B. Mallik, and D. Morikis. 2007. Immunophysical exploration of C3d–CR2(CCP1-2) interaction using molecular dynamics and electrostatics. J. Mol. Biol. 369: 567–583.

20. Gilson, M.K., A. Rashin, R. Fine, and B. Honig. 1985. On the calculation of electrostatic interactions in proteins. J. Mol. Biol. 184: 503–516.

21. Gilson, M., and K. Sharp. 1988. Calculating the electrostatic potential of molecules in solution: Method and error assessment. J. Comput. Chem. 9: 327–335.

22. Warshel, A., and A. Papazyan. 1998. Electrostatic effects in macromolecules: fundamental concepts and practical modeling. Curr. Opin. Struct. Biol. 8: 211–217.

23. Wagoner, J., and N.A. Baker. 2004. Solvation forces on biomolecular structures: A comparison of explicit solvent and Poisson-Boltzmann models. J. Comput. Chem. 25: 1623–1629.

24. Baker, N.A. 2005. Improving implicit solvent simulations: a Poisson-centric view. Curr. Opin. Struct. Biol. 15: 137–143.

25. Honig, B., K. Sharp, and A.S. Yang. 1993. Macroscopic models of aqueous solutions: biological and chemical applications. J. Phys. Chem. 97: 1101–1109.

26. Simonson, T. 2001. Macromolecular electrostatics: continuum models and their growing pains. Curr. Opin. Struct. Biol. 11: 243–252.

27. Simonson, T. 2003. Electrostatics and dynamics of proteins. Rep. Prog. Phys. 66: 737–787.

28. Bertonati, C., B. Honig, and E. Alexov. 2007. Poisson-Boltzmann calculations of nonspecific salt effects on protein-protein binding free energies. Biophys. J. 92: 1891–1899.

29. Simonson, T. 1999. Dielectric relaxation in proteins: Microscopic and macroscopic models. Int. J. Quant. Chem. 73: 45–57.

30. Schutz, C.N., and A. Warshel. 2001. What are the dielectric "constants" of proteins and how to validate electrostatic models? Proteins. 44: 400–417.

31. Georgescu, R., and E. Alexov. 2002. Combining conformational flexibility and continuum electrostatics for calculating pKas in proteins. Biophys. J. 83, 1731–1748.

32. Dong, F., M. Vijayakumar, and H.-X. Zhou. 2003. Comparison of calculation and experiment implicates significant electrostatic contributions to the binding stability of barnase and barstar. Biophys. J. 85: 49–60.

33. Archontis, G., and T. Simonson. 2005. Proton binding to proteins: A free-energy component analysis using a dielectric continuum model. Biophys. J. 88: 3888–3904.

34. Dong, F., and H.-X. Zhou. 2006. Electrostatic contribution to the binding stability of protein-protein complexes. Proteins. 65: 87–102.

35. Nymeyer, H., and H.-X. Zhou. 2008. A method to determine dielectric constants in nonhomogeneous systems: Application to biological membranes. Biophys J. 94: 1185–1193.

36. Tjong, H., and H.-X. Zhou. 2008. On the Dielectric Boundary in Poisson−Boltzmann Calculations. J. Chem. Theory Comput. 4: 507–514.

37. Gilson, M.K., and B.H. Honig. 1986. The dielectric constant of a folded protein. Biopolymers. 25: 2097–2119.

38. Sham, Y.Y., I. Muegge, and A. Warshel. 1998. The effect of protein relaxation on charge-charge interactions and dielectric constants of proteins. Biophys J. 74: 1744–1753.

39. Hofinger, S., and T. Simonson. 2001. Dielectric relaxation in proteins: A continuum electrostatics model incorporating dielectric heterogeneity of the protein and time-dependent charges. J. Comput. Chem. 22: 290–305.

40. Simonson, T. 2008. Dielectric relaxation in proteins: the computational perspective. Photosynth. Res. 97: 21–32.

41. Elcock, A.H., D. Sept, and J.A. McCammon. 2001. Computer simulation of protein−protein interactions. J. Phys. Chem. B. 105: 1504–1518.

42. Lee, L.P., and B. Tidor. 2001. Optimization of binding electrostatics: charge complementarity in the barnase-barstar protein complex. Protein Sci. 10: 362–377.

43. Sheinerman, F. 2002. On the role of electrostatic interactions in the design of protein–protein interfaces. J. Mol. Biol. 318: 161–177.

44. Wang, T., S. Tomic, R.R. Gabdoulline, and R.C. Wade. 2004. How optimal are the binding energetics of barnase and barstar? Biophys. J. 87: 1618–1630.

45. Qin, S., and H.-X. Zhou. 2007. Do electrostatic interactions destabilize protein–nucleic acid binding? Biopolymers. 86: 112–118.

46. Cheung, A.S., C.A. Kieslich, J. Yang, and D. Morikis. 2010. Solvation effects in calculated electrostatic association free energies for the C3d-CR2 complex and comparison with experimental data. Biopolymers. 93: 509–519.

47. Kieslich, C.A., D. Morikis, J. Yang, and D. Gunopulos. 2011. Automated computational framework for the analysis of electrostatic similarities of proteins. Biotechnol. Prog. 27: 316–325.

48. Gorham, R.D., C.A. Kieslich, and D. Morikis. 2010. Electrostatic Clustering and Free Energy Calculations Provide a Foundation for Protein Design and Optimization. Ann. Biomed. Eng. 39: 1252–1263.

49. Thorn, K.S., and A.A. Bogan. 2001. ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. Bioinformatics. 17: 284–285.

50. Schreiber, G., and A.R. Fersht. 1993. Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. Biochemistry. 32: 5145–5150.

51. Schreiber, G., and A.R. Fersht. 1995. Energetics of protein-protein interactions - Analysis of the barnase-barstar interface by single mutations and double mutant cycles. J. Mol. Biol. 248: 478–486.

52. Wallis, R., K.Y. Leung, M.J. Osborne, R. James, G.R. Moore, et al. 1998. Specificity in protein-protein recognition: conserved Im9 residues are the major determinants of stability in the colicin E9 DNase-Im9 complex. Biochemistry. 37: 476–485.

53. Shapiro, R., and B.L. Vallee. 1992. Identification of functional arginines in human angiogenin by site-directed mutagenesis. Biochemistry. 31: 12477–12485.

54. Chen, C.Z., and R. Shapiro. 1997. Site-specific mutagenesis reveals differences in the structural bases for tight binding of RNase inhibitor to angiogenin and RNase A. Proc. Natl. Acad. Sci. U.S.A. 94: 1761–1766.

55. Shapiro, R. 2000. Analysis of the interactions of human ribonuclease inhibitor with angiogenin and ribonuclease A by mutagenesis: importance of inhibitor residues inside versus outside the C-terminal "hot spot." J. Mol. Biol. 302: 497–519.

56. Cunningham, B., and J. Wells. 1989. High-resolution epitope mapping of hgh-receptor interactions by alanine-scanning mutagenesis. Science. 244: 1081–1085.

57. Bass, S., M. Mulkerrin, and J. Wells. 1991. A systematic mutational analysis of hormone-binding determinants in the human growth-hormone receptor. Proc. Natl. Acad. Sci. U.S.A. 88: 4498–4502.

58. Ashkenazi, A., L.G. Presta, S.A. Marsters, T.R. Camerato, K.A. Rosenthal, et al. 1990. Mapping the CD4 binding site for human immunodeficiency virus by alanine-scanning mutagenesis. Proc. Natl. Acad. Sci. U.S.A. 87: 7150–7154.

59. Berman, H.M., J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, et al. 2000. The protein data bank. Nucleic Acids Res. 28: 235–242.

60. Buckle, A., G. Schreiber, and A. Fersht. 1994. Protein-protein recognition - Crystal structural-analysis of a barnase barstar complex at 2.0-angstrom resolution. Biochemistry. 33: 8878–8889.

61. Kleanthous, C., A.M. Hemmings, U.C. Kühlmann, A.J. Pommer, N. Ferguson, et al. 1999. Structural and mechanistic basis of immunity toward endonuclease colicins. Nat. Struct. Biol. 6: 243–252.

62. Papageorgiou, A.C., R. Shapiro, and K.R. Acharya. 1997. Molecular recognition of human angiogenin by placental ribonuclease inhibitor—an X-ray crystallographic study at 2.0 Å resolution. EMBO J. 16: 5162–5177.

63. de Vos, A.M., M. Ultsch, and A.A. Kossiakoff. 1992. Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. Science. 255: 306–312.

64. Kwong, P.D., R. Wyatt, J. Robinson, R.W. Sweet, J. Sodroski, et al. 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature. 393: 648–659.

65. Pettersen, E.F., T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, et al. 2004. UCSF Chimera: A visualization system for exploratory research and analysis. J. Comput. Chem. 25: 1605–1612.

66. Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics. 23: 2947–2948.

67. Marti-Renom, M.A., A.C. Stuart, and A. Fiser. 2000. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29: 291-325.

68. Vriend, G. 1990. WHAT IF: a molecular modeling and drug design program. J. Mol. Graphics. 8: 52–56.

69. Kieslich, C.A., R.D. Gorham Jr, and D. Morikis. 2011. Is the rigid-body assumption reasonable? Insights into the effects of dynamics on the electrostatic analysis of barnase–barstar. J. Non-Cryst. Sol. 357: 707–716.

70. Dolinsky, T.J., J.E. Nielsen, J.A. McCammon, and N.A. Baker. 2004. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res. 32: W665–W667.

71. Sitkoff, D., K.A. Sharp, and B. Honig. 1994. Accurate calculation of hydration free energies using macroscopic solvent models. J. Phys. Chem. 98: 1978–1988.

72. Baker, N.A., D. Sept, S. Joseph, M.J. Holst, and J.A. McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. U.S.A. 98: 10037–10041.

73. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: Visual molecular dynamics. J Mol Graphics. 14: 33–38.

74. Rogers, N.K., and M. Sternberg. 1984. Electrostatic interactions in globular proteins: Different dielectric models applied to the packing of α-helices. J. Mol Biol. 127: 527–542.

CHAPTER 3: COMPLEMENT INHIBITION BY STAPHYLOCOCCUS AUREUS: ELECTROSTATICS OF C3D-EFB-C AND C3D-ECB ASSOCIATION

3.1 Introduction

While the human immune system has evolved an intricate network of defense mechanisms to prevent infection, bacteria, viruses, and other microbial pathogens have developed numerous strategies to overcome these barriers. *Staphylococcus aureus* expresses a myriad of virulence factors that facilitate immune evasion, leading to infections and inflammatory response (1-4). To promote its survival, S. aureus expresses several proteins that interfere with complement activation (see Chapter 1.4). Efb and Ecb are two secreted proteins that mediate complement function on the bacterial surface (3, 5-7). In contrast to other staphylococcal virulence factors (i.e. SCIN, CHIPS), their mechanism of action remains somewhat unclear. Recent structural work elucidated the interaction of Efb-C (the C-terminal domain of Efb) and Ecb with complement C3d (7, 8). Biophysical and experimental data shows that Efb and Ecb bind to all C3d-containing proteins (C3, C3b, iC3b, and C3d), and that these interactions are consistently mediated through the TED (3, 7-10). Interestingly, these virulence factors inhibit complement via distinct mechanisms. Binding to C3b has been shown to stabilize interactions between proenzyme Factor B and C3b, preventing the formation of the alternative pathway C3 convertase (C3bBb) (10). More recently, Ecb was reported to recruit FH to the bacterial surface and downregulate complement activation (11). Alternatively, binding of Efb or Ecb to iC3b or C3d disrupts the C3d-CR2 interaction (6, 12, 13), which prevents C3d-mediated recruitment of B cells to the site of infection. These two virulence factors effectively inhibit both innate and adaptive immunological responses against *S. aureus*.

The C3d-Efb-C and C3d-Ecb interactions are of interest for therapeutic design. First, since Efb-C and Ecb are strong contributors to staphylococcal virulence, their interactions with C3d may serve as targets for the design of anti-infective molecules. Furthermore, since Efb-C and Ecb inhibit complement activation, they could provide templates for complement-targeted therapeutic design. In this analysis, we examine the virulence mechanism of Efb-C and Ecb, in light of recently published cocrystal structures (7, 8). Since both Efb-C and Ecb are excessively cationic (+7e and +11e respectively), we expect both overall

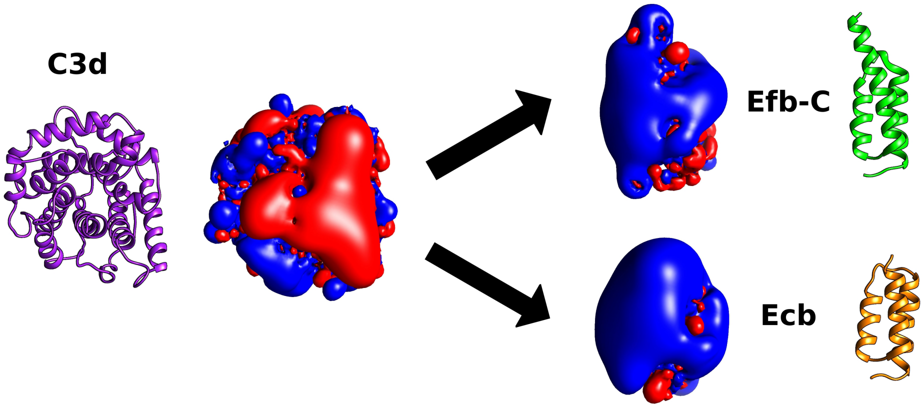


Figure 3.1 Electrostatically-driven association of C3d-Efb-C and C3d-Ecb complexes. Ribbon representations of C3d (purple), Efb-C (green), and Ecb (orange) are shown. Next to the ribbon structures, the spatial distributions of electrostatic potential of each protein are shown. Isopotential contours correspond to +1 k_bT/e (blue) and -1 k_bT/e (red), respectively. The proteins are oriented such that their binding sites face each other. Note the complementarity between basic regions of Efb-C/Ecb and the acidic region of C3d.

electrostatic potential and interfacial interactions will govern protein association and stability (Figure 3.1). According to the two-step model of protein association (14), protein-protein recognition is governed solely by long-range electrostatic interactions. Excessive charge differential between a protein and its target receptor is believed to enhance the probability of encounter complex formation (14-19). As a result, we predict that mutations of ionizable residues will influence protein complex formation, irrespective of location within Efb-C or Ecb. By analyzing the physicochemical properties of the C3d-Efb-C and C3d-Ecb complexes, we expect to delineate the crucial interactions that stabilize the complexes. We performed systematic mutagenesis *in silico*, and clustered mutants based on electrostatic similarity. We hypothesize that electrostatically similar mutants will exhibit similar functional characteristics as well. As a result, our analysis will aid in understanding the molecular basis for bacterial complement evasion by *S. aureus*, as well as providing a foundation for drug design in future studies.

3.2 Methods

3.2.1 Preparation and Equilibration of Molecular Structures

Three-dimensional coordinates were obtained from the protein data bank (PDB) (20) for the C3d-Efb-C (PDB code 2gox) (8) and C3d-Ecb (PDB code 2noj) (7) complexes. PDB files were downloaded, edited, and examined for missing residues. The C3d-Efb-C structure contained two complete heterodimers, of which one was selected (Chains A and B). The C3d-Ecb structure contained four heterodimers, with each missing numerous residues in the C3d portion of the complex. Chains G and H were chosen, since this dimer possessed a lower number of missing residues. Seven missing residues (in four separate locations of C3d) were reconstructed using Modeller (21) to generate a complete structure of the protein complex. Visual Molecular Dynamics (VMD) (22) was used to generate a protein structure file for each complex based on the CHARMM 27 forcefield (23). Complexes were solvated in explicit water boxes and periodic boundary conditions were applied to maintain continuity during the simulations. The system was neutralized using NaCl counterions, with the total number of ions selected to attain a total concentration of 150 mM. Each protein complex was subjected to 1000 steps of conjugate gradient energy minimization, followed by a 10 ps molecular dynamics equilibration run using Nanoscale Molecular Dynamics (NAMD) (24). Minimization and equilibration were performed on all atoms in each complex, with no restraints. Simulations were performed using PM Ewald electrostatics with 1 Å grid spacing. The SHAKE algorithm was enabled to maintain rigid bonds between hydrogen and heavy atoms, allowing the use of 2 fs time steps. The total energy and backbone RMSD of each protein complex was examined over the course of the dynamics simulation, to ensure adequate equilibration and structural relaxation.

3.2.2 Alanine Scanning Mutagenesis

Alanine scan mutants of each protein complex were generated using equilibrated structures to comparatively examine the effects of electrostatics in protein association. Alanine mutations were used to efficiently examine the role of individual charged residues within each protein, while minimizing structural perturbations. Charges and atomic radii were added to PDB files for each of the wild-type protein complexes using PDB2PQR (25) and the PARSE force field (26). No further optimization in the structures of the mutants was performed. A recent study showed that although post-mutation dynamic relaxation led to noticeable structural fluctuations, it did not yield significant improvement in correlation between calculated and experimental data for another set of single-alanine protein mutants (27). The resulting PQR files contain atomic coordinates from the PDB files in addition to charges and radii for each atom. Each

PQR file was processed using a custom R script, which singly mutated each charged amino acid (arginine, lysine, aspartic acid, and glutamic acid) to alanine, and split each mutated protein complex into its component proteins. Mutations were generated by removing all side chain atoms with the exception of Cβ and its hydrogen atoms, while placing a third hydrogen atom along the previous Cβ-Cγ vector. Mutations were performed in the complex structure, which was subsequently separated to its components for free energy and clustering calculations.

3.2.3 Calculation of Electrostatic Potential

The Adaptive Poisson-Boltzmann Solver (APBS) (28) was used to calculate spatial distribution of electrostatic potentials for each set of PQR files (six total). The protein molecular surface was calculated using a probe sphere with radius 1.4 Å and the ion accessibility surface was defined using a probe sphere with radius 2.0 Å. Calculations were carried out using a dielectric constant of 40 within the protein interior, and 78.54 in the solvent. The value of 40 for interior protein dielectric was selected based on the results of our previous analysis and references therein (29-32). Ion concentrations of 0 mM and 150 mM were used to simulate pure solvent and physiological serum, respectively. Grid sizes and spacing were determined using the psize Python script included with APBS. For specifics on grid size and spacing throughout this analysis, refer to Table C.1 in the Supporting Material. Mutants of C3d, ligand, and complex were centered within each grid based on their respective parent structures.

3.2.4 Clustering of Mutants

Clustering techniques can be used for comparison of protein structures and characteristics (33-35). All protein mutants were clustered based on electrostatic similarity, using an localized difference (LD) measure and hierarchical clustering as discussed previously (27, 36-38). LD values were calculated according to:

$$LD = \frac{1}{N} \sum_{i,j,k} \frac{|\varphi_B(i,j,k) - \varphi_A(i,j,k)|}{max(|\varphi_B(i,j,k)|,|\varphi_A(i,j,k)|)} \quad (1).$$

In Eq. 1, φ_A and φ_B refer to electrostatic potential at point (i,j,k) in mutant proteins A and B, respectively, and N represents the total number of grid points at which electrostatic potential has been calculated. LD values were calculated for mutants and parent structures of C3d and ligands (but not the complex), using all data points from the electrostatic potential output from APBS. An n × n matrix was generated containing the LD values of all possible mutant pairs, where n is the number of mutants plus the parent. Mutants were clustered using the hclust hierchical clustering function in R utilizing average linkage and dendrograms were generated. Based on the LD metric, electrostatically similar mutants cluster together and dissimilar mutants cluster apart. Similar mutants had LD values near zero, and LD magnitude increased with dissimilarity.

3.2.5 Calculation of Electrostatic Free Energy

Free energy calculations were performed using APBS. PQR files were used as the input files for APBS, and electrostatic potential was calculated. All structures were centered based on the coordinates of the parent protein complex, in order to eliminate grid artifacts and ensure accurate comparison of relative association of protein mutants. Free energy was calculated from partial charges and electrostatic potential according to:

$$G_{elec} = \frac{1}{2}\sum_i q_i \varphi_i \qquad (2).$$

In Eq. 2, q_i and φ_i refer to the charge and electrostatic potential at point i within the grid containing the protein. Free energy of complex association and solvation was calculated according to the thermodynamic cycle in Figure 2.1. The free energy was calculated for a given protein complex and its components in both a reference and solvated state, and reported free energy values were calculated according to:

$$\Delta G^{ref} = G_{AB}^{ref} - G_A^{ref} - G_B^{ref} \qquad (3),$$

$$\Delta G^{solu} = G_{AB}^{solu} - G_A^{solu} - G_B^{solu} \qquad (4),$$

$$\Delta\Delta G^{solvation} = \Delta G_{AB}^{solvation} - \Delta G_{A}^{solvation} - \Delta G_{B}^{solvation} = \Delta G^{solu} - \Delta G^{ref} \quad (5).$$

In these equations, *AB* refers to the protein complex (C3d-Efb-C or C3d-Ecb), *A* refers to free C3d, and *B* refers to the free ligand (Efb-C or Ecb). The quantity described by Eq. 5 incorporates the effects of both association and solvation to the overall process of complex formation. Additionally, calculation of $\Delta\Delta G^{solvation}$ values eliminate grid artifacts, which arise due to self-energies from distribution of a single charge across multiple grid points. In order to determine accurate values for ΔG^{ref} and ΔG^{solu} without grid artifacts, we used the APBS external program Coulomb to calculate Coulombic potential of each protein and protein complex, with $\Delta G^{Coulombic}$ calculated according to:

$$\Delta G^{Coulombic} = G_{AB}^{Coulombic} - G_{A}^{Coulombic} - G_{B}^{Coulombic} \quad (6).$$

Using this value in place of ΔG^{ref}, we were able to calculate $\Delta G^{solution}$ according to:

$$\Delta G^{solution} = \Delta\Delta G^{solvation} + \Delta G^{Coulombic} \quad (7).$$

It should be noted that since $\Delta\Delta G^{solvation}$ is the difference between ΔG^{solu} and ΔG^{ref}, sign convention dictates that higher values indicate more favorable association, and lower values indicate less favorable association.

3.3 Results

We examined the interaction between two proteins from *S. aureus* (Efb-C and Ecb) and the thioester domain of complement system protein C3 (C3d). Through a systematic theoretical alanine scan of charged residues, we elucidated the role of electrostatics in the association of each protein complex. There are numerous intermolecular salt bridges and long-range Coulombic interactions that stabilize the C3d-Efb-C and C3d-Ecb complexes (Tables 3.1 and 3.2). While electrostatic interactions may facilitate formation of protein complexes in excessively and oppositely charged systems, they often have a destabilizing effect upon overall complex formation, due to the energetic penalty of desolvation associated with binding (17, 18).

Table 3.1 Intermolecular Coulombic interactions in C3d-Efb-C complex.

Interaction	Distance (Å)
K1050-D156	3.3
E1160-K110	3.5
E1032-R165	3.7
E1156-K135	3.9
E1030-R131	3.9
D1029-R131	4.0
E1035-R165	4.1
E1160-K106	5.6
E1047-D156	**6.0**
E1159-K106	6.2
D1096-K148	6.2
R1042-R131	**6.6**
K1284-K135	**6.8**
E1159-K110	7.8
K1284-R131	**7.9**

Table 3.2 Intermolecular Coulombic interactions in C3d-Ecb complex.

Interaction	Distance (Å)
K1050-D100	3.6
E1160-K54	3.6
E1032-K109	3.7
D1156-R79	3.9
D1029-R75	3.9
E1030-R75	4.4
D1096-K92	6.6
R1042-R75	**7.3**
E1035-K109	7.3
K1284-R75	**7.7**

We calculated electrostatic free energies of association and solvation for each wild-type protein complex and their mutants, in order to assess the roles of solvation and electrostatics in protein complex formation. Specifically, we calculated $\Delta G^{Coulombic}$, $\Delta G^{solution}$, and $\Delta\Delta G^{solvation}$ values to assess the influence of solvation on relative free energy values. A recent comprehensive analysis of Poisson-Boltzmann free energy calculations applied to single-alanine mutants revealed a strong correlation between calculated values and experimental association data for excessively and oppositely charged protein complexes (27, 29, 30, 37, 39), thus we expect that our calculations will serve as a predictor of binding properties of Efb-C and Ecb mutants.

The structures of C3d-Efb-C and C3d-Ecb are shown in Figure 3.2. It is evident that both ligands share analogous interactions with C3d, with Ecb sharing less interactions due to truncation of helix 1. Ecb possesses only 4 acidic residues, yielding a net charge of +11e. Its highly positive electrostatic potential distribution facilitates its association with the acidic interface of C3d. We clustered spatial distribution of electrostatic potential for single alanine mutants of C3d, Efb-C, and Ecb. Figure 3.3 shows the clustering dendrograms for Efb-C under conditions of 0 mM and 150 mM ionic strength, as well as the relative free energy values of each mutant. Mutants of acidic residues to alanine (acidic mutants) and mutants of basic

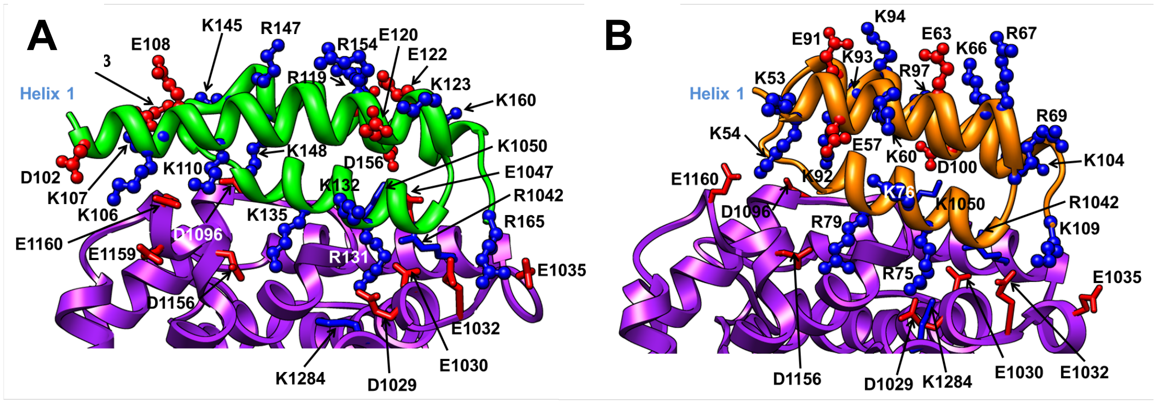

A
B
Helix 1
3
E108
K145
R147
R154
E120
E122
R119
K123
K160
K1050
K148
D156
E1047
D102
K107
K110
K132
R1042
K106
K135
R165
E1160
D1096
E1159
R131
E1035
D1156
E1032
K1284
D1029
E1030
Helix 1
E91
K94
K93
E63
R67
K66
K53
R97
R69
K54
E57
K60
D100
K104
E1160
K92
K76
D1096
K1050
R1042
R79
K109
R75
E1035
D1156
D1029
K1284
E1030
E1032

Figure 3.2 Molecular graphics representations of C3d-Efb-C and C3d-Ecb. (a) Molecular representation of the binding interface of C3d-Efb-C. C3d (purple) and Efb-C (green) protein backbones are shown as ribbons. (b) Molecular representation of the binding interface of C3d-Ecb. C3d (purple) and Ecb (orange) protein backbones are shown as ribbons. Acidic (red) and basic (blue) residues are shown in a ball-and-stick representation for Efb-C and Ecb, and in stick representation for C3d. Helix 1 is labeled in both Efb-C (residues 101-124) and Ecb (residues 52-68). Only C3d residues that participate in intermolecular interactions < 8 Å are shown on C3d for clarity.

residues to alanine (basic mutants) cluster together, respectively. The differences in clustering between 0 mM and 150 mM ionic strength become evident in the subclusters within the acidic and basic superclusters. Since ions surrounding the protein are interacting with atoms at the protein surface, the charge distribution in a physiological ionic strength environment (150 mM) changes from the pure solvent case. As a result of this ionic screening, mutations have a much smaller effect on overall electrostatic potential of the protein. This phenomenon was previously shown experimentally, where association slowed with increasing ionic strength (9).

It is clear that Efb-C mutants involving residues that participate in intermolecular interactions with C3d cluster together, and have free energy values that differ significantly from wild-type Efb-C (parent). Based on an analysis of free energy values, mutants K110A, R131A, K135A, and R165A have a significantly reduced electrostatic affinity for C3d (Figures 3.3C-3.3H). This result is similar to the finding in Haspel et al., which indicated that interactions involving these residues were important for the formation of an energetically stable complex (9). As expected, the mutated residues in these mutants are involved in medium-strength salt bridging interactions (Figures 3.2 and 3.3, Table 3.1). Based on the clustering dendrograms, the above mutants are separated into two clusters within the basic supercluster; R131A, K132A, K135A, R165A and K106A, K107A, K110A (Figures 3.3A and 3.3B). Both subclusters contain

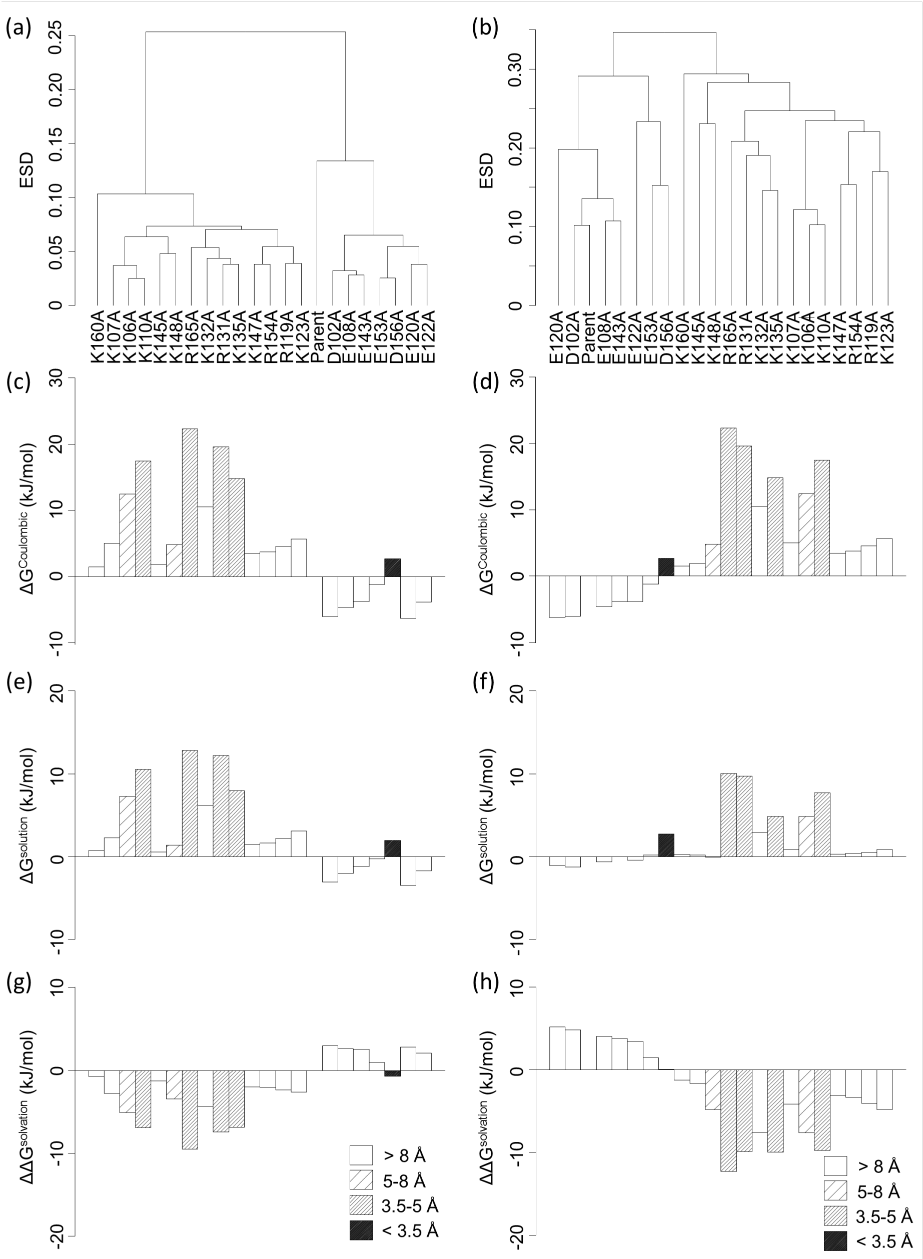


Figure 3.3 Clustering and free energy results for C3d-Efb-C. (a) Clustering of Efb-C mutants at 0 mM ion concentration. (b) Clustering of Efb-C mutants at 150 mM ion concentration. (c) $\Delta G^{\text{Coulombic}}$ values of C3d-Efb-C complex mutants at 0 mM ion concentration. (d) $\Delta G^{\text{Coulombic}}$ values of C3d-Efb-C complex mutants at 150 mM ion concentration. (e) $\Delta G^{\text{solution}}$ values of C3d-Efb-C complex mutants at 0 mM ion concentration. (f) $\Delta G^{\text{solution}}$ values of C3d-Efb-C complex mutants at 150 mM ion concentration. (g) $\Delta\Delta G^{\text{solvation}}$ values of C3d-Efb-C complex mutants at 0 mM ion concentration. (h) $\Delta\Delta G^{\text{solvation}}$ values of C3d-Efb-C complex mutants at 150 mM ion concentration. All free energy values are plotted relative to parent Efb-C. Interacting residues are denoted by bar patterns according to the distances in the legend.

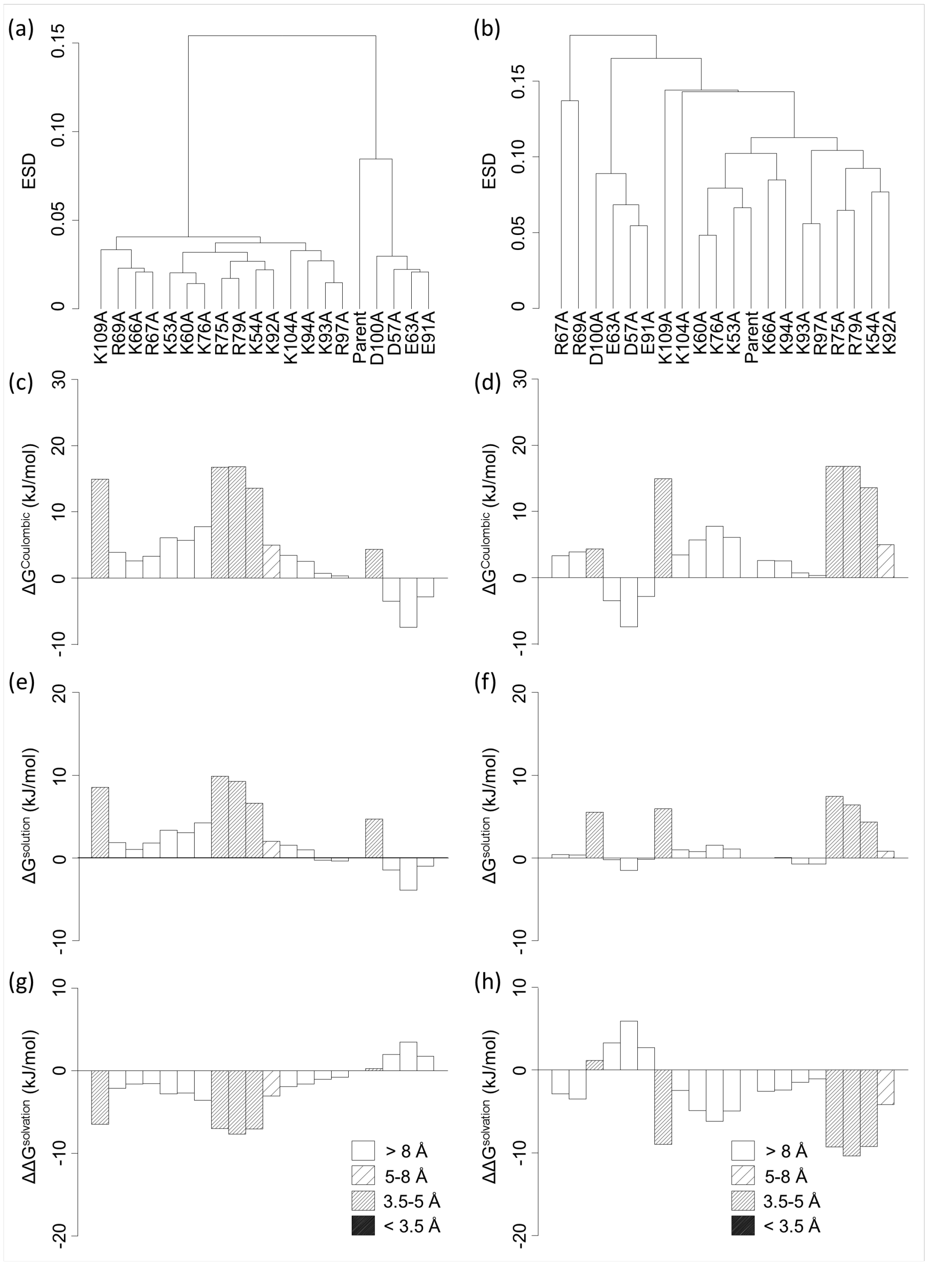


Figure 3.4 Clustering and free energy results for C3d-Ecb. (a) Clustering of Ecb mutants at 0 mM ion concentration. (b) Clustering of Ecb mutants at 150 mM ion concentration. (c) $\Delta G^{Coulombic}$ values of C3d-Ecb complex mutants at 0 mM ion concentration. (d) $\Delta G^{Coulombic}$ values of C3d-Ecb complex mutants at 150 mM ion concentration. (e) $\Delta G^{solution}$ values of C3d-Ecb complex mutants at 0 mM ion concentration. (f) $\Delta G^{solution}$ values of C3d-Ecb complex mutants at 150 mM ion concentration. (g) $\Delta\Delta G^{solvation}$ values of C3d-Ecb complex mutants at 0 mM ion concentration. (h) $\Delta\Delta G^{solvation}$ values of C3d-Ecb complex mutants at 150 mM ion concentration. All free energy values are plotted relative to parent Ecb. Interacting residues are denoted by bar patterns according to the distances in the legend.

residues within a respective local regions of positive electrostatic potential, thus mutations of residues within the same cluster will likely yield proteins with similar binding affinity for their target or receptor. Our results illustrate that clustering of electrostatic potential incorporates both structural and electrostatic proximity effects and the capacity of mutations to disrupt intermolecular interactions upon complex formation. While all basic mutants have higher $\Delta G^{Coulombic}$ and $\Delta G^{solution}$ values than the parent and $\Delta\Delta G^{solvation}$ values lower than the parent, the converse is typically true for acidic mutants.

Although charge differential is an indicator of the magnitude of attraction between two proteins, it is the spatial distribution of electrostatic potential and specific intermolecular interactions that contribute to the fine differences observed in clustering and free energy values of each mutant. For instance, D102A, E108A, and E143A cluster together under both 0 mM and 150 mM ionic strength conditions and have similar free energy values, whereas mutants E153A and D156A cluster on their own in both cases (Figures 3.3A and 3.3B). This result is due to the interaction network between E153, D156, and K1050 of C3d. While both mutants may modestly strengthen binding of C3d-Efb-C due to alteration of electrostatics, there is a balancing effect caused by the D156-K1050 salt bridge (Table 3.1) that will cause destabilization of C3d-Efb-C when D156 is mutated to alanine. Although E153 does not participate in intermolecular interactions with C3d, it is involved in a repulsive interaction with D156 (Table C.3), which effectively strengthens the D156-K1050 salt bridge. Residues E153 and D156 act together to stabilize C3d-Efb-C, despite their disruption of basicity near the Efb-C interface. Additionally, we observe differences between clustering of acidic mutants at 0 mM and 150 mM ionic strength. In 0 mM ionic strength, acidic mutants are divided among three subclusters, containing D102A, E108A, E143A (cluster 1); E120A and E122A (cluster 2); and E153A and E156A (cluster 3), respectively. In 150 mM ionic strength, clusters 1 and 3 remain intact, but E120A and E122A cluster separately. This may be due to the fact that although E120 and E122 are close in proximity, their free energy values differ, reflecting their respective locations relative to C3d and differences in intramolecular interactions within Efb-C.

Clustering and free energy trends are highly similar for Ecb mutants compared to mutants of Efb-C (Figure 3.4). This result should be expected, since Ecb shares nearly identical intermolecular interactions with C3d. There is a higher net charge differential between Ecb and C3d (11e) compared to Efb-C and C3d

(7e), which is responsible for an increased electrostatic free energy of the free Ecb and C3d proteins. Free energies of association, however, are of comparable magnitude to C3d-Efb-C, since the more favorable charge differential in C3d-Ecb is compensated by a lower number of stabilizing intermolecular interactions. There are four mutants of Ecb (K54A, R75A, R79A, K109A) that significantly reduce C3d-Ecb association (Fig. 3.4C-3.4H). The four mutated residues are analogous to K110, R131, R135, and R165 of Efb-C, as they share identical interactions and are located in the same relative positions. Unlike the case for Efb-C, where all four significant mutants cluster relatively closely, K109A does not cluster with K54A, R75A, and R79A (which cluster together). In both 0 mM and 150 mM ionic strength, K109A clusters on its own, but near other proximal basic residues. K109 does participate in a salt bridge with Ecb, but does not possess a bifurcated interaction observed for R165 of Efb-C. Also, it is closer to other basic mutants than R165, thus the similarity of K109 and R165 extends no further than a single intermolecular interaction. The acidic mutants of Ecb follow a similar trend to Efb-C as well, having lower $\Delta G^{Coulombic}$ and $\Delta G^{solution}$ and higher $\Delta\Delta G^{solvation}$ values than the parent. Mutant D100A is an exception to this trend, as it shares a salt bridge with K1050, similarly to D156 of Efb-C.

In addition to Efb-C and Ecb, clustering and free energy analyses were performed on C3d from each complex, yielding highly similar results (Figures C.1-C.16). Since intermolecular interactions are nearly identical in both complexes, C3d mutants have similar relative free energy values in both cases, with the magnitudes being slightly higher for C3d from C3d-Ecb. Mutations involving residues that participate in intermolecular interactions with Efb-C and Ecb cluster together in dendrograms of both C3d proteins, and have higher $\Delta G^{Coulombic}$ and $\Delta G^{solution}$ and lower $\Delta\Delta G^{solvation}$ values than other C3d mutants. This result indicates that both Coulombic and Poisson-Boltzmann calculations are capable of incorporating not only the effects of alterations in electrostatic potential distribution, but also specific interactions across the protein-protein interface.

3.4 Discussion

3.4.1 Electrostatic Association and Binding of C3d-Efb-C and C3d-Ecb

Recent studies show that Efb-C and Ecb from *S. aureus* inhibit complement activation, resulting in bacterial infection (7-10). Research efforts aimed at combating these types of bacterial infections often involve analysis of protein-protein interactions at a molecular level. In this analysis, we elucidated the role of electrostatic interactions in the association of complement protein fragment C3d with bacterial inhibitors Efb-C and Ecb from *S. aureus*. Since Efb-C and Ecb are excessively positively charged, we hypothesized that electrostatic interactions are important in driving recognition and binding to C3d. Previous studies have shown the efficacy of using electrostatic mutagenesis to improve binding, activity, and function of electrostatically complementary proteins (19, 29, 40-45), thus we employed this method here as a first step in therapeutic design. We found that electrostatically similar mutations clustered together, based on relative free energy values for mutants. In addition, a concurrent study showed that Poisson-Boltzmann free energy calculations are highly predictive of relative binding and activity of alanine mutants of excessively and oppositely charged protein complexes, providing further evidence supporting the efficacy of mutagenic clustering as a predictive tool for protein function (29, 30).

Our results give insight into the role of electrostatics in complex association and stability. First, and most intuitively, we observed that mutations of residues involved in intermolecular electrostatic interactions have the largest effects on free energy, indicating that the presence of these residues and their interacting partners on the target protein is vital for the formation of a thermodynamically stable protein complex. Secondly, we found that while mutations that reduce complex association are located within the complex interface, mutations that enhance association are typically located at larger distances from the protein-protein interface. There are two potential reasons for this finding. Some of the enhancing mutations involve residues that comprise part of an interfacial network of electrostatic interactions. These mutations typically strengthen interactions between each bacterial protein and C3d, thus stabilizing the bound complex. Other mutations are located opposite the binding interface, and contribute to complex association by altering the overall electrostatic potential of the protein, irrespective of their location within the protein.

Our results are supported by the findings of previous work with the C3d-Efb-C interaction. Hammel et al. showed that residues R131 and N138 formed a large number of intermolecular interactions with C3d. They pointed out that these residues are crucial for complex stability (8). Furthermore, later studies thoroughly investigated the characteristics of these mutants in comparison to wild-type Efb-C (9, 46). In accordance with the two-step binding model of protein association, the results of surface plasmon resonance experiments indicated that these mutants have little effect on association rates, but drastically increase dissociation rates for the complex with C3d. This indicates that highly-specific interfacial interactions are responsible for maintaining complex stability, while complex association is governed by the overall electrostatic potential distributions of C3d and Efb-C.

Further support of electrostatic association rate enhancement is given by data indicating that the IC_{50} value for native Ecb is about three times lower than that of Efb. Using isothermal titration calorimetry experiments, it was demonstrated that Ecb likely has two binding sites for C3d, with the second binding site proposed to involve helix 1 (Figure 3.2B) (7). The authors propose that the additional C3 binding capability of Ecb explains its enhanced complement inhibitory activity compared to Efb, however ELISA data was not shown for the mutant (N63E) that abolishes the second, weak-affinity binding site. We hypothesize that the increased potency of Ecb is, at least in part, due to its increased basicity compared to Efb-C. The crystal structure of Ecb contains the N63E mutation, to abolish the second low-affinity binding site for C3d. We examined our free energy results for mutant E63A, which approximately restores the electrostatic potential distribution of native Ecb. Our results illustrate a lower $\Delta G^{Coulombic}$ value (Figure 3.4D), indicating enhanced long-range interaction between Ecb and C3d, however the $\Delta G^{solution}$ value suggests that residue 63 has no effect on complex specificity (Figure 3.4F). Additional evidence comes from the residues missing at the N-terminus of Ecb in the crystal structure. Although the structure of this 22-residue region is unknown, it contains five basic and two acidic residues, thus native Ecb possesses a larger positive electrostatic potential than observed in the structure. Since the overall positive electrostatic potential is much larger for Ecb, we expect that it will associate faster and bind more strongly to C3d, enhancing its inhibitory function. It would be interesting to validate this hypothesis experimentally to demonstrate the importance of electrostatic potential as a predictive tool for protein function. Furthermore, based on the observation that

Ecb is electrostatically enhanced to bind C3d (compared to Efb-C), we can introduce additional alanine and charge reversal mutations (i.e. glutamic acid to lysine) away from the protein interface, in order to accelerate diffusional encounter of C3d and Ecb in solution.

Our comparison of clustering and free energy data for single-alanine mutants is capable of being used in the prediction of relative binding affinities, and in turn immunological activity. Based on the results in this analysis, we are able to select mutations that have a significant enhancing or reducing effect on complex binding, in order to achieve a gain or loss of function. This study serves as a guide in the selection of mutants for experimental analysis, as well as the foundation for further computational design of combinatorial mutants with significantly enhanced or reduced affinity for their targets. Additionally, using more rigorous structural and dynamics tools, we can design variants containing non-alanine mutations within the protein-protein interface, in order to improve geometric and physicochemical complementarity.

3.4.2 Immunological Activity of Efb-C and Ecb

Previous work indicates that both Efb-C and Ecb can interact with C3d, C3(H_2O), native C3, and C3b, with decreasing affinity for each target (7, 8). Interaction with C3d is implicated in blocking adaptive immune response by preventing CR2 from binding (12, 13), and subsequently precluding recruitment of B- and T-cells to the site of *S. aureus* infection. Interaction with hydrolyzed C3 and C3b results in complement inactivation by preventing formation of the alternative pathway C3 convertase. Hammel et al. pointed out that Efb-C and Ecb are sterically blocked from binding free C3b in its native conformation, despite experimental evidence suggesting otherwise (8). We examined the structure of free C3b, noting that the thioester domain interacts via part of its concave acidic face with macroglobulin domain 1 (MG1), which has a region of positive electrostatic potential distribution. It is possible that the excessive charge of Efb-C and Ecb can compete with MG1 interactions, causing a minor conformational change in C3b that will allow binding of the *S. aureus* proteins. A recent study confirmed that Efb-C is capable of separating the C3d domain from MG1 using small-angle X-ray scattering experiments (47). It is also unclear how Efb-C and Ecb can affect Factor B binding to C3b. Experimental evidence suggests that proenzyme Factor B takes an open conformation when initially binding C3b, with complement control protein (CCP) modules contacting several domains of C3b, including the CUB domain (48-53). With CUB being adjacent to the thioester domain, the alteration in local electrostatic potential of C3b upon binding of Efb-C or Ecb may facilitate stronger intermolecular interactions between Factor B and C3b. New evidence suggests that Efb-C induces conformational changes in the C3d domain that indeed propagate through domains to which the Ba fragment of Factor B is proposed to bind (47). Additionally, we observed that shared contacts exist between C3d-Efb-C, C3d-Ecb, and C3b-Factor H (CCP 4) (52), indicating that Efb and Ecb may play a role in preventing Factor H-mediated inactivation of C3b. Finally, it is unclear how binding of Efb-C and Ecb to native C3 will inactivate complement activation. It is known that these interactions prevent C3 cleavage, but likely via a different mechanism than in C3b-Efb-C and C3b-Ecb interactions. We posit that binding of Efb-C or Ecb to substrate C3 in the convertase exosite (54) may have an electrostatic allosteric effect on the proposed dynamic movement of the Factor B serine protease domain to the C3 cleavage site.

3.5 Concluding Remarks

Staphylococcus aureus possesses numerous proteins that facilitate evasion of the immune system. In the case of virulence factors Efb-C and Ecb, mutations that enhance or reduce association can potentially be used as a basis for the development of a therapeutic agent that will disrupt the interaction of C3d with Efb and Ecb, reducing the severity of hospital-associated MRSA infections. Furthermore, these bacterial proteins can serve as a template for the design of a potent inhibitor of complement activation in cases of autoimmune disorders, in which complement is activated in an unregulated manner.

3.6 References

1. Foster, T.J. 2005. Immune evasion by staphylococci. Nat. Rev. Micro. 3: 948–958.

2. Sibbald, M.J.J.B., A.K. Ziebandt, S. Engelmann, M. Hecker, A. de Jong, et al. 2006. Mapping the pathways to Staphylococcal pathogenesis by comparative secretomics. Microbiol. Mol. Biol. Rev. 70: 755–788.

3. Jongerius, I., J. Kohl, M.K. Pandey, M. Ruyken, K.P.M. van Kessel, et al. 2007. Staphylococcal complement evasion by various convertase-blocking molecules. J. Exp. Med. 204: 2461–2471.

4. Wyatt, M.A., W. Wang, C.M. Roux, F.C. Beasley, D.E. Heinrichs, et al. 2010. Staphylococcus aureus nonribosomal peptide secondary metabolites regulate virulence. Science. 329: 294–296.

5. Lee, L.Y.L., X.W. Liang, M. Hook, and E.L. Brown. 2004. Identification and characterization of the C3 binding domain of the Staphylococcus aureus extracellular fibrinogen-binding protein (Efb). J. Biol. Chem. 279: 50710–50716.

6. Lee, L.Y.L., M. Höök, D. Haviland, R.A. Wetsel, E.O. Yonter, et al. 2004. Inhibition of complement activation by a secreted Staphylococcus aureus protein. J. Infect D. 190: 571–579.

7. Hammel, M., G. Sfyroera, S. Pyrpassopoulos, D. Ricklin, K.X. Ramyar, et al. 2007. Characterization of Ehp, a secreted complement inhibitory protein from Staphylococcus aureus. J. Biol. Chem. 282: 30051–30061.

8. Hammel, M., G. Sfyroera, D. Ricklin, P. Magotti, J.D. Lambris, et al. 2007. A structural basis for complement inhibition by Staphylococcus aureus. Nat. Immunol. 8: 430–437.

9. Haspel, N., D. Ricklin, B.V. Geisbrecht, L.E. Kavraki, and J.D. Lambris. 2008. Electrostatic contributions drive the interaction between Staphylococcus aureus protein Efb-C and its complement target C3d. Protein Sci. 17: 1894–1906.

10. Jongerius, I., B.L. Garcia, B.V. Geisbrecht, J.A.G. van Strijp, and S.H.M. Rooijakkers. 2010. Convertase inhibitory properties of Staphylococcal extracellular complement-binding protein. J Biol. Chem. 285: 14973–14979.

11. Amdahl, H., I. Jongerius, T. Meri, T. Pasanen, S. Hyvärinen, et al. 2013. Staphylococcal Ecb protein and host complement regulator factor H enhance functions of each other in bacterial immune evasion. J. Immunol. doi:10.4049/jimmunol.1300638

12. Isenman, D.E., E. Leung, J.D. Mackay, S. Bagby, and J.M.H. van den Elsen. 2010. Mutational analyses reveal that the Staphylococcal immune evasion molecule Sbi and complement receptor 2 (CR2) share overlapping contact residues on C3d: Implications for the controversy regarding the CR2/C3d cocrystal structure. J. Immunol. 184: 1946–1955.

13. Ricklin, D., S.K. Ricklin-Lichtsteiner, M.M. Markiewski, B.V. Geisbrecht, and J.D. Lambris. 2008. Cutting Edge: Members of the Staphylococcus aureus extracellular fibrinogen-binding protein family inhibit the interaction of C3d with complement receptor 2. J. Immunol. 181: 7463–7467.

14. McCammon, J.A., S.H. Northrup, and S.A. Allison. 1986. Diffusional dynamics of ligand-receptor association. J. Phys. Chem. 90: 3901–3905.

15. Dong, F., and B. Olsen. 2008. Computational methods for biomolecular electrostatics. Methods in cell biology. 84: 843–870.

16. Gabdoulline, R.R., and R.C. Wade. 1997. Simulation of the diffusional association of barnase and barstar. Biophys. J. 72: 1917–1929.

17. Lee, L.P., and B. Tidor. 2001. Optimization of binding electrostatics: charge complementarity in the barnase-barstar protein complex. Protein Sci. 10: 362–377.

18. Sheinerman, F., and R. Norel. 2000. Electrostatic aspects of protein-protein interactions. Curr. Opin. Struct. Biol. 10: 153–159.

19. Zhang, L., B. Mallik, and D. Morikis. 2007. Immunophysical exploration of C3d–CR2(CCP1-2) interaction using molecular dynamics and electrostatics. J. Mol. Biol. 369: 567–583.

20. Berman, H.M., J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, et al. 2000. The protein data bank. Nucleic Acids Res. 28: 235–242.

21. Marti-Renom, M.A., A.C. Stuart, and A. Fiser. 2000. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 28: 291–325.

22. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: Visual molecular dynamics. J. Mol. Graphics. 14: 33–38.

23. MacKerell, A.D., D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, et al. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B. 102: 3586–3616.

24. Phillips, J., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, et al. 2005. Scalable molecular dynamics with NAMD. J. Comput. Chem. 26: 1781–1802.

25. Dolinsky, T.J., J.E. Nielsen, J.A. McCammon, and N.A. Baker. 2004. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res. 32: W665–W667.

26. Sitkoff, D., K.A. Sharp, and B. Honig. 1994. Accurate calculation of hydration free energies using macroscopic solvent models. J. Phys. Chem. 98: 1978–1988.

27. Kieslich, C.A., R.D. Gorham Jr, and D. Morikis. 2011. Is the rigid-body assumption reasonable? Insights into the effects of dynamics on the electrostatic analysis of barnase–barstar. J. Non-Cryst. Sol. 357: 707–716.

28. Baker, N.A., D. Sept, S. Joseph, M.J. Holst, and J.A. McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. U.S.A. 98: 10037–10041.

29. Gorham, R.D., C.A. Kieslich, and D. Morikis. 2010. Electrostatic clustering and free energy calculations provide a foundation for protein design and optimization. Ann. Biomed. Eng. 39: 1252–1263.

30. Gorham Jr, R.D., C.A. Kieslich, A. Nichols, N.U. Sausman, M. Foronda, et al. 2011. An evaluation of poisson-boltzmann electrostatic free energy calculations through comparison with experimental mutagenesis data. Biopolymers. 95: 746–754.

31. Simonson, T. 2003. Electrostatics and dynamics of proteins. Rep. Prog. Phys. 66: 737–787.

32. Warshel, A., P.K. Sharma, M. Kato, and W.W. Parson. 2006. Modeling electrostatic effects in proteins. Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics. 1764: 1647–1676.

33. DiMaggio, P.A., S.R. McAllister, and C.A. Floudas. 2010. Enhancing molecular discovery using descriptor-free rearrangement clustering techniques for sparse data sets. AIChE J. 56: 405–418.

34. Wade, R.C., and R.R. Gabdoulline. 2001. Protein interaction property similarity analysis. Int. J. Quant. Chem. 83: 122–127.

35. Zhang, X., C.L. Bajaj, B. Kwon, T.J. Dolinsky, J.E. Nielsen, et al. 2006. Application of new multiresolution methods for the comparison of biomolecular electrostatic properties in the absence of global structural similarity. Multiscale Model. Simul. 5: 1196–1213.

36. Jain, A.K., M.N. Murty, and P.J. Flynn. 1999. Data clustering: a review. ACM Comput. Surv. 31: 264–323.

37. Kieslich, C.A., D. Morikis, J. Yang, and D. Gunopulos. 2011. Automated computational framework for the analysis of electrostatic similarities of proteins. Biotechnol. Prog. 27: 316–325.

38. Petke, J. 1993. Cumulative and discrete similarity analysis of electrostatic potentials and fields. J. Comput. Chem. 8: 928–933.

39. Cheung, A.S., C.A. Kieslich, J. Yang, and D. Morikis. 2010. Solvation effects in calculated electrostatic association free energies for the C3d-CR2 complex and comparison with experimental data. Biopolymers. 93: 509–519.

40. Mallik, B., L. Zhang, S. Koide, and D. Morikis. 2008. pH Dependence of stability of the 10th human fibronectin type III domain: A computational study. Biotechnol. Prog. 24: 48–55.

41. Morikis, D., A.K. Rizos, D.A. Spandidos, and E. Krambovitis. 2007. Electrostatic modeling of peptides derived from the V3-loop of HIV-1 gp120: implications of the interaction with chemokine receptor CCR5. Int. J. Mol. Med. 19: 343–351.

42. Pyaram, K., C.A. Kieslich, V.N. Yadav, D. Morikis, and A. Sahu. 2010. Influence of electrostatics on the complement regulatory functions of Kaposica, the complement inhibitor of Kaposi's Sarcoma-associated herpesvirus. J. Immunol. 184: 1956–1967.

43. Qureshi, N.N., D. Morikis, and N.L. Schiller. 2011. Contribution of specific amino acid changes in penicillin binding protein 1 to amoxicillin resistance in clinical Helicobacter pylori isolates. Antimicrob. Agents Chemother. 55: 101–109.

44. Sfyroera, G., M. Katragadda, D. Morikis, S.N. Isaacs, and J.D. Lambris. 2005. Electrostatic modeling predicts the activities of orthopoxvirus complement control proteins. J Immunol. 174: 2143–2151.

45. Zhang, L., and D. Morikis. 2006. Immunophysical properties and prediction of activities for vaccinia virus complement control protein and smallpox inhibitor of complement enzymes using molecular dynamics and electrostatics. Biophys. J. 90: 3106–3119.

46. Haspel, N., B.V. Geisbrecht, J. Lambris, and L. Kavraki. 2009. Multi-scale characterization of the energy landscape of proteins with application to the C3D/Efb-C complex. Proteins. 78: 1004–1014.

47. Chen, H., D. Ricklin, M. Hammel, B.L. Garcia, W.J. McWhorter, et al. 2010. Allosteric inhibition of complement function by the Staphylococcal immune evasion protein Efb. Proc. Natl. Acad. Sci. U.S.A. 107: 17621–17626.

48. Gros, P., F.J. Milder, and B.J.C. Janssen. 2008. Complement driven by conformational changes. Nat. Rev. Immunol. 8: 48–58.

49. Ruiz-Gómez, G., J. Lim, M.A. Halili, G.T. Le, P.K. Madala, et al. 2009. Structure−activity relationships for substrate-based inhibitors of human complement factor B. J. Med. Chem. 52: 6042–6052.

50. Torreira, E., A. Tortajada, T. Montes, S. Rodriguez de Cordoba, and O. Llorca. 2009. 3D structure of the C3bB complex provides insights into the activation and regulation of the complement alternative pathway convertase. Proc. Natl. Acad. Sci. U.S.A. 106: 882–887.

51. Torreira, E., A. Tortajada, T. Montes, S. Rodriguez de Cordoba, and O. Llorca. 2009. Coexistence of closed and open conformations of complement factor B in the alternative pathway C3bB(Mg2+) proconvertase. J. Immunol. 183: 7347–7351.

52. Wu, J., Y.-Q. Wu, D. Ricklin, B.J.C. Janssen, J.D. Lambris, et al. 2009. Structure of complement fragment C3b–factor H and implications for host protection by complement regulators. Nat. Immunol. 10: 729–734.

53. Xu, Y., S. Narayana, and J. Volanakis. 2001. Structural biology of the alternative pathway convertase. Immunol. Rev. 180: 123–135.

54. Rooijakkers, S.H.M., J. Wu, M. Ruyken, R. van Domselaar, K.L. Planken, et al. 2009. Structural and functional implications of the alternative complement pathway C3 convertase stabilized by a staphylococcal inhibitor. Nat. Immunol. 10: 722–728.

CHAPTER 4: MOLECULAR ANALYSIS OF THE INTERACTION BETWEEN STAPHYLOCOCCAL VIRULENCE FACTOR SBI-IV AND COMPLEMENT C3D

4.1 Introduction

Staphylococcus aureus (*S. aureus*) produces numerous toxins and virulence factors known to disrupt a wide array of biological functions (1, 2). Recent work has identified secreted proteins that disrupt function of the complement system (3-8), a branch of innate immunity responsible for phagocytosis of pathogens in the bloodstream. Sbi, a four-domain protein associated with the cell wall of *S. aureus* (9), contains two domains (III and IV) that bind complement component 3 (C3) and its thioester domain (C3d)-containing fragments (6). Functional studies indicate that an Sbi construct comprised only of domains III and IV (Sbi-III/IV) leads to rapid consumption of C3 in the bloodstream, preventing C3 from being activated to C3b and opsonizing the bacterial surface. Interestingly, domain IV of Sbi (Sbi-IV) on its own specifically inhibits activation of the complement alternative pathway (AP) by preventing active convertase formation, representing an entirely different mechanism of action (6).

The structures of Sbi-IV, both free and in complex with C3d, were recently reported (10). Dual binding modes were reported for Sbi-IV, with binding observed at the concave acidic region (Complex 1) and at the convex thioester-containing region (Complex 2). Mutagenesis studies of Sbi-IV binding and function, together with astounding similarity to binding modes of staphylococcal complement inhibitors Efb-C and Ecb at the concave acidic region (11, 12), suggest that Complex 1 is responsible for AP inhibition (10, 13). Nearly all known ligands of C3d bind at or near the concave acidic region (Figure D.1). This region is suggested to be evolutionarily conserved, facilitating complement-mediated recruitment of B cells to bacterial surfaces via interaction with complement receptor 2 (CR2) (14). The second proposed binding site of Sbi-IV (Complex 2) represents a novel C3d interaction site (Figure D.1). If Sbi-IV indeed binds to the convex thioester-containing region, this binding mode could represent a previously uncharacterized complement evasion tactic by pathogens. However, both the physiological significance and existential probability of Complex 2 remain unclear.

Macromolecular crystallography provides invaluable insight into the structural basis of protein interactions and how these interactions translate into functional phenomena observed *in vitro* and *in vivo*. However, one major limitation involves capturing molecules of interest in static conformations. Crystallization of proteins with their binding partners may induce nonphysiological binding conformations due to crystallization conditions and crystal packing effects. Indeed, recent cocrystal structures involving complement proteins suggest binding arrangements that are somewhat contradictory (15-21). We describe herein the integration of a variety of computational tools to analyze two recently published crystallographic structures of C3d in complex with Sbi. We seek to answer the question of whether both Complexes 1 and 2 exist physiologically, by examining their dynamic stability and associative and dissociative properties.

4.2 Methods

4.2.1 Structure preparation

Three-dimensional cocrystal structures of each C3d-Sbi Complex 1 (PDB code 2WY8) and Complex 2 (PDB code 2WY7) were obtained from the Protein Data Bank (PDB) (22). In cases where multiple crystallographic positions were reported for residues, position A was selected.

4.2.2 Analysis of Electrostatic Similarities Of Proteins (AESOP)

AESOP (23-26) was used to quantify the effects of specific charged residues on C3d-Sbi association in Complexes 1 and 2. First, PDB2PQR (27) was used to add hydrogen atoms, partial charges, and atomic radii to each complex PDB structure, using the PARSE force field (28). Theoretical alanine scan mutagenesis was performed for all charged residues of the protein complexes, one at a time. For each mutated protein and protein complex, electrostatic potential maps and free energies were calculated using APBS (29), as described previously (23, 24, 26, 30). For electrostatics calculations, we used protein and solvent dielectric values of 20 and 78.54, respectively, and ionic strength values of 0 and 150 mM. The choice of the value of protein dielectric is justified in an earlier study (25). Electrostatic similarity clustering and calculation of electrostatic free energies of association and solvation were performed as described previously (23, 26).

4.2.3 Explicit solvent molecular dynamics (MD) simulations

MD simulations of Complexes 1 and 2 were performed using NAMD (31). PDB structures of each complex were prepared using the PSFGEN utility in VMD (32) and the CHARMM27 force field (33). Complexes 1 and 2 were solvated in a TIP3P water boxes measuring 87 Å × 87 Å × 87 Å and 82 Å × 82 Å × 82 Å respectively. Sodium and chloride counterions were added to neutralize the charge of each system, with an ionic strength of 150mM. Each system was minimized using 25000 steps of conjugate gradient energy minimization, followed by heating from 0 to 300 K (62 ps total), four equilibration stages (50 ps each) with all protein atoms harmonically constrained (using force constants of 10, 5, 2, and 1 kcal/mol/Å^2 respectively) to their post-minimization positions, a final equilibration stage (50 ps) with protein backbone atoms harmonically constrained (using force constant of 1 kcal/mol/Å^2), followed by the production run. All simulations were performed using periodic boundary conditions and particle mesh Ewald electrostatics with a nonbonded interaction cutoff and switching distance of 12 and 10 Å, respectively. The SHAKE algorithm was employed to fix the length of bonds involving hydrogen atoms, enabling the use of 2 fs integration time steps. Langevin temperature and pressure controls were used for all NPT ensemble simulations. For each system, one 100 ns trajectory and three 20 ns trajectories were performed, using the energy minimized coordinates of each system as the starting point for all simulations, with different random number generated seeds.

4.2.4 Steered molecular dynamics (SMD) simulations

SMD simulations were also performed using NAMD. For each system, the solvent accessible surface of C3d atoms within 8 Å of Sbi and its normal vectors were calculated using UCSF Chimera (34). The mean of the normal vectors was used as the direction of induced Sbi dissociation from C3d. In order to facilitate SMD simulations and analysis, each complex was rotated such that the mean normal vector was aligned with the +z axis. Complexes 1 and 2 were solvated in TIP3P water boxes measuring 76 Å × 87 Å × 105 Å and 75 Å × 78 Å × 104 Å respectively, which provided additional space for Sbi dissociation in the +z direction. Each system was minimized and heated as described above, followed by a 500 ps equilibration stage with all protein atoms harmonically restrained using a force constant of 10 kcal/mol/Å^2.

Equilibrated structures were used as input for all SMD simulations. Five independent 1.6 ns SMD simulations were performed for each complex. Residues of C3d > 12 Å from Sbi were harmonically constrained during the simulations with a force constant of 10 kcal/mol/Å^2, and the center of mass of all Sbi atoms was constrained (with a force constant of 10 kcal/mol/Å^2) to a point moving at a constant velocity of 10 Å/ns in the +z direction.

4.2.5 Brownian dynamics (BD) simulations

BD simulations for Complexes 1 and 2 were performed using BrownDye (35). For each complex, PQR files and electrostatic potential maps were generated for C3d and Sbi as described above. In all simulations, C3d was the stationary molecule and Sbi was the diffusing molecule. Reaction criteria were selected based on a pilot comparative study using mutants of barnase-barstar and corresponding experimental k_{on} data (data not shown). Pairwise residue interactions were calculated from energy minimized structures of each complex, using cutoff distances of either 3.5 or 5.0 Å. Successful reactions in BD simulations required formation of at least 2 (or 3) interactions from the pairwise list, with interatomic distance < 6.0 (or 6.5) Å. The program 'bd_top' was used to generate all necessary input files for BD simulations, and 100000 trajectories were performed for each system under each set of reaction conditions. Reaction probabilities and corresponding k_{on} rates were calculated by BrownDye.

4.2.6 MD simulation analysis

MD trajectories were analyzed using the Bio3D package (36) and additional custom scripts in R. Solvent accessible surface area (SASA) calculations, analysis of hydrogen bonds, and visual PDB and MD trajectory analysis were performed with UCSF Chimera. SASA calculations were performed using structures without hydrogen atoms.

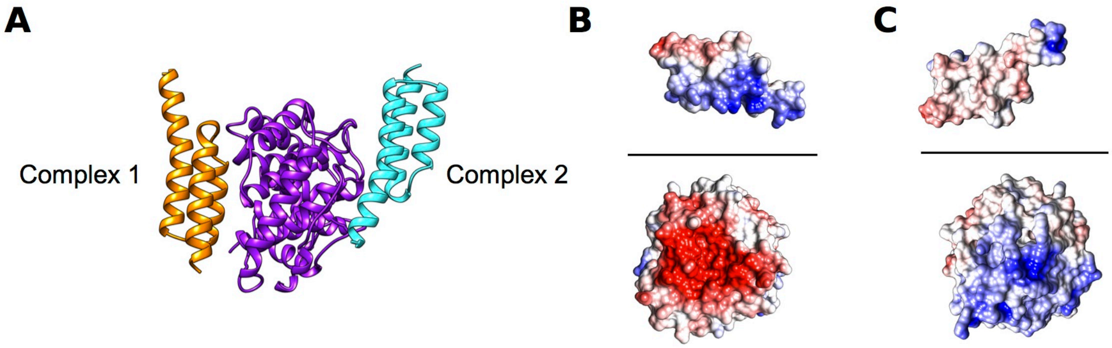


Figure 4.1 Molecular graphics representations of C3d-Sbi-IV complexes. (A) The two binding modes of Sbi-IV (orange and cyan) are shown bound to C3d (purple). Open-book representations of C3d-Sbi-IV Complex 1 (B) and Complex 2 (C) are shown, which illustrates electrostatic complementarity at each interface. C3d and Sbi-IV are rotated about the horizontal axis (black line) by -90 and +90 degrees, respectively. Molecular surfaces are colored based on electrostatic potential values, with a gradient from -5 k_BT/e (red) to +5 k_BT/e (blue).

4.3 Results

4.3.1 Electrostatic analysis

Sbi-IV is a three-helix bundle that exhibits binding to C3d via the acidic concave region (Complex 1) and the convex thioester-containing region (Complex 2), according to crystallographic structures (10). Figure 4.1A shows a comparison of both binding modes, with the Sbi-IV molecule from Complexes 1 and 2 on the left and right sides of C3d, respectively. The interfacial SASA of both complexes is similar (763 and 646 Å^2, respectively). There is electrostatic complementarity at both binding interfaces, however the complementarity is much more pronounced in Complex 1 (Figures 4.1B and 4.1C). To quantify the role of electrostatics in C3d-Sbi-IV association at each binding site, we performed AESOP analyses. Based on the magnitudes and perturbations of electrostatic free energies upon mutation of ionizable residues to alanine, it is clear that Complex 1 (Figure D.2) is much more electrostatically driven than Complex 2 (Figure D.3). There are six C3d mutations and seven Sbi-IV mutations that experience > 5 kJ/mol gain in association free energy in Complex 1, whereas in Complex 2, no C3d mutations and only two Sbi-IV mutations result in considerable changes in energy. These data reflect that the Complex 1 interface is significantly more polar than for Complex 2.

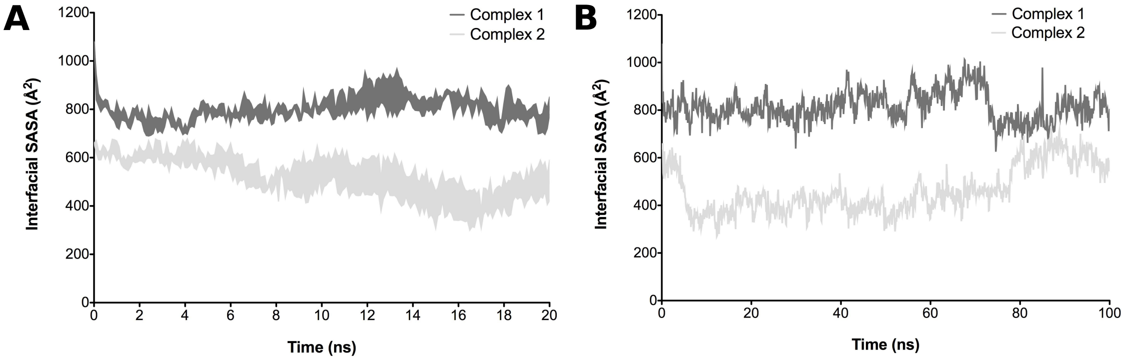


Figure 4.2 Change in interfacial SASA during MD trajectories of C3d-Sbi-IV complexes. (A) Interfacial SASA versus time for three independent 20 ns MD trajectories of Complexes 1 (dark gray) and 2 (light gray). (B) Interfacial SASA versus time for 100 ns MD trajectories of Complexes 1 and 2. The data points in (A) represent mean +/- SEM of interfacial SASA data.

4.3.2 MD simulations

We used explicit solvent molecular dynamics simulations to examine the dynamic stability of Complex 1 and 2 binding modes. Previous data suggested that Complex 1 is the primary binding mode for C3d-Sbi-IV (10). We observe that in general, interfacial SASA of Complex 1 remains constant over time, while Complex 2 undergoes significant loss of interfacial SASA (Figure 4.2). Despite these data, we did not observe complete dissociation of Complex 2 at any point, and the interfacial SASA tends to reach a steady-state within ~10 ns in all trajectories. Additionally, observed losses in interfacial SASA were not always permanent. Near the end of the 100 ns MD trajectory for Complex 2, we observed a sudden increase in interfacial SASA. Examination of this trajectory revealed unwinding of the N-terminus of Sbi-IV helix 1, which facilitated formation of charge-charge interactions that guided the N-terminus back to C3d.

Weakening of Complex 2 (relative to the crystallographic structure) can be also observed through the number of nonpolar interactions, hydrogen bonds, and salt bridges throughout MD trajectories (Figures D.4-D.6). Furthermore, while both complexes are stabilized by a comparable number of nonpolar interactions (Figures 4.3A and 4.3B), polar interactions are much more dominant at the Complex 1 interface (Figures 4.3C-4.3F). In the initial configuration of Complex 2, there is a relative physicochemical mismatch between residues at the N-terminus of Sbi-IV and the N-terminal loops and helices of C3d. N-terminal Sbi-IV residues preferentially engage in intramolecular interactions, including salt bridges and a hydrophobic cluster. As a consequence, N-terminal residues 198-203 of Sbi-IV tend to move away from

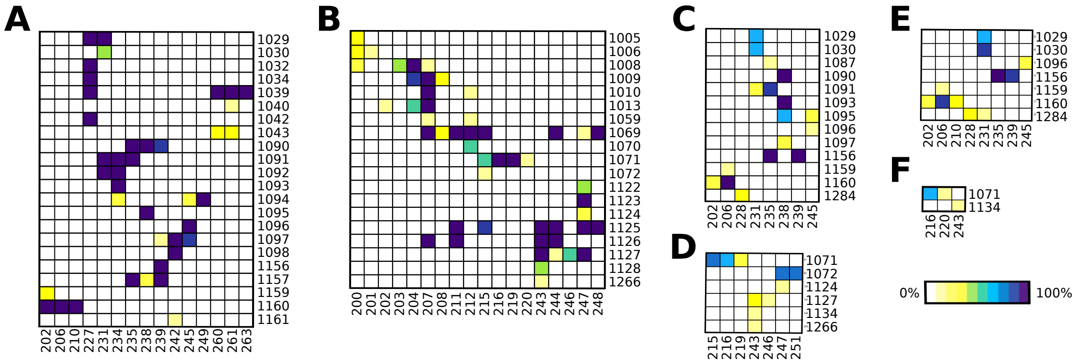


Figure 4.3 Interaction occupancy during MD trajectories of C3d-Sbi-IV complexes. Interactions were calculated at 100 ps intervals throughout one 100 ns and three 20 ns MD trajectories for Complexes 1 and 2. Occupancies represent the percentage of snapshots (across all trajectories) in which specific interactions were present. Residue numbers on the x- and y-axes represent residues of C3d and Sbi-IV, respectively. The legend shows the color scale from white (0% occupancy) to purple (100% occupancy). Interactions calculated include nonpolar interactions for Complex 1 (A) and Complex 2 (B), hydrogen bonds for Complex 1 (C) and Complex 2 (D), and salt bridges for Complex 1 (E) and Complex 2 (F).

C3d, while nonpolar interactions in the C-terminal region of helix 1 and N-terminal region of helix 3 hold the complex together (Figure 4.3B). In contrast, intermolecular interactions are well conserved throughout the MD trajectories for Complex 1. The Sbi-IV N-terminus is less prone to dynamic instability; Arg and Lys residues (K202, R206, and R210) switch between stabilizing helix 1 of Sbi-IV and forming salt bridging interactions with C3d residues. Notably, Sbi-IV residues responsible for interaction with C3d in Complex 1 are more dispersed, consisting of positively charged residues in the N-terminal region of helix 1, all of helix 2, and the N- and C-terminal regions of helix 3 (Figure 4.3A, 4.3C, 4.3F).

4.3.3 BD simulations

Since equilibrium binding of protein complexes is comprised of association and dissociation phases, we used computational tools to probe each of these phases. Brownian dynamics simulations show that Complex 1 associates more than an order of magnitude faster than Complex 2 (Figure 4.4). To verify that these data were not dependent on selected reaction criteria, we first calculated k_{on} values for mutants of barnase-barstar via BD simulations, using a diverse set of reaction criteria. We compared our results to experimental k_{on} data for barnase-barstar, and identified reaction criteria that yielded the closest agreement between calculated and experimental data (37). While BD simulations depict electrostatically-driven diffusion accurately, there are inherent limitations in such simulations. BD simulations are necessarily

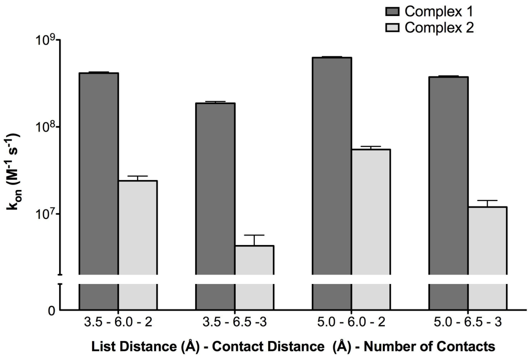


Figure 4.4 Comparison of association rates of C3d-Sbi-IV complexes. Rates were calculated by Brownian dynamics simulations under four conditions, with varied BD parameters including the cutoff distance for listing candidate atomic contact pairs (list distance), the distance required for successful atomic contact (contact distance), and the number of atomic contacts (number of contacts). Dark gray and light gray bars represent calculated k_{on} values (with 95% confidence intervals) for Complex 1 and Complex 2, respectively.

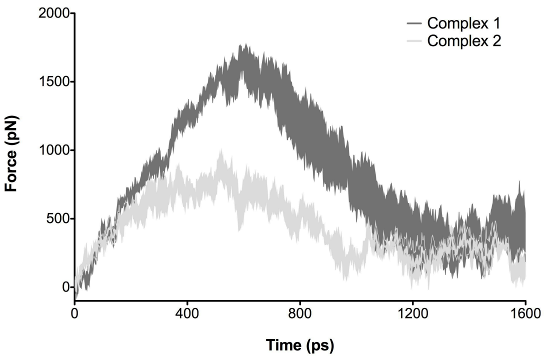


Figure 4.5 Unbinding forces for C3d-Sbi-IV complexes. Steered MD simulations were performed for each complex, in which the center of mass of Sbi was pulled away from C3d at a rate of 10 Å/ns, and exerted force was reported every 20 fs. The area of the curves represent mean +/- SEM of the force on Sbi in Complex 1 (dark gray) and Complex 2 (light gray) for five independent SMD trajectories.

"rigid-body" simulations, in which each binding partner remains in its static (bound) conformation. This assumption is adequate for long-range encounter complex formation. Once proteins encounter each other, in a near-native orientation, successful binding is almost inevitable, since residue side chains are locked in their binding configurations. Thus, the difference in association rates between Complexes 1 and 2 may be larger, especially since our MD data suggest that interactions observed from the crystal structure of Complex 2 may be inadvertently strengthened by crystal packing effects.

4.3.4 SMD simulations

To examine the dissociation phase of C3d-Sbi-IV interactions, we induced unbinding using SMD simulations (38, 39). In each simulation, the center of mass of Sbi-IV was pulled away from C3d at a constant velocity, and the force acting on Sbi-IV was calculated throughout the trajectories. The force required to "pull" Sbi-IV from C3d is ~2-fold larger for Complex 1 compared with Complex 2, indicating that the latter interaction is less stable. This result is confirmed in multiple SMD trajectories (Figure 4.5). Additionally, by examining the number of intermolecular interactions over time, we observe that loss of interactions occurs gradually (Figures D.7-D.9), which may explain the gentle slope of the force curves after the critical force.

4.4 Discussion

We have implemented a computational framework comprised of electrostatic free energy calculations, MD, BD, and SMD simulations to shed light on the binding mode between staphylococcal virulence factor Sbi and complement fragment C3d. In accordance with extensive biophysical analyses for the C3d-Sbi-IV interaction (10), our data suggest that Sbi-IV likely binds via the mode observed in Complex 1. Complex 1 is stabilized by numerous polar and nonpolar interactions over a large area, most of which are conserved throughout MD trajectories and are energetically favorable. Additionally, the interfacial SASA of Complex 1 remains constant. In contrast, there is significant loss of interfacial SASA for Complex 2; the N-terminus of Sbi-IV tends to move away from C3d, while the C-terminus of helix 1 and N-terminus of helix 3 hold the complex together, primarily via nonpolar interactions. BD and SMD simulations illustrate that Complex 1 is a markedly more stable complex, as it is characterized by more favorable association and dissociation kinetics compared to Complex 2.

While binding and mutagenesis studies showed that Complex 1 is likely the physiologically relevant binding mode, NMR chemical shift perturbation experiments identified interacting residues in both binding modes (10). Interestingly, none of the N-terminal residues (E201, V205, and E209) that were shown to have large chemical shifts in previous NMR experiments participated in persistent interactions with C3d in our analysis, however V244 and L248 (which also exhibited large chemical shifts) do indeed share persistent interactions with I1125 on C3d. Other Sbi-IV residues deemed to be most critical to C3d interaction in Complex 2 by our analysis (N215, D216, S219, D243, H247, and Q251) did not show large chemical shifts.

Although Sbi-IV is an AP complement inhibitor on its own, full-length Sbi can inhibit all three pathways of complement. Previous studies have shown that domains I and II of Sbi bind immunoglobulin G (IgG) molecules via their Fc region, conferring protection of *S. aureus* against antibody-mediated opsonization, while domains III and IV are involved in C3 binding (6). Recent structural analyses were unsuccessful in determining structure of Sbi-III/IV, and it was reported that Sbi-III has an inherently disordered structure (10, 13). Since Sbi-III/IV (and full-length Sbi) exhibit a different mechanism of complement evasion compared to Sbi-IV alone, it is necessary to examine our results in light of this information.

Similarly to Sbi-IV, Sbi-III/IV contains an interaction site on the thioester domain of C3, and exhibits similar binding preferences for thioester domain-containing C3 fragments (iC3b ~ C3d > C3 > C3b). Clark et al. suggested that the secondary binding site of Sbi-IV on C3d (Complex 2) may represent the binding mode when other domains of Sbi are also present (10). Based on the observation that the N-terminus of Sbi-IV (residues 198-205) make strong contacts with C3d in the crystallographic structure, they suggest that presence of Sbi-III may stabilize this weak binding site. We observed that the N-terminus of Sbi-IV moves away from C3d almost immediately in all MD simulations of Complex 2, indicating that intermolecular interactions of these residues are not strong enough to keep the complex together.

In context of structural information for C3 and C3b, it is unlikely that Sbi-III/IV binds to C3 via Complex 2 binding mode, as there are extensive steric clashes. The thioester bond on C3d is intrinsically shielded from solvent in native C3, precluding Sbi-IV from binding at that site, unless major conformational changes occur. In the Complex 1 site, Sbi-IV binds easily without clashes (Figure D.10A). However, in C3b, the Complex 2 binding site appears to be the preferred mode, as Complex 1 shows clashes with C3b (Figure D.10B). Additionally, Sbi-III/IV binds to both C3d and C3a in native C3, which is likely simultaneous based on the slow biphasic association kinetics observed in SPR experiments (6). Although the Sbi-IV molecule in Complex 2 would be spatially closer to C3a than in Complex 1, steric clashes and Sbi-IV orientation likely preclude simultaneous binding. Through construction of a basic homology model of Sbi-E (domains I-IV), it is plausible that Sbi-III and Sbi-IV (in Complex 1 binding mode) can simultaneously bind C3a and C3d, respectively (Figure D.10C). Since Sbi-III/IV can form a covalent adduct to nascent C3b (6), binding of Sbi-III/IV in an arrangement bridging the region of C3 containing the thioester bond would facilitate rapid transacylation upon C3 cleavage.

Although our data demonstrate much higher stability for Complex 1 compared to Complex 2, and together with experimental data suggest that the structure of Complex 1 is physiological, Sbi-IV alone may be capable of binding to C3d via the Complex 2 binding mode. Despite weak interactions at the termini of Sbi-IV, there is a strong central core of polar and nonpolar interactions that remains bound during MD simulations. We suggest that Sbi-IV can provide a foundation for therapeutic development. C3b recognizes and covalently attaches to both host and pathogen cell surfaces, and host protection is reliant on tight regulation by other complement factors. Several autoimmune diseases are propagated by inadequate regulation, and in turn, deposition of complement on host cells. Rational structure-based design of Sbi-IV variants may lead to novel therapeutics targeting C3b deposition on host surfaces.

4.5 References

1. Foster, T.J. 2005. Immune evasion by staphylococci. Nat. Rev. Micro. 3: 948–958.

2. Finlay, B.B., and G. McFadden. 2006. Anti-Immunology: Evasion of the host immune system by bacterial and viral pathogens. Cell. 124: 767–782.

3. Lee, L.Y.L., X.W. Liang, M. Hook, and E.L. Brown. 2004. Identification and characterization of the C3 binding domain of the Staphylococcus aureus extracellular fibrinogen-binding protein (Efb). J. Biol. Chem. 279: 50710–50716.

4. Jongerius, I., J. Kohl, M.K. Pandey, M. Ruyken, K.P.M. van Kessel, et al. 2007. Staphylococcal complement evasion by various convertase-blocking molecules. J. Exp. Med. 204: 2461–2471.

5. Rooijakkers, S.H.M., M. Ruyken, A. Roos, M.R. Daha, J.S. Presanis, et al. 2005. Immune evasion by a staphylococcal complement inhibitor that acts on C3 convertases. Nat. Immunol. 6: 920–927.

6. Burman, J.D., E. Leung, K.L. Atkins, M.N. O'Seaghdha, L. Lango, et al. 2008. Interaction of human complement with Sbi, a staphylococcal immunoglobulin-binding protein - Indications of a novel mechanism of complement evasion by Staphylococcus aureus. J. Biol. Chem. 283: 17579–17593.

7. Langley, R., B. Wines, N. Willoughby, I. Basu, T. Proft, et al. 2005. The Staphylococcal superantigen-like protein 7 binds IgA and complement C5 and inhibits IgA-FcαRI binding and serum killing of bacteria. J. Immunol. 174: 2926-2933.

8. de Haas, C.J.C., K.E. Veldkamp, A. Peschel, F. Weerkamp, W.J.B. van Wamel, et al. 2004. Chemotaxis inhibitory protein of Staphylococcus aureus, a bacterial antiinflammatory agent. J. Exp. Med. 199: 687–695.

9. Zhang, L., K. Jacobsson, J. Vasi, M. Lindberg, and L. Frykberg. 1998. A second IgG-binding protein in Staphylococcus aureus. Microbiology. 144: 985–991.

10. Clark, E.A., S. Crennell, A. Upadhyay, A.V. Zozulya, J.D. Mackay, et al. 2010. A structural basis for Staphylococcal complement subversion: X-ray structure of the complement-binding domain of Staphylococcus aureus protein Sbi in complex with ligand C3d. Mol. Immunol. 48: 452–462.

11. Hammel, M., G. Sfyroera, D. Ricklin, P. Magotti, J.D. Lambris, et al. 2007. A structural basis for complement inhibition by Staphylococcus aureus. Nat. Immunol. 8: 430–437.

12. Hammel, M., G. Sfyroera, S. Pyrpassopoulos, D. Ricklin, K.X. Ramyar, et al. 2007. Characterization of Ehp, a secreted complement inhibitory protein from Staphylococcus aureus. J. Biol. Chem. 282: 30051–30061.

13. Upadhyay, A., J.D. Burman, E.A. Clark, E. Leung, D.E. Isenman, et al. 2008. Structure-function analysis of the C3 binding region of Staphylococcus aureus immune subversion protein Sbi. J. Biol. Chem. 283: 22113–22120.

14. Kieslich, C.A., and D. Morikis. 2012. The two sides of complement C3d: Evolution of electrostatics in a link between innate and adaptive immunity. PLoS Comput Biol. 8: e1002840.

15. Szakonyi, G., J.M. Guthridge, D. Li, K. Young, V.M. Holers, and X.S. Chen. 2001. Structure of complement receptor 2 in complex with its C3d ligand. Science. 292: 1725–1728.

16. Isenman, D.E., E. Leung, J.D. Mackay, S. Bagby, and J.M.H. van den Elsen. 2010. Mutational analyses reveal that the Staphylococcal immune evasion molecule Sbi and complement receptor 2 (CR2) share overlapping contact residues on C3d: Implications for the controversy regarding the CR2/C3d cocrystal structure. J. Immunol. 184: 1946–1955.

17. van den Elsen, J.M.H., and D.E. Isenman. 2011. Complement receptor complex structure. Science. 332: 608–611.

18. Morgan, H.P., C.Q. Schmidt, M. Guariento, B.S. Blaum, D. Gillespie, et al. 2011. Structural basis for engagement by complement factor H of C3b on a self surface. Nat. Struct. Mol. Biol. 18: 463–470.

19. Kajander, T., M.J. Lehtinen, S. Hyvärinen, A. Bhattacharjee, E. Leung, et al. 2011. Dual interaction of factor H with C3d and glycosaminoglycans in host-nonhost discrimination by complement. Proc. Natl. Acad. Sci. U.S.A. 108: 2897–2902.

20. Gros, P. 2011. In self-defense. Nature. 18: 401–402.

21. Morikis, D. F1000.com/10371956.

22. Berman, H., J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, et al. 2000. The protein data bank. Nucleic Acids Res. 28: 235-242.

23. Kieslich, C.A., D. Morikis, J. Yang, and D. Gunopulos. 2011. Automated computational framework for the analysis of electrostatic similarities of proteins. Biotechnol. Prog. 27: 316–325.

24. Kieslich, C.A., R.D. Gorham Jr, and D. Morikis. 2011. Is the rigid-body assumption reasonable? Insights into the effects of dynamics on the electrostatic analysis of barnase–barstar. J. Non-Cryst Sol. 357: 707–716.

25. Gorham Jr, R.D, C.A. Kieslich, A. Nichols, N.U. Sausman, M. Foronda, et al. 2011. An evaluation of poisson-boltzmann electrostatic free energy calculations through comparison with experimental mutagenesis data. Biopolymers. 95: 746–754.

26. Gorham, R.D., C.A. Kieslich, and D. Morikis. 2010. Electrostatic clustering and free energy calculations provide a foundation for protein design and optimization. Ann. Biomed. Eng. 39: 1252–1263.

27. Dolinsky, T.J., J.E. Nielsen, J.A. McCammon, and N.A. Baker. 2004. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res. 32: W665–W667.

28. Sitkoff, D., K.A. Sharp, and B. Honig. 1994. Accurate calculation of hydration free energies using macroscopic solvent models. J. Phys. Chem. 98: 1978–1988.

29. Baker, N.A., D. Sept, S. Joseph, M.J. Holst, and J.A. McCammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. U.S.A. 98: 10037–10041.

30. Gorham Jr, R.D., C.A. Kieslich, and D. Morikis. 2011. Complement inhibition by Staphylococcus aureus: Electrostatics of C3d–EfbC and C3d–Ehp association. Cel. Mol. Bioeng. 5: 32–43.

31. Phillips, J., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, et al. 2005. Scalable molecular dynamics with NAMD. J. Comput. Chem. 26: 1781–1802.

32. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: Visual molecular dynamics. J. Mol. Graphics. 14: 33–38.

33. MacKerell, A.D., D. Bashford, M. Bellott, R.L. Dunbrack Jr, J.D. Evanseck, et al. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B. 102: 3586-3616.

34. Pettersen, E.F., T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, et al. 2004. UCSF Chimera: A visualization system for exploratory research and analysis. J. Comput. Chem. 25: 1605–1612.

35. Huber, G.A., and J.A. McCammon. 2010. Browndye: A software package for Brownian dynamics. Comp. Phys. Comm. 181: 1896–1905.

36. Grant, B.J., A.P.C. Rodrigues, K.M. ElSawy, J.A. McCammon, and L.S.D. Caves. 2006. Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics. 22: 2695–2696.

37. Gabdoulline, R.R., and R.C. Wade. 1997. Simulation of the diffusional association of barnase and barstar. Biophys. J. 72: 1917–1929.

38. Cuendet, M.A., and O. Michielin. 2008. Protein-protein interaction investigated by steered molecular dynamics: the TCR-pMHC complex. Biophys. J. 95: 3575–3590.

39. Randjelović, J., S. Erić, and V. Savić. 2013. Computational study and peptide inhibitors design for the CDK9 - cyclin T1 complex. J. Mol. Model. 19: 1711–1725.

CHAPTER 5: IDENTIFICATION OF LOW MOLECULAR MASS INHIBITORS OF C3D-LIGAND INTERACTIONS: A STEP TOWARD NEW COMPLEMENT-TARGETED THERAPEUTICS

5.1 Introduction

The thioester domain of complement component 3 (C3d) is a multifaceted protein involved in pathogen recognition and complement regulation. The convex basic region contains a reactive thioester, which allows its precursor C3b to covalently attach to pathogenic cell surfaces. The concave acidic region of C3d serves as a hub for a variety of ligands, including host complement regulator factor H (FH) and complement receptor 2 (CR2), as well as several secreted proteins from *S. aureus* (Efb, Ecb, and Sbi). C3d is of interest as a therapeutic target for several reasons. First, pathogens exploit C3d to evade complement. *Staphylococcus aureus* (*S. aureus*) expresses three homologous proteins (Efb, Ecb, and Sbi) that contain C3d-binding domains, which have all been shown to inhibit alternative pathway activation of complement, and to interfere with C3d binding to CR2 (see Chapter 1.3). Identifying molecules to block these interactions could aid in development of novel anti-infective therapeutics against strains of *S. aureus* resistant to traditional antibiotics (i.e. MRSA). In addition, inappropriate attachment of C3d to host proteins has been shown to lead to the production of autoantibodies, leading to development and progression of autoimmune diseases, including rheumatoid arthritis and systemic lupus erythmaeosus (see Chapter 1.2).

There are currently only two complement-targeted therapeutics on the market, approved for treatment of only a small percentage of complement-mediated diseases. Several others are currently in preclinical or clinical development (see Chapter 1.4). The vast majority are biologic-type therapeutics, which exhibit high specificity for their targets, and show high efficacy in early testing. Unfortunately, biopharmaceuticals face many challenges during development, including high production cost, poor oral bioavailability, immunogenicity, and proteolytic degradation (1). Low molecular mass inhibitors typically do not suffer from these challenges, but often lack target specificity and high potency. Furthermore, infections from emerging strains of *S. aureus* require new therapeutic approaches.

Recent advances in structural biology and computational chemistry have renewed interest in designing targeted small molecular therapeutics for a variety of applications. Conformational sampling methods and affinity prediction provide a foundation for the design of peptides and peptidomimetics, while virtual high-throughput screening can identify small organic molecules with strong binding affinity for specific biological targets. Due to limited computational resources and time, there is a trade off between speed and accuracy, often resulting in large numbers of false positives and false negatives during such screens. We describe herein an integration of two approaches to the design of low molecular mass complement therapeutics. First, we describe computational methodology for the structure-based rational design of peptides with predicted C3d binding affinity. Furthermore, we discuss an integration of molecular dynamics (MD) simulations, pharmacophore screening, and molecular docking to identify drug-like organic molecules that bind C3d. Together, these methods provide a foundation for the development of new anti-staphylococcal and complement-targeted therapeutics.

5.2 Methods

5.2.1 Preparation of protein structures

Initial crystallographic structures of C3d (PDB code 1C3D, (2)) and the C3d-Efb-C (PDB code 2GOX, (3)), C3d-Ehp (PDB code 2NOJ, (4)), and C3d-CR2 (PDB code 3OED, (5)) complexes were obtained from the protein data bank (PDB) (6). C3d-derived peptides were based on the cocrystal structure of the C3d-Efb-C complex. Efb- and Ecb-derived peptides were based on the cocrystal structures of C3d-Efb-C and C3d-Ehp, respectively. Proteins (from which the peptide was derived) were truncated to desired length, and in some cases, mutations were performed using Chimera (7).

5.2.2 MD simulations

Explicit solvent MD simulations were performed using NAMD (8). Missing heavy atoms and hydrogens were added, and molecules were parameterized using the CHARMM27 force field (9), in VMD (10). Structures were placed in TIP3P water cubes, with a minimum distance of 12 Å between any protein atom and the cube boundary, and sodium and chloride counterions were added to achieve 150 mM ionic strength and neutralization of protein charges. Each system was energy minimized using 25000 steps of conjugate gradient minimization, followed by heating and equilibration steps. After minimization, all protein atoms were harmonically constrained. A force constant of 10 kcal/mol/Å^2 was used during heating (62 ps) and the first equilibration stage (50 ps), followed by three equilibration stages (50 ps each) using force constants of 5, 2, and 1 kcal/mol/Å^2, respectively. In the final equilibration stage (50 ps), only backbone atoms were constrained (at 1 kcal/mol/Å^2). Equilibrated systems were simulated for 20 ns with no harmonic constraints. In all simulations, periodic boundary conditions and particle mesh Ewald electrostatics were employed, with nonbonded interaction cutoff and switching distance of 12 and 10 Å, respectively. Bonds involving hydrogen atoms were fixed using SHAKE, and 2 fs integration time steps were used. Langevin temperature and pressure controls were used for all NPT ensemble simulations. Resulting MD trajectories were analyzed using the Bio3D package (11) in R and UCSF Chimera (7).

5.2.3 Pharmacophore models derived from C3d-Efb-C interaction (Virtual Screen A)

Two peptide ligands (derived from the structure of Efb-C) were used as templates for deriving initial pharmacophore models. Ligands I (residues 127-131) and II (residues 134-139) were centered around Arg131 and Asn138, respectively, which were previously shown to be crucial to interaction with C3d (3, 12, 13). We used LigandScout 3.0.1 (14) for generation of structure-based pharmacophore models for Ligands I and II independently. PDB files containing C3d and either Ligand I or II were imported into LigandScout, and simplified pharmacophore models were generated with no more than one pharmacophore feature per heavy atom. Pharmacophore I contained five features, including two hydrogen bond acceptors, one hydrogen bond donor, one positive ionizable region, one hydrophobic feature, and ten exclusion volumes. Exclusion volumes allow the shape of the receptor binding pocket to be accounted for implicitly during pharmacophore screening. Pharmacophore II contained six features, including one hydrogen bond acceptor, two hydrogen bond donors, one positive ionizable region, two hydrophobic features, and eleven exclusion volumes. The drug-like purchasable subset of the ZINC database (~5.6 million compounds at time of initial screen) was used for screening. Catalyst 4.11 (Accelrys) was used to generate a conformer library for the database and subsequent screening of the conformer library against Pharmacophores I and II.

5.2.4 Dynamic pharmacophore models derived from C3d-CR2 interaction (Virtual Screen B)

The MD trajectory of C3d-CR2 was used to develop dynamic pharmacophore models. CR2 residues participating in persistent intermolecular salt bridges, hydrogen bonds, or nonpolar interactions (> 50% occupancy) were selected as tentative pharmacophore points. Relevant atoms of each selected residue were identified, and mean positions of centers of mass were calculated. A freestyle pharmacophore hypothesis was generated using Phase (Schrödinger, LLC), with 24 features corresponding to mean center-of-mass positions of specified CR2 residues. Feature types were defined based on the physicochemical properties of the CR2 atoms used to define them. The conformational flexibility of each feature was calculated from the MD trajectory, and used to define the tolerance radii of each corresponding feature of the pharmacophore model. Three hydrophobic/aromatic features were manually placed at the bottom of the C3d cavity as "anchoring points" for small molecules. From a total of 27 pharmacophore points, sets of 3-5 features were defined, using interfeature distances, diversity of feature types and experimental mutagenesis/binding/functional data as selection criteria. A total of 84 pharmacophore models were selected for screening. The source of screening molecules in this study was the drug-like purchasable subset (15) of the ZINC 12 database (16), consisting of ~7 million molecules at the time of writing. Conformer generation was performed using Phase, resulting in ~160 conformers per molecule on average. Conformers were aligned with each pharmacophore submodel, with exclusion volumes to prevent overlap with C3d residues.

5.2.5 Docking

All molecules identified in pharmacophore screens were docked to C3d using AutoDock Vina (17). Receptor and ligand structures were processed using AutoDockTools. For Virtual Screen A, pharmacophore hits were docked within a grid that was centered around their respective pharmacophore model on the C3d surface. In Virtual Screen B, we followed a protocol similar to the relaxed complex scheme (RCS) (18, 19), in which molecules were docked to a conformational ensemble of C3d. A total of 400 MD snapshots (200 from each trajectory) were clustered based on the all-atom RMSD of C3d residues involved in CR2 binding. The resulting dendrogram was separated into five clusters using the 'cutree' function in R, and representative snapshots from distinct clusters were selected, resulting in five C3d structures (four from the C3d-CR2 simulation and one from the free C3d simulation) with ≥ 1.75 Å RMSD from each other. Ligands were docked within a grid surrounding the entire CR2 binding site on C3d (measuring 40 Å x 40 Å x 40 Å) using an increased exhaustiveness parameter to enhance search accuracy within the enlarged grid. The top 20 docked poses of each molecule were returned after docking.

5.2.6 Scoring

All docked molecule poses were scored using the Vina scoring function. In Virtual Screen A, molecules were rescored using the Autodock4RAP scoring function, which makes use of AM1-BCC charge types for more accurate energy prediction in some cases (20). Molecules were ranked based on these scores, and 10 molecules from the top 20 docked molecules from each model were selected for experimental testing. For Virtual Screen B, predicted binding energies (from Vina) were reported for docking to each receptor conformation, and mean energies (across all receptor conformations) were used as a ranking metric. In addition, docked poses were rescreened against all pharmacophore models "in-place", to identify molecules that docked in positions that overlapped with two or more pharmacophore features. Docked poses were also subjected to RMSD-based clustering to identify molecules with low-energy poses that bind in similar conformations to all receptor conformations. A combination of all scoring methods was used for selection of 50 molecules for experimental testing.

5.2.7 Peptide and small molecule preparation

Peptides were synthesized with free N- and C-termini by Genscript. Stock solutions were prepared by dissolving peptides in deionized water to a final concentration of 2 mM, and were stored at -20°C until used. Selected drug-like compounds were obtained from Enamine, Pharmeks, ChemDiv, ChemBridge, and Maybridge. Stock solutions were prepared by dissolving molecules in DMSO to concentrations of 4 mM (with no visible precipitation), and stored at -20°C until used. In all assays, molecules were dissolved further in appropriate aqueous buffer, to minimize effects of DMSO, and DMSO (to an equivalent final concentration) was used in control samples.

5.2.8 Circular dichroism

Circular dichroism (CD) spectra were collected for each peptide using a Jasco J-815 CD Spectrophotometer. Measurements were performed at 25°C in a 1 cm path length cuvette, over a wavelength range of 190-260 nm. Peptides were measured at a final concentration of 40 µM in water, and C3d (20 µM) was used as a positive control for helicity.

5.2.9 C3d-Efb/Ecb binding ELISA

Peptides and top-ranked molecules were evaluated for competitive binding to C3d. Nunc Maxisorp 96-well ELISA plates were coated with either 0.5 µg Efb-C or Ecb (or peptide) for ~18 h at room temperature. Plates were washed three times with PBS containing 0.05% Tween-20 (PBS-T) between each step. Washed plates were blocked with 1% BSA/PBS-T for 1 h at 37°C. Serial dilutions of C3d were prepared (10-0.015 nM), in the presence or absence of each compound (200 µM). Samples were preincubated at room temperature for 15 mins. Preincubated samples were transferred to the washed plates, and incubated at 37°C for 1 h. After the binding reactions, plates were incubated with anti-C3d (Quidel; 1:1000 in 1% BSA/PBS-T), followed by incubation with horseradish peroxidase-conjugated anti-mouse (BioRad; 1:5000 in 1% BSA/PBS-T), both at 37°C for 1 h. Bound C3d was quantified using a 3,3',5,5'-tetramethylbenzidine substrate solution containing urea hydrogen peroxide in 0.11 M sodium acetate buffer, followed by a 1 N H_2SO_4 acid stop. Plates were measured spectrophotometrically at 450 nm (see Appendix G.1 for detailed protocol).

5.2.10 Microscale thermophoresis

Direct binding of molecules was assayed using a Monolith NT.115 (Nano Temper Technologies GMBH). Recombinant C3d (20 µM, plasmid kindly provided by Jean van den Elsen) was labeled with NT647 dye using the kit from Nano Temper Technologies. A dilution series of each molecule was prepared (800-0.39 µM) and preincubated with NT647-labeled C3d (200 nM) for 30 mins at room temperature in the dark. 4 µl of each dilution was loaded into hydrophobic capillary tubes, and measurements were obtained.

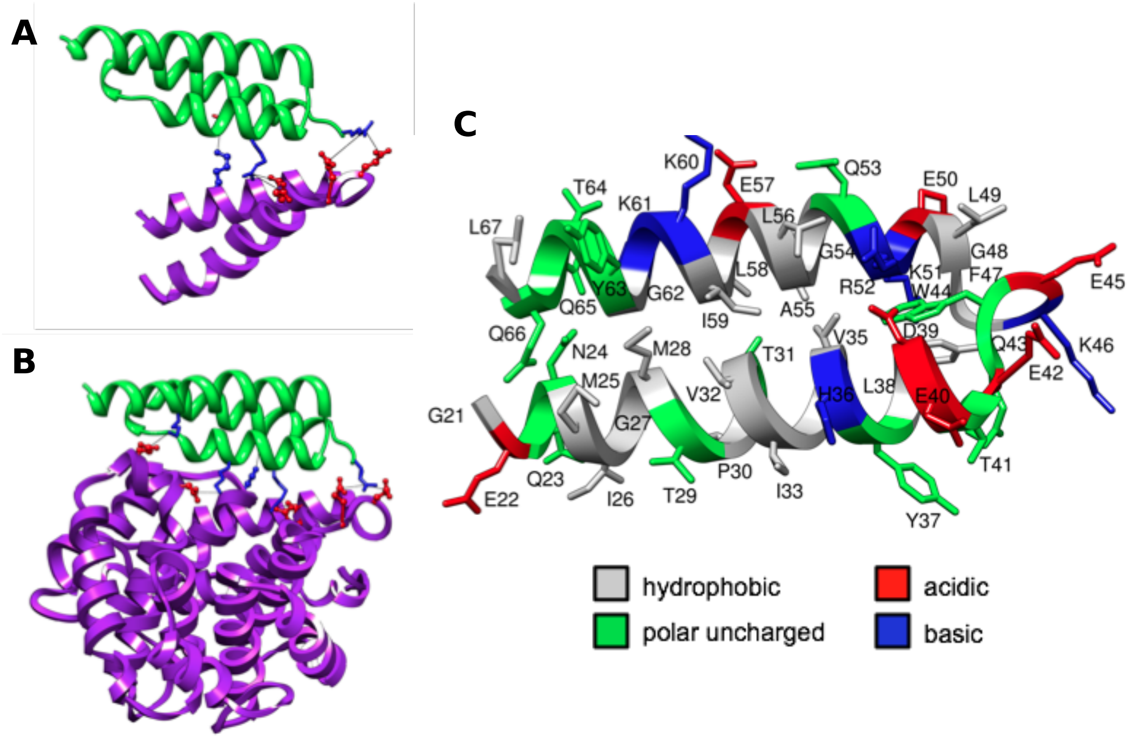

A
B
C
K60
Q53
E50
T64
K61
E57
L49
L56
L67
G54
G48
L58
K51
F47
E45
Y63
G62
R52
W44
Q65
I59
A55
D39
Q66
V35
Q43
K46
N24
M28
T31
E42
M25
V32
L38
G21
H36
E40
G27
Q23
P30
T41
E22
I26
T29
I33
Y37
hydrophobic
acidic
polar uncharged
basic

Figure 5.1 Molecular graphics representations of the C3d-derived peptide (C3dp) template. Efb-C (green ribbon) is shown in complex with (A) C3dp and (B) full-length C3d. The location from which C3dp was derived is evident here. (C) A ribbon diagram of C3dp and the physicochemical properties of its residues is shown.

Table 5.1 C3d-derived peptide sequences.

Peptide	Sequence[a]
C3dpv1	GEQNMIGMTPTVIAVHYLDETEQWEKFGLEKRQGALELIKKGYTQQL
C3dpv2	-------MTATVIAVHYLDETEQWEKFGLEKRQGALELIKKGYTQQL
C3dpv3	-------MTATVIAVHYLDETEQWEKFGLEKRQGALELIKKGY----
C3dpv4	-------ETATVIAVHYLDETEQWEKFGLEKRQGALELIKKGYTQR-
C3dpv5	-------DTATVIAVHYLDETEQWEKFGLEKRQGALELIKKGYTQR-
C3dpv7	-------ETATVEAVHYLDETEQWEKFGLEKRQGALEEIKKGYTQR-
C3dpv8	------------------EETEQWEKFGLEKRQ--------------
C3dpv9	-------ETATVEAVHYLDETEQWEKFGEEKRQGAEEEIKKGYTQR-

[a] Sequences are colored by residue type; nonpolar (black), polar uncharged (green), positive (blue), and negative (red).

Table 5.2 Efb- and Ecb-derived peptide sequences.

Peptide	Sequence[a]
Efbp1	VSAHRKAQKAVNL
Efbp2	-SAHRKAQKAVN-
Ecbp1	VATHRKAQRAVNI
Ecbp2	-ATHRKAQRAVN-

[a] Sequences are colored by residue type; nonpolar (black), polar uncharged (green), positive (blue), and negative (red).

5.3 Results and Discussion

5.3.1 Peptide design

We sought to design peptides to inhibit the interaction between C3d and Efb-C/Ecb as potential staphylococcal therapeutics. Our initial efforts were aimed at designing C3d-derived peptides. We isolated a peptide fragment of C3d (C3dp), which includes several ionizable residues involved in electrostatic interactions with Efb-C and Ecb. An advantage of C3d-derived peptides is the low likelihood of immunogenicity. The initial template for C3dp was comprised of a 47-residue helix-loop-helix motif (residues 18-66) from C3d (Figure 5.1). Examination of C3dp within the native C3d structure indicated that most of the peptide was exposed to solvent, with the exception of the N-terminus of the peptide. In order to evaluate structural stability and binding of C3dp, we performed MD simulations of C3dp alone (free) and in complex with Efb-C (bound). Visual inspection of structures from MD simulations showed a generally stable structure, with high flexibility at the N-terminus. Quantitative secondary structure analysis confirmed this observation. Moreover, the bound simulation illustrated strengthening of intermolecular salt bridges. We generated difference interaction occupancy plots, which show the percentage of snapshots in which particular C3dp-Efb-C interactions are present relative to the wild-type C3d-Efb-C complex. These plots present a quantitative analysis of peptide binding compared to the native protein-protein interaction. Combination of secondary structure and binding analyses yielded important information for the selection of optimized peptides for experimental testing.

Several rounds of sequence optimization and dynamics simulations led to a 39-residue template peptide (C3dpv4) predicted to exhibit structural stability and tight binding to bacterial ligands, however it was poorly soluble and prone to aggregation in both water and DMSO, indicating that the hydrophobic character of this peptide would not be amenable to therapeutic development (Figure 5.2). We attempted to further optimize the sequence of C3dp (Table 5.1), by introducing acidic residues and making terminal truncations to increase solubility, while maintaining strong interactions with positively-charged bacterial ligands. Computational results did not yield promising candidates, as larger deviation from the native structure introduced structural perturbations and altered binding to the targets (Figure 5.3).

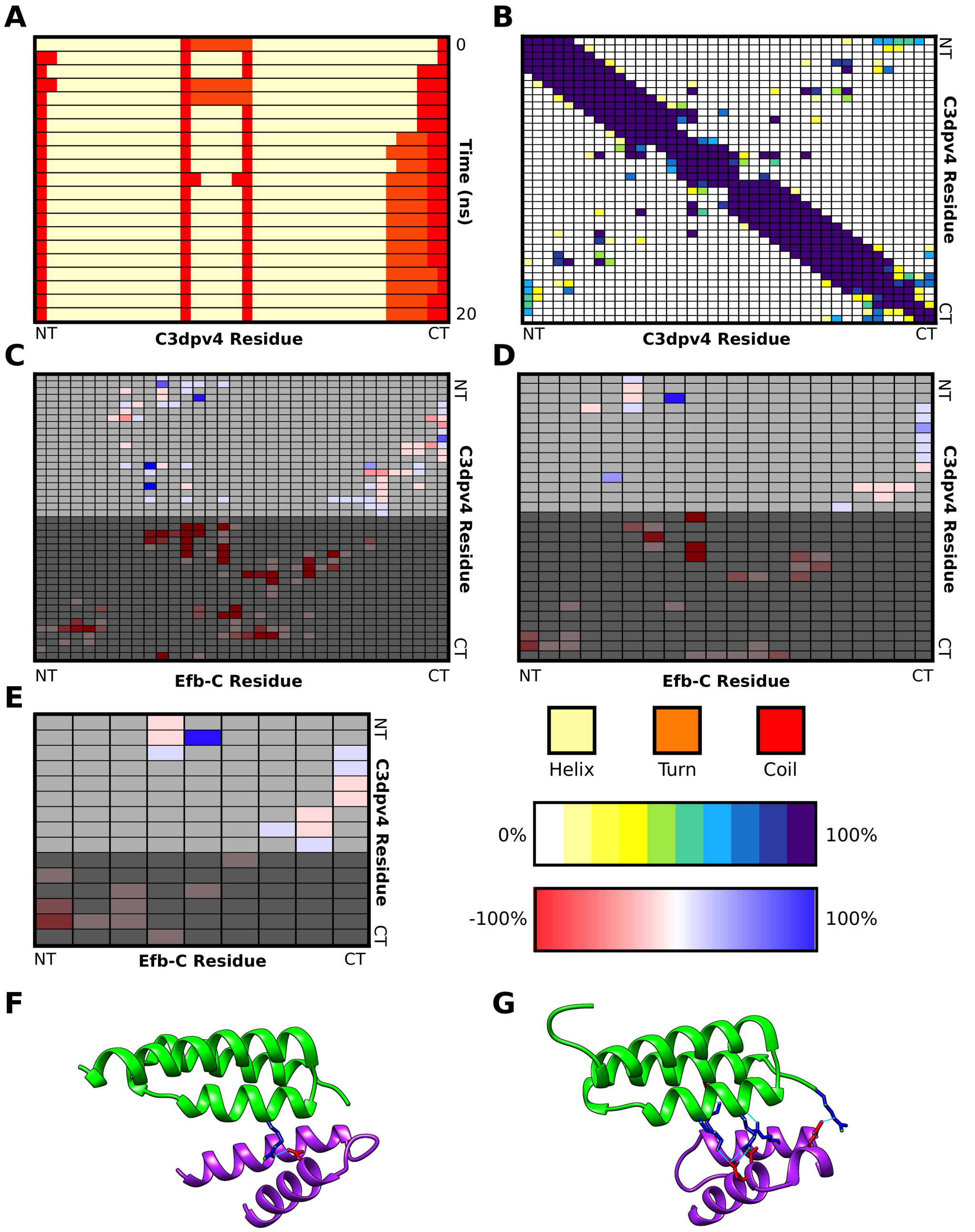


Figure 5.2 Quantitative comparison of C3dpv4 structural stability and binding. Analysis of secondary structure (A) and intramolecular residue-residue contacts (B) illustrate the stability of helices and the helix-loop-helix motif, respectively. Difference interaction contact maps are calculated by determining the percentage of time (during 20 ns MD trajectories) in which particular intermolecular interactions are present between C3dpv4 and Efb-C, and subsequently subtracting the occupancy of the same reaction pair from the wild-type C3d-Efb-C simulation. Interactions in red are weakened in the C3dpv4-Efb-C complex, and those in blue are strengthened. The plots show nonpolar interactions (C), hydrogen bonds (D), and salt bridges (E). The region shaded in darker gray represents residues from C3d that are not present in the peptide, such that interactions involving those residues are automatically lost. The legends and scale bars show colors corresponding to secondary structure elements for (A), interaction occupancy for (B), and difference interaction occupancy (C-E). The molecular graphics in (F) and (G) show the conformations of C3dpv4-Efb-C at the beginning and end of the 20 ns simulation, respectively.

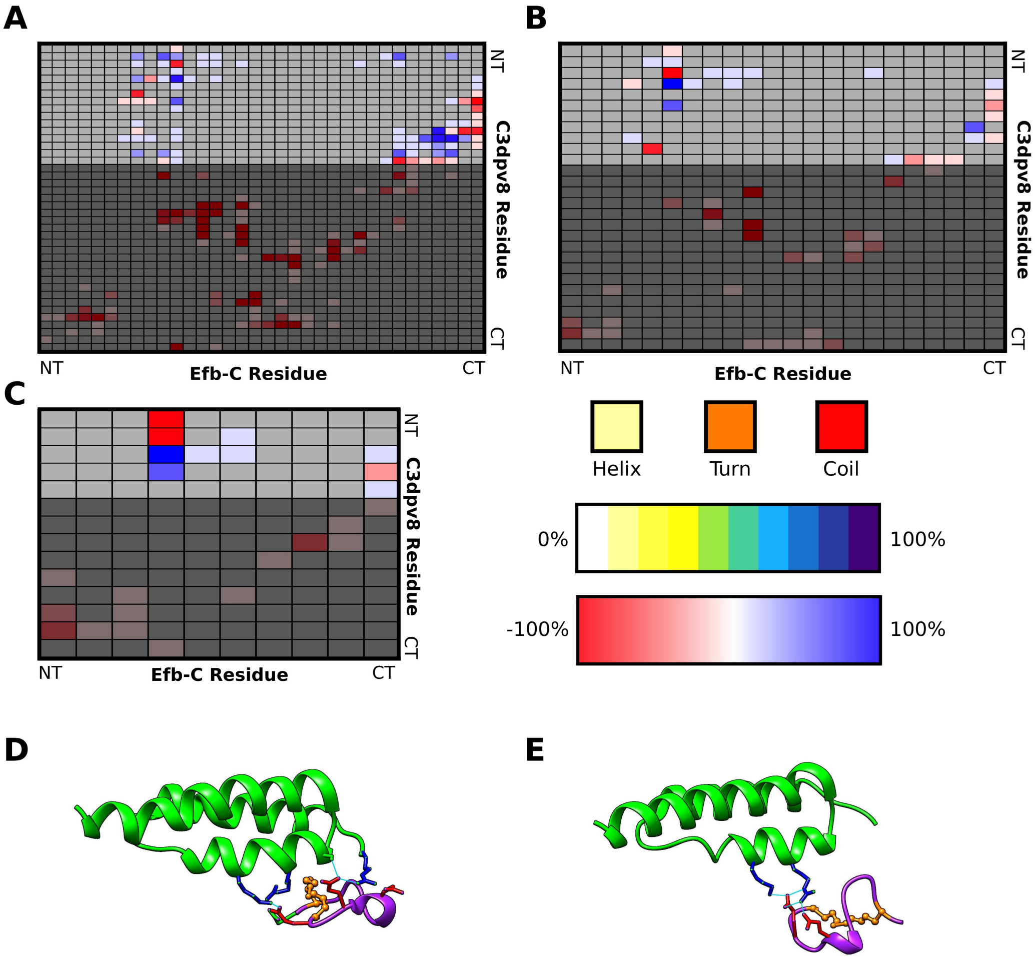


Figure 5.3 Quantitative comparison of C3dpv8 binding. Difference interaction contact maps are calculated by determining the percentage of time (during 20 ns MD trajectories) in which particular intermolecular interactions are present between C3dpv8 and Efb-C, and subsequently subtracting the occupancy of the same reaction pair from the wild-type C3d-Efb-C simulation. Interactions in red are weakened in the C3dpv8-Efb-C complex, and those in blue are strengthened. The plots show nonpolar interactions (A), hydrogen bonds (B), and salt bridges (C). The scale bar shows colors corresponding to difference interaction occupancy (A-C). The molecular graphics in (D) and (E) show the conformations of C3dpv8-Efb-C at the beginning and end of the 20 ns simulation, respectively.

We also designed peptides based on Efb-C and Ecb, which may competitively bind C3d and prevent bacterial ligand binding (Table 5.2). This approach holds the advantage that most of the interactions of Efb-C/Ecb/Sbi with C3d are confined to a single helix, thus binding is more likely to be maintained (Figure 5.4). We performed MD simulations with the second helix of Efb-C and Ecb, alone and in complex with C3d. Simulations showed that Efb-C and Ecb helices (which share nearly identical sequence) are highly stable, and maintain interactions with C3d. Efb and Ecb peptides were synthesized and tested for structure, binding, and activity. CD spectra, however, showed no evidence of helicity of

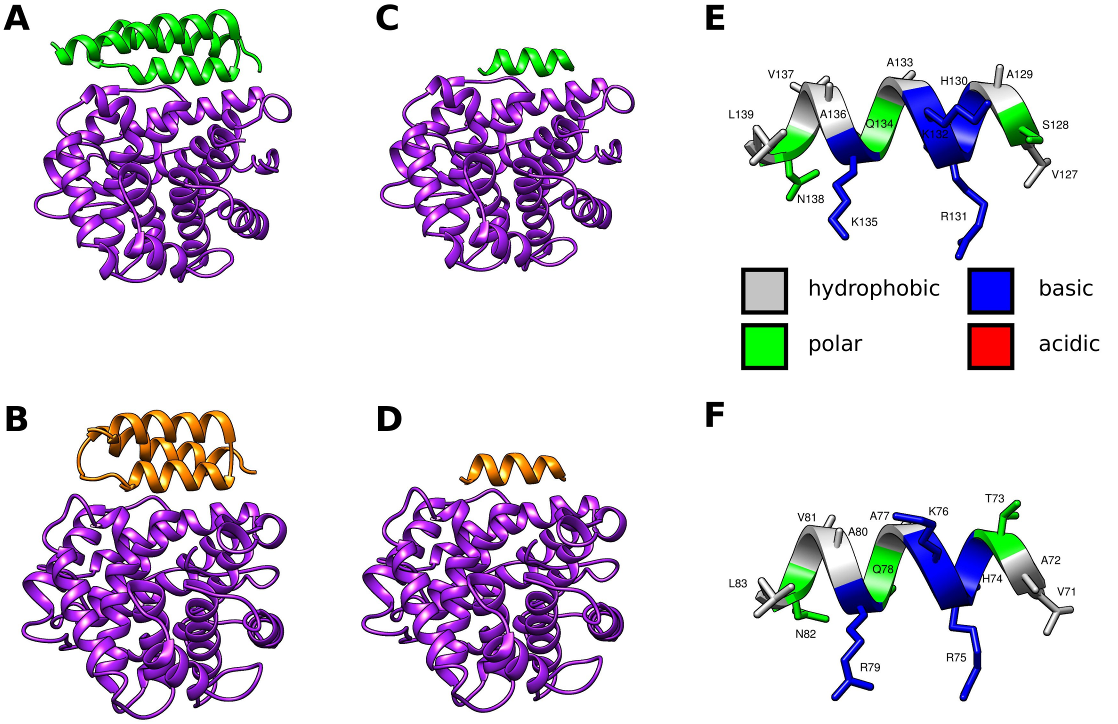


Figure 5.4 Molecular graphics representations of the Efb- and Ecb-derived peptides. C3d (purple ribbon) is shown in complex with full-length Efb-C (green; A), full-length Ecb (orange; B), Efb peptide (green; C), and Ecb peptide (orange; D). The physicochemical properties of the residues of Efb- and Ecb-derived peptides are shown in (E) and (F), respectively.

peptides (Figure 5.5A), as predicted in free peptide MD simulations. ELISA-based assays showed no direct binding of peptides to either C3b or C3d, and no competitive inhibition of bacterial ligand binding (Figure 5.5B-C).

These results are suggestive that our peptide design approaches require some improvements. First, our free MD simulations of peptides seem somewhat inadequate at predicting the equilibrium secondary structure of peptides. Explicit solvent MD cannot sufficiently explore unfolding of secondary structure elements in 20 ns. To efficiently explore peptide conformations, it may be necessary to use implicit solvent, replica exchange, or simulated annealing-type MD simulation methods. Regarding peptide binding, while we observed compensatory interactions among the residues shared among the peptide and wild-type protein, the fact that so many important interactions are lost when protein sequences are truncated makes lack of strong binding unsurprising (Figure 5.2C-E and 5.3A-C). Furthermore, cyclization methods (i.e. disulfide bridges, stapling) may aid in peptide structural stability (21, 22).

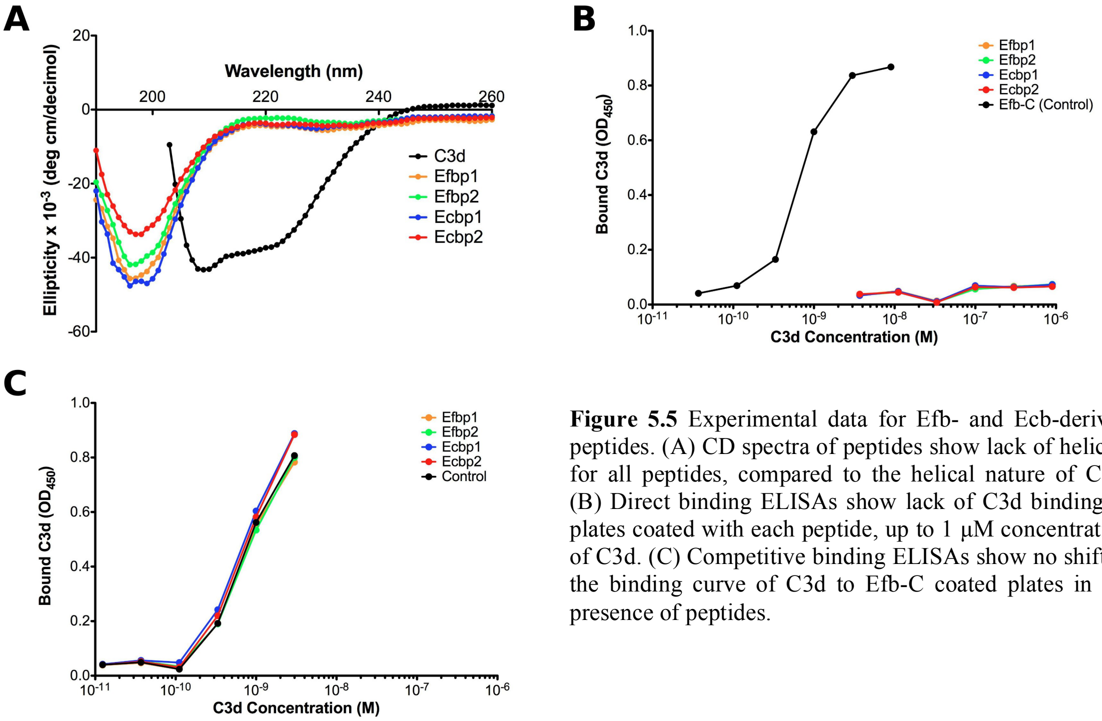


Figure 5.5 Experimental data for Efb- and Ecb-derived peptides. (A) CD spectra of peptides show lack of helicity for all peptides, compared to the helical nature of C3d. (B) Direct binding ELISAs show lack of C3d binding to plates coated with each peptide, up to 1 μM concentration of C3d. (C) Competitive binding ELISAs show no shift in the binding curve of C3d to Efb-C coated plates in the presence of peptides.

5.3.1 Virtual Screen A

We subsequently performed virtual screening based on the crystallographic structure of the C3d-Efb-C complex, in order to identify small drug-like organic molecules to inhibit this interaction. In this initial screen, we isolated two helical fragments of Efb-C (Figure 5.6A), and used them as templates to derive pharmacophore models (Figure 5.6B-C). The helical fragments consisted of residues of Efb-C (R131 and N138) previously determined to be critical for binding both experimentally and computationally. Conformers of molecules from the drug-like purchasable subset of the ZINC database were screened against the two pharmacophore models, resulting in ~30000 hits for Model I, and 77 hits for Model II. Each set of molecules was docked independently to a localized binding region on C3d surrounding the respective pharmacophore model (Figure 5.6D-E), and molecules were ranked according to their predicted binding affinities.

From this initial screen, many molecules were predicted to bind C3d with high affinity. We selected 10 molecules (7 from Model I docked hits, 3 from Model II docked hits) to test experimentally for

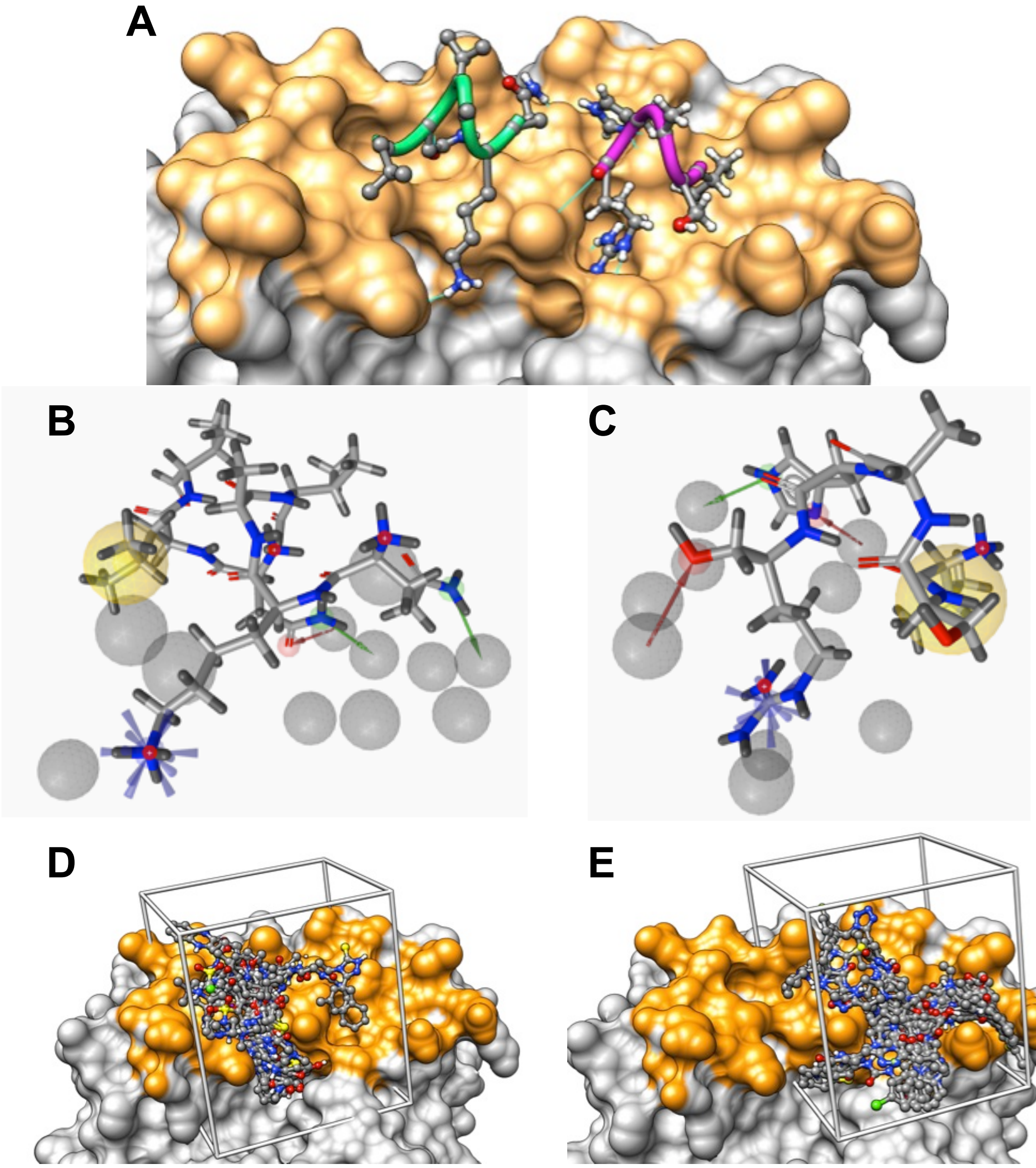


Figure 5.6 Molecular graphics and details of Virtual Screen A (pharmacophores and docking). Residues of C3d less than 5 Å from Efb-C are shown in orange. Two selected Efb-derived peptides are shown in magenta (A) and green (B). Pharmacophore models are shown based on Ligands II (C) and I (D), containing hydrogen bond donors (green arrows) and acceptors (red arrows), positive ionizable regions (blue starburst), hydrophobic features (yellow spheres), and excluded volumes (gray spheres). Top compounds docked within the Model II (D) and Model I (E) sites, with the corresponding grid within molecules were conformationally sampled.

binding. The properties and structures of these molecules are shown in Table 5.3 and Figure E.1, respectively. While none of the molecules exhibited competitive binding with Efb-C in binding ELISA experiments (Figure E.2), one molecule (Molecule A1) was identified to bind to C3d using microscale thermophoresis (Figure 5.7). This result motivated a more rigorous virtual screen, in which we incorporated

Table 5.3 Molecules selected from Virtual Screen A for experimental testing.

No	ZINC ID	Energy	Charge	MW	Vina Rank	Autodock4^{RAP} Rank	Docking Site
A1	ZINC24528440	-8.92	1	498.7	4558	2	I
A2	ZINC08753517	-8.84	1	452.9	1096	6	I
A3	ZINC36548357	-8.66	0	468.6	330	9	I
A4	ZINC14124661	-8.48	1	456.6	102	10	I
A5	ZINC09589829	-8.47	1	441.6	322	11	I
A6	ZINC58054534	-8.47	1	466.5	271	12	I
A7	ZINC15070155	-8.41	2	499.7	260	13	I
A8	ZINC12245128	-7.88	1	420.5	2	4	II
A9	ZINC19897404	-7.85	0	456.6	3	5	II
A10	ZINC12545339	-7.03	1	396.5	35	13	II

The table shows ZINC IDs of selected molecules from Virtual Screen A, predicted (recalculated) binding energies to a single C3d structure, charge, molecular weight (MW), and rankings. Vina rank is based on the energy reported by the default Vina scoring function, and Autodock4^{RAP} rank is based on recalculated energies. The docking site refers to whether the molecules came from docking to site I or II.

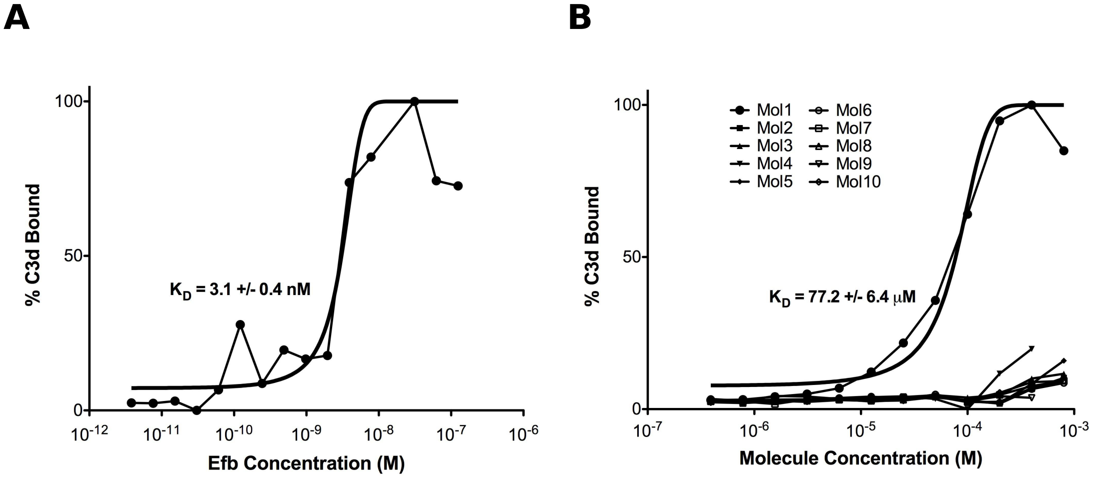


Figure 5.7 Microscale thermophoresis experiments for direct molecule binding to C3d. Ten top selected molecules were tested for direct binding to C3d. (A) Efb binds to C3d with strong affinity, while (B) Mol3 binds C3d with weaker affinity. All other molecules did not bind. Mol1-10 correspond to A1-A10 in Table 5.3.

ligand and receptor flexibility, as well as additional postprocessing methods, to identify additional compounds with predicted C3d binding capability.

5.3.1 Virtual Screen B

Following the success of the initial screen, we made significant optimizations to our screening protocol in order to identify an enriched set of C3d-binding molecules. In Virtual Screen A, we focused on identifying inhibitors of the C3d-Efb-C interaction. Since this interaction is of high affinity (K_d ~ 1 nM), we turned our attention to the C3d-CR2 interaction, which is significantly weaker (K_d ~ 500 nM) and more likely inhibited by small drug-like molecules. Inhibition of C3d-CR2 is of particular interest in the

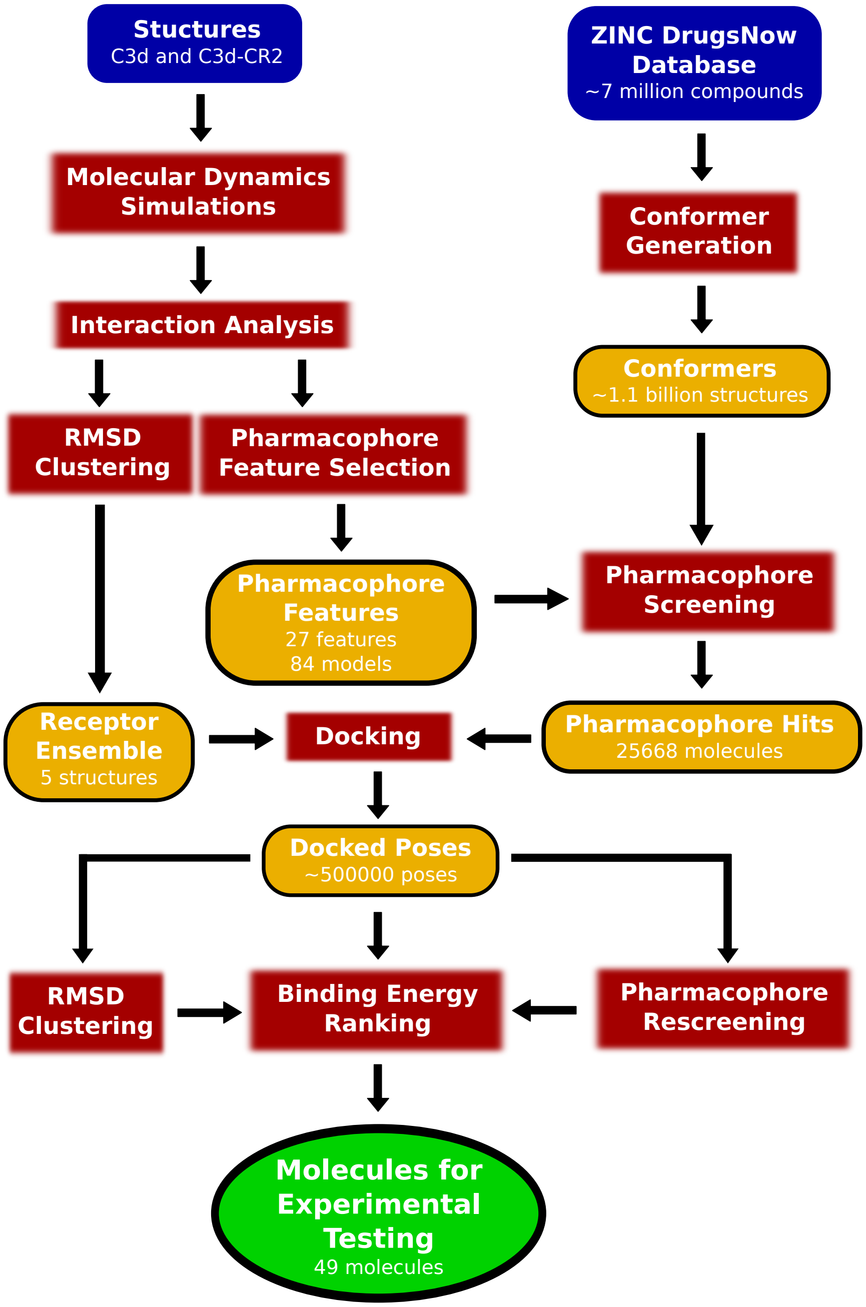

Stuctures
C3d and C3d-CR2
ZINC DrugsNow Database
~7 million compounds
Molecular Dynamics Simulations
Conformer Generation
Interaction Analysis
Conformers
~1.1 billion structures
RMSD Clustering
Pharmacophore Feature Selection
Pharmacophore Features
27 features
84 models
Pharmacophore Screening
Receptor Ensemble
5 structures
Docking
Pharmacophore Hits
25668 molecules
Docked Poses
~500000 poses
RMSD Clustering
Binding Energy Ranking
Pharmacophore Rescreening
Molecules for Experimental Testing
49 molecules

Figure 5.8 Overview of improved computational protocol for Virtual Screen B. Required inputs are shown in blue boxes, computational methods in red, intermediate outputs in yellow, and the final selected molecules in green (for experimental testing.

identification of complement therapeutics, and the presence of a deep cavity on C3d in the center of the CR2 binding site improves amenability to small molecule binding. Furthermore, since the CR2 binding site

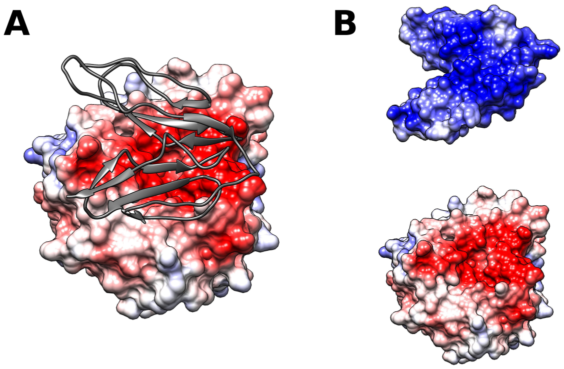


Figure 5.9 Electrostatically-driven association of C3d and CR2. (A) CR2 (gray ribbon) binds over the acidic concave region on the surface of C3d. (B) An openbook representation (rotation about horizontal axis) of C3d and CR2 surfaces show the complementarity between positively charged CR2 (top) and negatively charged C3d (bottom). Surfaces are colored by electrostatic potential using isovalues of +5 k_bT/e (blue) and -5 k_bT/e (red).

overlaps significantly with Efb-C (and Ecb and Sbi) (Figure E.3), our new virtual screening protocol serves to potentially identify staphylococcal inhibitors as well.

We employed a multi-step protocol for the identification of C3d-CR2 inhibitors (Figure 5.8). First, we performed MD simulations of free C3d and the C3d-CR2 complex, and analyzed interactions and conformational flexibility in each trajectory. Interaction data, in conjunction with available experimental mutagenesis data for C3d and CR2, were used to define pharmacophore features and derive sets of pharmacophore models for the initial compound screen. We generated conformers of ZINC compounds in order to account for compound flexibility, and screened these conformers against 84 pharmacophore models, yielding ~25000 unique compounds that matched three or more pharmacophore features in at least one model. Each of the pharmacophore hits were docked to five distinct conformations of C3d, derived from RMSD based clustering of binding site residues from MD trajectory snapshots of C3d, both free and in complex with CR2. Lowest energy binding poses for each molecule (and their corresponding binding energies) were reported. Since molecules were free to dock anywhere within the acidic concave region of C3d, many binding poses varied considerably from one another. We used RMSD clustering of poses to identify compounds that docked in a similar conformation to all receptor conformations. Furthermore, we re-screened docked poses against pharmacophore models, to identify compounds that exhibited low energy binding and desired physicochemical property distribution simultaneously.

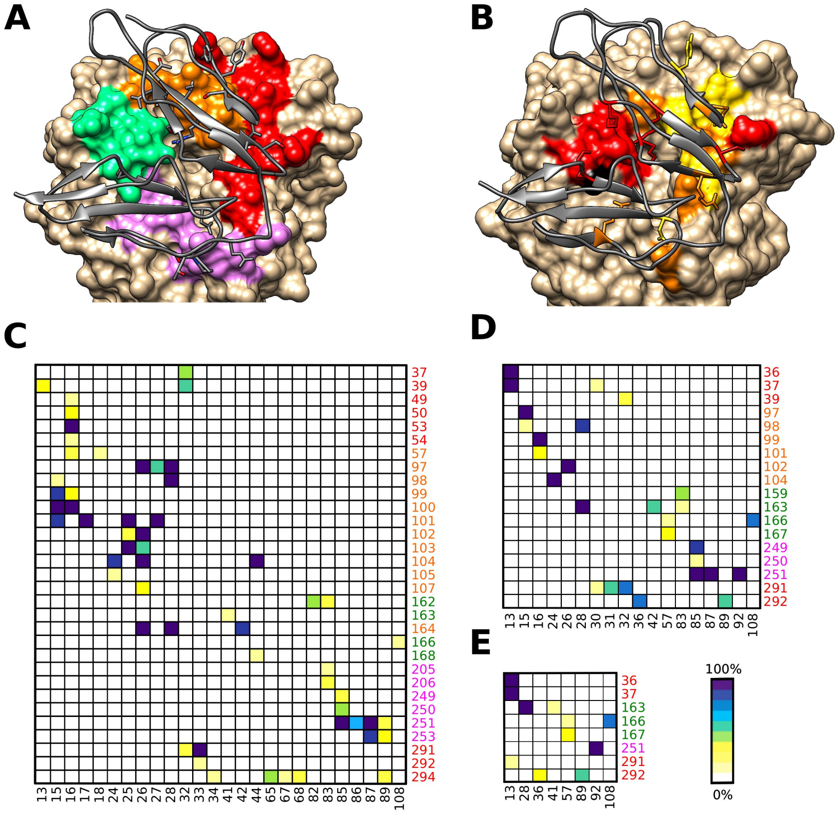


Figure 5.10 Analysis of interactions between C3d and CR2. (A) Examination of intermolecular interactions from the C3d-CR2 crystallographic structure (PDB code 3OED) reveals a ring of interactions surrounding a deep cavity on the C3d surface. The figure shows interacting regions of C3d, divided into four sectors; acidic cluster 1 (red), nonpolar cluster (orange), acidic cluster 2 (green), and polar cluster (purple). (B) The C3d surface and CR2 side chains are colored by their effect on experimental binding affinity compared to the wild-type C3d-CR2 complex; < 25% binding (red), 25-59% binding (orange), 60-84% binding (yellow), and > 125% binding (black). C3d-CR2 interaction occupancies were calculated throughout a 20 ns MD trajectory, and nonpolar interactions (C), hydrogen bonds (D), and salt bridges (E) are shown here. CR2 residues are on the x-axes, and C3d residues on the y-axes. C3d residue labels are colored according to their corresponding sector (from Panel A).

5.3.2 Dynamics of C3d-CR2 interaction

We performed MD simulations of the C3d-CR2 complex, in order to evaluate the conformational flexibility of the C3d-CR2 binding site, and to elucidate which residues are crucial to the C3d-CR2 interaction. Similar to most known ligands of C3d, CR2 binds at the concave acidic region of C3d (Figure 5.9A), and the interaction is dominated by electrostatic interactions (Figure 5.9B). Through initial examination, a deep cavity is observed in the concave acidic region of C3d, and this cavity persists during MD simulations of both free C3d and the C3d-CR2 complex. Interestingly, CR2 itself does not "enter" the cavity, but rather resides atop the cavity, forming polar and nonpolar interactions with a ring of C3d

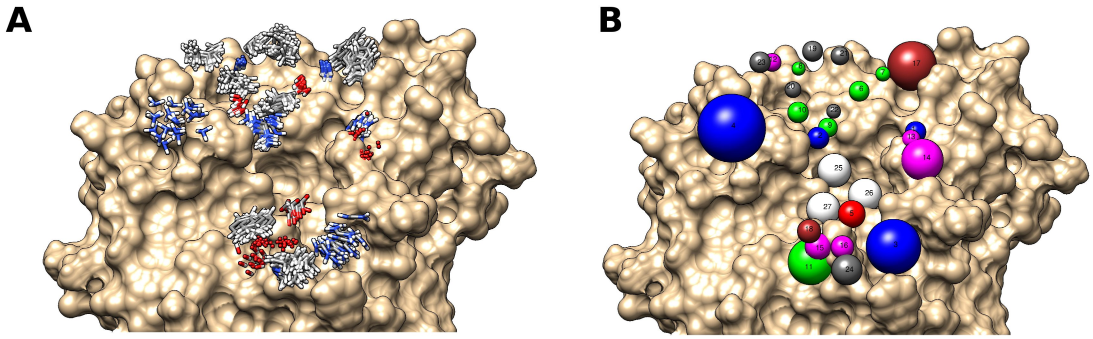


Figure 5.11 Development of a dynamic pharmacophore model for C3d-CR2. (A) Positions of CR2 residues and groups interacting with C3d, at 10 different time points during the MD trajectory. Atoms are colored by element type (carbon, gray; nitrogen, blue; oxygen, red; hydrogen, white). The physicochemical properties and spatial distribution of the atom groups were used to define the 27 pharmacophore features observed in (B). Features are colored by pharmacophore type (hydrogen bond donor, green; hydrogen bond acceptor, magenta; positive charge, blue; negative charge, red; aromatic, brown; and hydrophobic, gray).

residues lining the perimeter of the cavity (Figure 5.10A). The C3d interaction ring can be divided into four sectors, based on physicochemical interaction type and spatial distribution. Sectors 1 (red) and 3 (green) consist primarily of electrostatic interactions, while sectors 2 (orange) and 4 (purple) are involved in hydrogen bonding and nonpolar interactions. Residues in sectors 1-3 are also crucial for interactions between C3d and staphylococcal virulence factors Efb-C, Ecb, and Sbi-IV, all of which have been shown (both structurally and functionally) to competitively inhibit C3d-CR2 interaction (Figure E.3).

The interaction between C3d and CR2 has been well studied in the last 15 years (2, 5, 23-33). A large amount of mutagenesis data is available, which can facilitate our vHTS efforts. Mutations of C3d and CR2, and their relative effects on binding, are shown in Figure 5.10B. Naturally, the locations of inhibitory mutations (red, orange, and yellow) correspond to the ring of interactions seen in Figure 5.9A. Whereas inhibitory mutations are spatially disbursed, we noticed that the most inhibitory mutations are observed in sectors 1 and 3, and involve primarily mutations of charged residues. Therefore, while all persistent interactions from MD simulations were considered during vHTS of C3d (Figure 5.10C-E), charge-charge interactions (Figure 5.10E) were given special consideration in development of pharmacophore models and postprocessing/re-ranking of lead compounds after docking.

Table 5.4 Properties of dynamic pharmacophore features for C3d-CR2.

Feature	Type[1]	Residue[2]	Residue Atoms[3]	C3d Residue Partners	Tolerance (Å)[4]
1	P(D)	R13	CZ,NH1,NH2	D36,E37	1.94
2	P(D)	R28	CZ,NH1,NH2	D163	1.57
3	P(D)	R89	CZ,NH1,NH2	D292	3.93
4	P(D)	K108	NZ	E166	2.71
5	N(A)	D92	CG,OD1,OD2	K251	2.02
6	D	S15	OG	V97,N98	0.59
7	D	Y16	N	L99,A101	0.74
8	D	V26	N	I102	0.71
9	D	R28	NE	N98	0.91
10	D	S42	OG	D163	1.05
11	D	S85	OG	Q249,L250,K251	1.42
12	A	G24	O	S104	1.43
13	A	C31	O	K291	1.29
14	A	S32	O	E39,K291	2.36
15	A	S85	O	Q249,L250,K251	1.15
16	A	P87	O	K251	1.00
17	R	Y16	CG,CD1,CE1,CZ,CD2,CE2	R49,Q50,L53,E54,K57,L99,I100	4.03
18	R	Y88	CG,CD1,CE1,CZ,CD2,CE2	---	4.03
19	H	T25	CG2	A101,I102,D103	1.78
20	H	V26	CG1,CG2	V97,I102,D103,S104,L107,I164	1.78
21	H	I27	CB,CG2,CG1,CD	V97,A101	1.80
22	H	R28	CB,CG,CD	V97,N98,I164	2.68
23	H	L44	CD1,CD2	S104,Q168	1.59
24	H	P87	N,CD,CA,CB,CG	K251,F253	2.62
25	R	---	---	S94,L95,N98,E160	2.00
26	R	---	---	H33,Q284,Q288	2.00
27	R	---	---	K91,E160,Y201,Q205	2.00

[1] Feature Types: D (hydrogen-bond donor), A (hydrogen-bond acceptor), P (positive), N (negative), H (hydrophobic), R (aromatic).

[2] Residue types and numbers correspond to CR2 residues and numbering (from PDB code 3OED).

[3] Atoms from residues used to derive features (also used associated hydrogen atoms - not listed).

[4] Tolerance reflects size of pharmacophore feature used for molecule screening.

5.3.3 Pharmacophore screening

The first step in our vHTS protocol involved searching for molecules with similar geometric and spatial arrangement of physicochemical features to CR2. Based on the MD simulation of C3d-CR2, we selected several CR2 residues important for its interaction with C3d. The arrangement of selected functional moieties (isolated from CR2 residues) is shown in Figure 5.11A, with respect to C3d. These moieties were used to define 24 pharmacophore features of varying types (i.e. positive, negative, hydrogen bond acceptor or donor, aromatic, hydrophobic), based on the physicochemical properties of each moiety itself (Table 5.4 and Figure 5.11B). Since these atoms fluctuated to varying degrees over time, tolerance spheres were generated with radii large enough to incorporate 95% of their occupied positions. In addition, three "anchoring" features (Figure 5.11B, features 25-27) were defined in the bottom of the C3d cavity.

We selected pharmacophore models from subsets of the 27 features, in order to optimize the number of hits, specificity, and screen time. The first parameter in model selection was the number of pharmacophore features. Using too few features often yields large numbers of false positives in pharmacophore screens, whereas using too many features may be too stringent during the search, yielding few or no hits at all. Selected pharmacophore models contained between three and five features, achieving a balance between specificity and diversity in molecule hits. Since the interfeature distances in our initial model are quite variable, we also applied distance criteria in submodel selection. Selected models had interfeature distances of at least 4 Å and less than 15 Å. These criteria alleviate fortuitous matching of multiple features by a single chemical moiety, while constraining the search to distances that can reasonably be spanned by a drug-like small molecule (MW < 500). Finally, we restricted model selection based on properties of specific features. Submodels were selected to contain one hydrophobic feature from the cavity base (Features 25, 26, or 27), one positive feature (Features 1-4), one feature representing a crucial interaction based on experimental binding data (Features 1, 2, 5, 6, 9-11, 13-16, 18, 20, 22, or 24), and no more than one additional hydrophobic feature. Submodels including sector 2 or sector 4 features were exempted from necessarily including base features (Features 25, 26, or 27). These criteria result in physicochemically diverse submodels that explore multiple regions of the C3d-CR2 interaction site.

In total, 84 pharmacophore models were selected, and generated conformers for ZINC compounds were screened against each model. We found that three-feature models were highly nonspecific and yielded very large numbers of hits. Models containing features from sector 4 also yielded large numbers of hits. Since sector 4 is less important than other sectors for C3d-ligand interactions, we focused on models comprised of features from sectors 1-3, with either four or five features. We narrowed the search to 51 models (27 of which yielded at least 1 hit), resulting in a total of 25668 unique compounds.

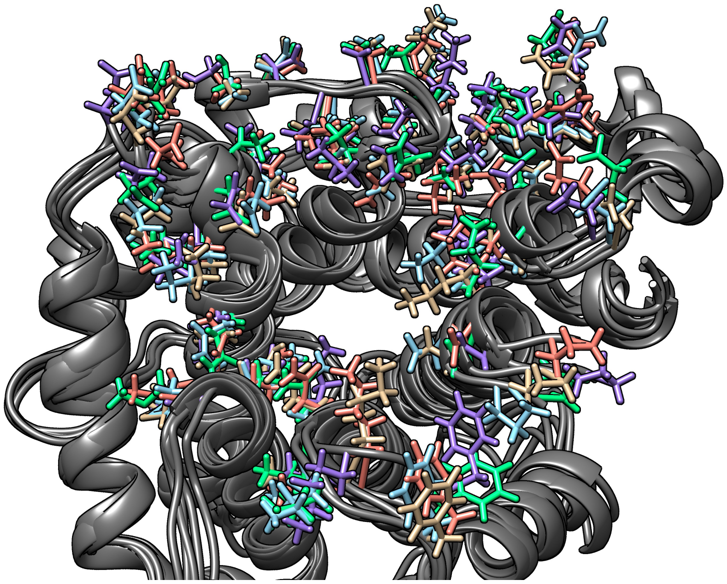


Figure 5.12 C3d receptor ensemble from MD simulation and RMSD-based clustering. Residue side chains used for RMSD clustering criteria are shown as sticks, and the different colors correspond to conformations of the side chains in five different receptor structures.

5.3.4 Docking and postprocessing

We used an RCS-type docking procedure, to incorporate receptor flexibility into docking simulations. From MD simulations of free C3d and the C3d-CR2 complex, we clustered C3d structures based on RMSD of interacting residues, as determined from the C3d-CR2 MD simulation (Figures 5.10C-E). This method allowed for selection of C3d structures with the most diverse CR2 binding modes. We obtained five structures, with RMSD for binding site residues between 2 and 3 Å for all structure pairs. The low pairwise RMSD values indicate robust structure within the binding site. Figure 5.12 shows the superposition of the five selected C3d structures from RMSD clustering, which shows similar positions for side chains of CR2-interacting residues, regardless of whether C3d is free or complexed.

Since we are exploring a large protein-protein interface as the molecular target in this vHTS project, it is of interest to examine where compounds dock within the entire binding site. Thus, during docking simulations, we allowed compounds to dock within a large (40 × 40 × 40 Å) grid surrounding the acidic concave region of C3d. Molecules were docked to the five C3d conformations mentioned above, and predicted binding energies were reported for low-energy docked poses, based on the AutoDock Vina scoring function.

Table 5.5 Molecules selected from Virtual Screen B for experimental testing.

No	ZINC ID	R1	R2	R3	R4	R5	Mean	Charge	MW	RMSD Rank	Rank
B1	ZINC02520169	-9.7	-8.2	-8.9	-8.8	-9.1	-8.94	0	500.0	3	1
B2	ZINC15962406	-9.6	-8.4	-8.1	-8.3	-9.4	-8.76	0	448.5	16	4
B3	ZINC09980865	-9.1	-8.2	-9.0	-8.5	-8.8	-8.72	0	491.6	1	7
B4	ZINC11933334	-9.6	-8.2	-8.6	-8.4	-8.8	-8.72	0	427.5	28	8
B5	ZINC14749391	-9.1	-8.4	-8.4	-8.5	-9.0	-8.68	1	499.7	7	11
B6	ZINC08826044	-9.0	-8.0	-8.7	-8.1	-9.5	-8.66	1	493.6	4	13
B7	ZINC09963258	-8.4	-8.3	-8.7	-9.1	-8.8	-8.66	0	485.5	6	14
B8	ZINC15825209	-8.7	-8.0	-8.8	-7.9	-9.9	-8.66	0	457.5	15	15
B9	ZINC09573345	-8.8	-8.2	-8.5	-7.8	-9.8	-8.62	0	441.5	35	19
B10	ZINC05093827	-8.8	-8.2	-8.1	-8.2	-9.5	-8.56	1	382.4	13	26
B11	ZINC09714883	-8.6	-7.7	-8.8	-8.5	-9.0	-8.52	0	460.4	22	33
B12	ZINC09609305	-9.1	-8.3	-8.1	-8.0	-9.0	-8.50	0	487.6	21	41
B13	ZINC17146977	-8.9	-7.9	-8.1	-7.9	-9.6	-8.48	1	461.3	40	46
B14	ZINC09336876	-9.1	-8.0	-8.0	-8.4	-8.8	-8.46	0	489.5	5	55
B15	ZINC24760364	-8.6	-7.9	-8.0	-8.3	-9.5	-8.46	0	463.5	12	60
B16	ZINC71749000	-8.9	-8.0	-8.1	-8.1	-9.2	-8.46	0	392.4	49	63
B17	ZINC12522154	-8.7	-8.0	-8.4	-8.4	-8.7	-8.44	0	459.5	48	68
B18	ZINC08826002	-8.4	-7.7	-8.4	-8.8	-8.7	-8.40	0	479.8	55	81
B19	ZINC32951574	-8.8	-8.1	-7.8	-8.5	-8.8	-8.40	0	445.5	31	83
B20	ZINC33037466	-9.0	-7.6	-7.9	-8.6	-8.9	-8.40	0	426.5	60	84
B21	ZINC09117544	-8.7	-8.0	-8.0	-8.2	-9.0	-8.38	1	446.6	26	88
B22	ZINC12912690	-8.8	-7.9	-8.5	-8.5	-8.1	-8.36	0	458.5	9	97
B23	ZINC14541866	-8.9	-8.0	-8.0	-7.9	-8.9	-8.34	1	445.6	30	110
B24	ZINC33116506	-8.5	-7.5	-8.7	-8.2	-8.8	-8.34	0	459.5	19	112
B25	ZINC06181558	-8.2	-7.9	-7.7	-8.5	-9.3	-8.32	0	454.9	46	121
B26	ZINC11964524	-8.6	-7.9	-8.6	-8.2	-8.3	-8.32	1	443.6	18	128
B27	ZINC09468232	-8.4	-8.1	-8.3	-7.9	-8.8	-8.30	0	479.5	10	143
B28	ZINC12548325	-8.4	-7.5	-8.4	-9.0	-8.2	-8.30	0	492.6	20	145
B29	ZINC02922837	-8.2	-7.6	-7.9	-8.0	-9.5	-8.24	0	426.5	45	195
B30	ZINC14992854	-8.9	-7.4	-8.1	-8.6	-8.2	-8.24	1	475.5	58	205
B31	ZINC58139425	-8.8	-8.0	-8.1	-7.9	-8.4	-8.24	0	445.5	61	210
B32	ZINC12486362	-8.5	-7.9	-8.3	-7.7	-8.7	-8.22	0	454.5	39	225
B33	ZINC12677534	-7.9	-7.8	-8.7	-8.5	-8.2	-8.22	0	486.6	64	226
B34	ZINC03642712	-8.8	-8.3	-7.4	-7.9	-8.6	-8.20	0	404.4	38	246
B35	ZINC33284369	-8.7	-7.6	-7.7	-8.6	-8.4	-8.20	0	449.5	67	257
B36	ZINC14980290	-8.5	-7.7	-8.1	-7.9	-8.7	-8.18	1	469.6	56	273
B37	ZINC72301224	-8.9	-7.9	-7.9	-8.4	-7.8	-8.18	0	494.6	37	292
B38	ZINC09604337	-8.4	-7.6	-8.1	-8.0	-8.7	-8.16	0	482.5	27	303
B39	ZINC33284430	-8.5	-7.7	-7.7	-8.9	-8.0	-8.16	0	469.9	68	318
B40	ZINC57270793	-8.3	-7.7	-8.6	-7.8	-8.4	-8.16	0	472.6	42	321
B41	ZINC12541787	-8.8	-7.7	-8.3	-7.6	-8.3	-8.14	0	466.5	29	339
B42	ZINC23360299	-8.0	-7.7	-8.0	-7.9	-8.6	-8.04	1	483.6	66	553
B43	ZINC20915400	-8.1	-7.4	-8.3	-7.6	-8.5	-7.98	0	442.6	584	734
B44	ZINC13362011	-8.3	-7.5	-7.3	-7.7	-8.0	-7.76	0	407.5	2435	1863
B45	ZINC02628238	-7.5	-7.2	-7.6	-8.0	-7.5	-7.56	0	410.9	1938	3793
B46	ZINC12151492	-8.1	-7.0	-7.1	-7.2	-7.8	-7.44	1	439.0	3167	5458
B47	ZINC06754062	-7.0	-7.7	-7.5	-7.2	-7.7	-7.42	0	392.5	4310	5723
B48	ZINC09588499	-8.1	-7.4	-7.1	-7.1	-7.3	-7.40	0	420.5	4140	6061
B49	ZINC55080042	-7.7	-7.4	-7.1	-6.9	-7.6	-7.34	1	407.5	2791	7239

The table shows ZINC IDs of selected molecules from Virtual Screen B, predicted binding energies to C3d structures R1-R5, the mean energy across all structures, charge, molecular weight (MW), and rankings. RMSD rank is the rank of the top binding pose consistent across all receptor structures. Molecules that matched features in pharmacophore rescreen are shown in blue.

While many docked poses had low predicted binding energies (as low as -10.1 kcal/mol), some compounds bound strongly to a small number of receptor conformations, and poorly to others. Some molecules docked in very different orientations to different receptor conformations, and others docked deep in the C3d cavity, such that inhibition of CR2 binding would be improbable. We clustered low energy compounds, based on RMSD of their docked conformations to the five diverse C3d structures, and identified compounds that had similar molecule binding modes (< 5 Å RMSD) to all five structures. Additionally, we rescreened all docked poses against the 84 pharmacophore models used in the initial screening round. Compounds were docked in-place; quick screening examined whether chemical groups in docked poses overlapped with pharmacophore features. We selected compounds that matched at least two pharmacophore features, while docked to at least three different C3d structures. We selected a list of hits with < -6.8 kcal/mol mean binding energy (corresponding to K_d = 10 μM) that met the RMSD and pharmacophore postprocessing criteria, yielding 14 compounds. In addition, we selected additional compounds that met RMSD criteria with energies < -8.0 kcal/mol, as top-binding compounds. In total, 50 newly identified molecules will be tested experimentally. These compounds and their properties are listed in Table 5.5, and their structures are shown in Figure E.4.

We plan to screen each of the selected molecules for direct binding to C3d using microscale thermophoresis, as described above for Molecule A1 from the initial virtual screen. Furthermore, we will screen molecules that directly bind to C3d for competitive inhibition with CR2, Efb-C, and Ecb. Molecules that bind C3d and competitively inhibit C3d-ligand interactions are amenable to optimization, and serve as a foundation for future drug design efforts.

5.4 References

1. Qu, H., D. Ricklin, and J.D. Lambris. 2009. Recent developments in low molecular weight complement inhibitors. Mol. Immunol. 47: 185–195.

2. Nagar, B, R.G. Jones, R.J. Diefenbach, D.E. Isenman, J.M. Rini. 1998. X-ray crystal structure of C3d: A C3 fragment and ligand for complement receptor 2. Science. 280: 1277–1281.

3. Hammel, M., G. Sfyroera, D. Ricklin, P. Magotti, J.D. Lambris, et al. 2007. A structural basis for complement inhibition by Staphylococcus aureus. Nat. Immunol. 8: 430–437.

4. Hammel, M., G. Sfyroera, S. Pyrpassopoulos, D. Ricklin, K.X. Ramyar, et al. 2007. Characterization of Ehp, a secreted complement inhibitory protein from Staphylococcus aureus. J. Biol. Chem. 282: 30051–30061.

5. van den Elsen, J.M.H., and D.E. Isenman. 2011. Complement receptor complex structure. Science. 332: 608–611.

6. Berman, H., J. Westbrook, and Z. Feng. 2000. The protein data bank. Nucleic Acids Res. 28: 235-242.

7. Pettersen, E.F., T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, et al. 2004. UCSF Chimera: A visualization system for exploratory research and analysis. J. Comput. Chem. 25: 1605–1612.

8. Phillips, J., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, et al. 2005. Scalable molecular dynamics with NAMD. J. Comput. Chem. 26: 1781–1802.

9. MacKerell, A.D., D. Bashford, M. Bellott, R.L. Dunbrack Jr, J.D. Evanseck, et al. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem B. 102: 3586-3616.

10. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: Visual molecular dynamics. J. Mol. Graphics. 14: 33–38.

11. Grant, B.J., A.P.C. Rodrigues, K.M. ElSawy, J.A. McCammon, and L.S.D. Caves. 2006. Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics. 22: 2695–2696.

12. Haspel, N., D. Ricklin, B.V. Geisbrecht, L.E. Kavraki, and J.D. Lambris. 2008. Electrostatic contributions drive the interaction between Staphylococcus aureus protein Efb-C and its complement target C3d. Protein Sci. 17: 1894–1906.

13. Gorham Jr, R.D., C.A. Kieslich, and D. Morikis. 2011. Complement inhibition by Staphylococcus aureus: Electrostatics of C3d–EfbC and C3d–Ehp association. Cel. Mol. Bioeng. 5: 32–43.

14. Wolber, G., and T. Langer. 2005. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. J. Chem. Inf. Model. 45: 160–169.

15. Lipinski, C.A. 2000. Drug-like properties and the causes of poor solubility and poor permeability. J. Pharmacol. Toxicol. Methods. 44: 235–249.

16. Irwin, J.J., T. Sterling, M.M. Mysinger, E.S. Bolstad, and R.G. Coleman. 2012. ZINC: A free tool to discover chemistry for biology. J. Chem. Inf. Model. 52: 1757-1768.

17. Trott, O., and A.J. Olson. 2009. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 31: 455–461.

18. Lin, J.-H., A.L. Perryman, J.R. Schames, and J.A. McCammon. 2003. The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme. Biopolymers. 68: 47–62.

19. Amaro, R.E., R. Baron, and J.A. McCammon. 2008. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. J. Comput. Aided Mol. Des. 22: 693–705.

20. Wang, J.-C., J.-H. Lin, C.-M. Chen, A.L. Perryman, and A.J. Olson. 2011. Robust scoring functions for protein-ligand interactions with quantum chemical charge models. J. Chem. Inf. Model. 51: 2528–2537.

21. Vanhee, P., A.M. van der Sloot, E. Verschueren, L. Serrano, F. Rousseau, et al. 2011. Computational design of peptide ligands. Trends Biotechnol. 29: 231–239.

22. Verdine, G.L., and G.J. Hilinski. 2012. Stapled peptides for intracellular drug targets. In: Protein Engineering for Therapeutics: Part B, New York, NY, Elsevier, US, pp. 3-30.

23. Szakonyi, G, J.M. Guthridge, D. Li, K. Young, V.M. Holers, and X.S. Chen. 2001. Structure of complement receptor 2 in complex with its C3d ligand. Science. 292: 1725–1728.

24. Clemenza, L., and D.E. Isenman. 2000. Structure-guided identification of C3d residues essential for its binding to complement receptor 2 (CD21). J. Immunol. 165: 3839–3848.

25. Morikis, D., and J.D. Lambris. 2004. The electrostatic nature of C3d-complement receptor 2 association. J. Immunol. 172: 7537–7547.

26. Zhang, L., B. Mallik, and D. Morikis. 2007. Immunophysical exploration of C3d–CR2(CCP1-2) interaction using molecular dynamics and electrostatics. J. Mol. Biol. 369: 567–583.

27. Cheung, A.S., C.A. Kieslich, J. Yang, and D. Morikis. 2010. Solvation effects in calculated electrostatic association free energies for the C3d-CR2 complex and comparison with experimental data. Biopolymers. 93: 509–519.

28. Hannan, J.P., K.A. Young, J.M. Guthridge, R. Asokan, G. Szakonyi, et al. 2005. Mutational analysis of the complement receptor type 2 (CR2/CD21)–C3d interaction reveals a putative charged SCR1 binding site for C3d. J. Mol. Biol. 346: 845–858.

29. Kovacs, J.M., J.P. Hannan, E.Z. Eisenmesser, and V.M. Holers. 2009. Mapping of the C3d ligand binding site on complement receptor 2 (CR2/CD21) using nuclear magnetic resonance and chemical shift analysis. J. Biol. Chem. 284: 9513–9520.

30. Shaw, C.D., M.J. Storek, K.A. Young, J.M. Kovacs, J.M. Thurman, et al. 2010. Delineation of the complement receptor type 2-C3d complex by site-directed mutagenesis and molecular docking. J. Mol. Biol. 404: 697–710.

31. Isenman, D.E., E. Leung, J.D. Mackay, S. Bagby, and J.M.H. van den Elsen. 2010. Mutational analyses reveal that the Staphylococcal immune evasion molecule Sbi and complement receptor 2 (CR2) share overlapping contact residues on C3d: Implications for the controversy regarding the CR2/C3d cocrystal structure. J. Immunol. 184: 1946–1955.

32. Kieslich, C.A., D. Morikis, J. Yang, and D. Gunopulos. 2011. Automated computational framework for the analysis of electrostatic similarities of proteins. Biotechnol. Prog. 27: 316–325.

33. Morikis, D. F1000.com/10371956.

CHAPTER 6: PERSPECTIVES AND FUTURE DIRECTIONS

Computational tools can provide new insights into our understanding of protein-protein interactions. Throughout recent years, major advances have been made in the development and application of computational methods toward biological systems, including electrostatic free energy calculations, molecular docking, and molecular dynamics simulations (1-4). In the application of these methods, there is a trade-off between speed and accuracy, and care must be taken in selecting appropriate methods for specific applications, and assessing the accuracy of these methods. In this thesis, we described the evaluation/parameterization of computational methods and application in understanding host regulation and bacterial evasion of complement immunity. We explored the use of computational methods to evaluate the physiological roles of binding interfaces from crystallographic structures and to provide insight in cases of experimental ambiguity. Furthermore, we discussed development and application of peptide design and virtual high-throughput screening methods for identification of low molecular mass complement-targeted therapeutics. In this chapter, we summarize the major outcomes from this thesis, as well as future work and ideas for improved implementation of computational methods toward both mechanistic understanding of protein-protein interactions and structure-based drug design.

Electrostatic interactions play an important role in the association of biomolecules under physiological conditions (5-8). Proteins are characterized by unique electrostatic properties, which stem from both primary sequence and structure. For many protein-protein interactions, the initial association step may be accelerated by electrostatic complementarity (3, 7, 9-11). Subsequently, short-range electrostatic interactions, including hydrogen bonds and salt bridges, contribute to energetically favorable contacts that stabilize the final bound protein complex. Consequently, calculation of protein electrostatics can provide insight into the associative properties of protein complexes (12-17). Poisson-Boltzmann electrostatics represents an efficient method for calculation of protein electrostatic potentials in a solvated environment, and serves as a predictive tool in the design of proteins and peptides with tailored electrostatic properties. In Chapter 2, we evaluated the accuracy of Poisson-Boltzmann electrostatic free energy calculations in predicting the relative association of protein mutants (18). Our data suggest that electrostatic free energy calculations are sufficient to predict association of protein complexes comprised of highly and oppositely charged species, but are not predictive for proteins of similar net charge, even if they possess electrostatic complementarity at their binding interfaces. Furthermore, we observed that calculation of Coulombic interactions between proteins (neglecting the effects of electrostatic screening by solvent) can predict association of highly/oppositely charged protein complexes with similar accuracy. We suggest that Coulombic and Poisson-Boltzmann free energy calculations can be used for rapid screening of designed protein/peptide variants. Since these free energy methods only reflect electrostatic contributions to protein interactions, more rigorous free energy calculations, including MM-GBSA or MM-PBSA free energies, should be used in conjunction to evaluate binding (17, 19, 20). Finally, entropic contributions can be estimated via independent molecular dynamics simulations for each binding partner and the protein complex (21). Together, these computational free energy methods provide a framework for protein and peptide design.

We used electrostatic free energy calculations to understand the molecular basis of complement system interactions. The complement system is a branch of innate immunity responsible for the rapid detection and elimination of pathogens in the bloodstream (see Chapter 1), and electrostatic interactions play an important role in driving the association of complement proteins with both host and pathogenic proteins in a high ionic strength environment (22-30). In Chapters 3 and 4, we specifically examined the interactions between secreted staphylococcal virulence factors (Efb, Ecb, and Sbi) and complement C3d (31). Implementation of the computational framework AESOP (30, 32) elucidated charged amino acid residues in these proteins crucial for complex association. We identified residues not only important for stabilizing bound C3d-Efb/Ecb/Sbi complexes at their interfaces, but also residues located away from binding interfaces, which may enhance or reduce electrostatically-mediated long-range association of these proteins. In conjunction, molecular and Brownian dynamics simulations provided insight into the dynamics and stability of these interactions, and helped to distinguish the physiological roles of two distinct binding modes of Sbi to C3d.

While electrostatics calculations and molecular dynamics simulations can aid in mechanistic understanding of protein interactions, they are also of vital importance in structure-based drug design. Our preliminary data for C3d-Efb/Ecb/Sbi complexes served as a foundation for the design of potential complement-targeted and anti-infective therapeutic molecules. Since most complement-targeted therapeutics (both in the pipeline and on the market) are large biomolecules, it is of interest to identify low molecular-mass therapeutics which alleviate many problems associated with biopharmaceuticals (see Chapter 1). Furthermore, targeting interactions between staphylococcal virulence factors and complement may aid in development of novel therapeutics against strains of *Staphylococcus aureus* resistant to traditional antibiotics.

We employed a multifaceted peptide design framework, in which we isolated small structural elements from crystallographic structures of the C3d-Efb/Ecb/Sbi protein complexes, and evaluated the resulting peptides for structural stability and binding to their respective targets (see Chapter 5). C3d-derived peptides held initial promise in the design of molecules aimed at disrupting staphylococcal interactions with complement, serving as an anti-infective strategy to reduce complement evasion (and in turn virulence) of *Staphylococcus aureus*. We used an iterative approach to make a series of truncations and mutations that simultaneously enhanced peptide polarity/solubility, maintained peptide structural integrity, and improved binding to bacterial ligands Efb, Ecb, and Sbi. We also designed peptides based on the structures of Efb, Ecb, and Sbi, which may also disrupt C3d-ligand binding, but may alternatively inhibit complement activation. While our initial peptide design efforts yielded peptides that lacked desired structural, binding, and functional properties, the peptide design process revealed steps necessary for improvement. First, we believe that explicit solvent molecular dynamics simulations may not be sufficient to explore unfolding or major structural changes of peptides. Alternative methods, such as implicit solvent molecular dynamics, replica exchange, or simulated annealing, are better suited to examine both structural stability and ligand binding of peptides (33). Second, since peptides are truncated from larger proteins, some desired ligand-binding interactions are lost right away, and in turn, remaining interfacial residues should be optimized to improve ligand interactions to compensate for this loss. Peptide design would benefit from a more rigorous search of sequence space through implementation of bioinformatics approaches and existing peptide design frameworks, in order to identify a large set of peptides with sequences amenable to desired secondary structure and potential ligand binding. Finally, cyclization/linkage methods (i.e. disulfide bonds, peptide stapling) may help to alleviate problems with peptide structural integrity in solution, and improve ligand binding (34, 35).

We also performed virtual high-throughput screening to identify drug-like molecules capable of binding to C3d and disrupting C3d-ligand interactions (see Chapter 5). C3d not only interacts with staphylococcal virulence factors, but also with complement receptor 2 (CR2), at the same binding site. We used a framework comprised of molecular dynamics simulations, pharmacophore screening, and docking to screen the ZINC drug-like (36, 37) database (consisting of ~7 million compounds) for molecules to inhibit C3d-ligand interactions. In our initial screen, we targeted the relatively flat binding interface between C3d and Efb, and identified one molecule with micromolar binding to C3d, but no inhibition of C3d-Efb interaction. Initial success motivated a more rigorous screen, in which we targeted the lower affinity C3d-CR2 interaction as a potential complement-targeted therapeutic. In this screen, we used molecular dynamics simulations and experimental binding/mutagenesis data to create dynamic pharmacophore models, based on residues of CR2 interacting with C3d. We screened ~1.1 billion generated conformers of ZINC drug-like molecules against 84 selected pharmacophore models, yielding ~25000 molecules with similar geometric and physicochemical properties to regions of CR2. We docked these molecules to five distinct C3d structures (extracted from MD simulations), and ranked molecules based on mean predicted binding energies to C3d structures, spatial consistency of docked binding modes, and overlap of docked binding modes with pharmacophore models. We selected 50 molecules, which we plan to test for direct C3d binding and inhibition of the C3d-CR2 interaction *in vitro* (using microscale thermophoresis and ELISA, respectively), followed by further functional studies with any molecules identified to bind C3d.

In summary, we examined the molecular basis for complement evasion by *Staphylococcus aureus* via interaction between C3d and virulence factors Efb, Ecb, and Sbi. New insights provided a foundation for structure-based drug design. We designed peptides and screened for molecules to inhibit these interactions, as a potential anti-infective strategy. Furthermore, we targeted the interaction between C3d and CR2, as a complement therapeutic strategy targeted at the production of autoantibodies in diseases such as systemic lupus erythematosus. Our computational frameworks are amenable to further development and optimization in order to enhance design and identification of therapeutics for other targets in future studies.

References

1. Honig, B., and A. Nicholls. 1995. Classical electrostatics in biology and chemistry. Science. 268: 1144–1149.

2. Baker, N.A. 2004. Poisson–Boltzmann Methods for Biomolecular Electrostatics. In: Methods in Enzymology. Elsevier. pp. 94–118.

3. Adcock, S.A., and J.A. McCammon. 2006. Molecular dynamics: Survey of methods for simulating the activity of proteins. Chem. Rev. 106: 1589–1615.

4. Jorgensen, W.L. 2004. The many roles of computation in drug discovery. Science. 303: 1813–1818.

5. Perutz, M. 1978. Electrostatic effects in proteins. Science. 201: 1187-1191.

6. Davis, M.E., and J.A. McCammon. 1990. Calculating electrostatic forces from grid-calculated potentials. J. Comput. Chem. 11: 401–409.

7. Sheinerman, F., and R. Norel. 2000. Electrostatic aspects of protein-protein interactions. Curr. Opin. Struct. Biol. 10: 153-159

8. WU, J., and D. Morikis. 2006. Molecular thermodynamics for charged biomacromolecules. Fluid Phase Equilibria. 241: 317–333.

9. Gabdoulline, R.R., and R.C. Wade. 1997. Simulation of the diffusional association of barnase and barstar. Biophys. J. 72: 1917–1929.

10. McCammon, J.A., S.H. Northrup, and S.A. Allison. 1986. Diffusional dynamics of ligand-receptor association. J. Phys. Chem. 90: 3901–3905.

11. Selzer, T., and G. Schreiber. 2001. New insights into the mechanism of protein-protein association. Proteins. 45: 190–198.

12. Gilson, M.K., A. Rashin, R. Fine, and B. Honig. 1985. On the calculation of electrostatic interactions in proteins. J. Mol. Biol. 184: 503–516.

13. Gilson, M., and K. Sharp. 1988. Calculating the electrostatic potential of molecules in solution: Method and error assessment. J. Comput. Chem. 9: 327-335.

14. Warshel, A., and A. Papazyan. 1998. Electrostatic effects in macromolecules: fundamental concepts and practical modeling. Curr. Opin. Struct. Biol. 8: 211–217.

15. Wagoner, J., and N.A. Baker. 2004. Solvation forces on biomolecular structures: A comparison of explicit solvent and Poisson-Boltzmann models. J. Comput. Chem. 25: 1623–1629.

16. Lu, B.Z., Y.C. Zhou, M.J. Holst, and J.A. McCammon. 2008. Recent progress in numerical methods for the Poisson-Boltzmann equation in biophysical applications. Commun. Comput. Phys. 3: 973-1009

17. Dong, F., and B. Olsen. 2008. Computational methods for biomolecular electrostatics. Methods in cell biology.

18. Gorham Jr, R.D., C.A. Kieslich, A. Nichols, N.U. Sausman, M. Foronda, et al. 2011. An evaluation of poisson-boltzmann electrostatic free energy calculations through comparison with experimental mutagenesis data. Biopolymers. 95: 746–754.

19. Kieslich, C.A., P. Tamamis, R.D Gorham Jr, A. López de Victoria, N.U. Sausman, et al. 2012. Exploring protein-protein and protein-ligand interactions in the immune system using molecular dynamics and continuum electrostatics. Curr. Phys. Chem. 2: 324-343.

20. Durrant, J.D., and J.A. McCammon. 2011. Molecular dynamics simulations and drug discovery. BMC Biol. 9: 71.

21. Tamamis, P., A. López de Victoria, R.D. Gorham, M.L. Bellows-Peterson, P. Pierou, et al. 2012. Molecular dynamics in drug design: new generations of compstatin analogs. Chem. Biol. Drug Des. 79: 703–718.

22. Morikis, D., and L. Zhang. 2006. An immunophysical study of the complement system: Examples for the pH dependence of protein binding and stability. J. Non-Cryst. Sol. 352: 4445–4450.

23. Ricklin, D., A. Tzekou, B.L. Garcia, M. Hammel, W.J. McWhorter, et al. 2009. A molecular insight into complement evasion by the Staphylococcal complement inhibitor protein family. J. Immunol. 183: 2565–2574.

24. Pyaram, K., C.A. Kieslich, V.N. Yadav, D. Morikis, and A. Sahu. 2010. Influence of electrostatics on the complement regulatory functions of Kaposica, the complement inhibitor of Kaposi's Sarcoma-associated herpesvirus. J. Immunol. 184: 1956–1967.

25. Morikis, D., and J.D. Lambris. 2004. The electrostatic nature of C3d-complement receptor 2 association. J. Immunol. 172: 7537–7547.

26. Sfyroera, G., M. Katragadda, D. Morikis, S.N. Isaacs, and J.D. Lambris. 2005. Electrostatic modeling predicts the activities of orthopoxvirus complement control proteins. J. Immunol. 174: 2143–2151.

27. Zhang, L., and D. Morikis. 2006. Immunophysical properties and prediction of activities for vaccinia virus complement control protein and smallpox inhibitor of complement enzymes using molecular dynamics and electrostatics. Biophys. J. 90: 3106–3119.

28. Zhang, L., B. Mallik, and D. Morikis. 2007. Immunophysical exploration of C3d–CR2(CCP1-2) interaction using molecular dynamics and electrostatics. J. Mol. Biol. 369: 567–583.

29. Cheung, A.S., C.A. Kieslich, J. Yang, and D. Morikis. 2010. Solvation effects in calculated electrostatic association free energies for the C3d-CR2 complex and comparison with experimental data. Biopolymers. 93: 509–519.

30. Kieslich, C.A., D. Morikis, J. Yang, and D. Gunopulos. 2011. Automated computational framework for the analysis of electrostatic similarities of proteins. Biotechnol. Prog. 27: 316–325.

31. Gorham Jr, R.D., C.A. Kieslich, and D. Morikis. 2011. Complement inhibition by Staphylococcus aureus: Electrostatics of C3d–EfbC and C3d–Ehp association. Cel. Mol. Bioeng. 5: 32–43.

32. Gorham Jr, R.D., C.A. Kieslich, and D. Morikis. 2010. Electrostatic clustering and free energy calculations provide a foundation for protein design and optimization. Ann. Biomed. Eng. 39: 1252–1263.

33. Stavrakoudis, A. 2010. Conformational flexibility in designing peptides for immunology: the molecular dynamics approach. Curr. Comput. Aided. Drug Des. 6: 207–222.

34. Vanhee, P., A.M. van der Sloot, E. Verschueren, L. Serrano, F. Rousseau, et al. 2011. Computational design of peptide ligands. Trends Biotechnol. 29: 231–239.

35. Verdine, G.L., and G.J. Hilinski. 2012. Stapled peptides for intracellular drug targets. In: Protein Engineering for Therapeutics: Part B, New York, NY, Elsevier, US, pp. 3-30.

36. Irwin, J.J., T. Sterling, M.M. Mysinger, E.S. Bolstad, and R.G. Coleman. 2012. ZINC: A Free Tool to Discover Chemistry for Biology. J. Chem. Inf. Model. 52: 1757-1768.

37. Lipinski, C.A. 2000. Drug-like properties and the causes of poor solubility and poor permeability. J. Pharmacol. Toxicol. Methods. 44: 235–249.

APPENDIX A: THEORETICAL BASIS FOR COMPUTATIONAL METHODS

A.1 Electrostatics calculations

Electrostatics plays a crucial role in the association of complement proteins with their targets. In the analysis of electrostatics, the fundamental governing equation is Coulomb's Law. This equation allows calculation of electrostatic potential based on charge (q), dielectric coefficient (ε), and distance (r), as shown below,

$$V(r) = \frac{q}{4\pi\varepsilon_0 \varepsilon r} \tag{1}$$

In this equation, ε_0 represents the vacuum permittivity constant. The applicability of this expression to large molecules requires the calculation of Coulomb's Law between one charged atom and every other charged atom in the molecule, followed by summation of all values to yield an overall electrostatic potential value for that point in the molecule. Since proteins are not present in an isolated vacuum environment, it becomes necessary to represent the effects of solvent molecules on the electrostatic potential of the protein. Unfortunately, this significant addition of molecules to the simulated environment causes the calculation of Coulomb's Law to become computationally expensive, especially for large systems.

As an alternative to Coulombic calculations in an explicit solvated environment, Poisson-Boltzmann calculations are widely accepted and represent the forefront of accurate and efficient calculation of electrostatic potential for biomolecules. Using this methodology, water molecules and ions are not explicitly modeled in the simulation, but their presence is implicitly accounted for in the linearized Poisson-Boltzmann equation (LPBE) shown below,

$$-\nabla \cdot \varepsilon(r) \nabla \varphi(r) + \varepsilon_0 \, \varepsilon(r) \kappa^2(r) \varphi(r) = \frac{4\pi e^2}{\varepsilon_0 k_b T} \sum_{i=1}^{F} z_i \delta(r - r_i) \tag{2}$$

This equation allows for the calculation of electrostatic potential (φ) as a function of distance (r), dielectric coefficient ($\varepsilon(r)$), ion accessibility ($\kappa(r)$), and ionic valence (z). The constant terms in the equation represent vacuum permittivity (ε_0), electron unit charge (e), the Boltzmann constant (k_b), and temperature (T). Since water molecules are not explicitly modeled, the effects of solvent polarizability on the electrostatic potential of the protein are incorporated into the above equation via a distance-dependent dielectric coefficient. The effect of ionic screening is accounted for by the ion accessibility function $\kappa(r)$, which incorporates ionic strength as shown by the following equations

$$\kappa(r) = \frac{4e^2 I}{\varepsilon_0 \varepsilon k_b T} \quad (3),$$

$$I = \frac{1}{2}\sum_{i=1}^{M} z_i^2 n_i \quad (4).$$

Ionic strength (I) is dependent on ionic valence (z) and concentration (n) of each ion species. By defining an appropriate boundary for dielectric coefficient and ionic accessibility, the Poisson-Boltzmann equation can be solved numerically for electrostatic potential. Electrostatic potential values can be translated into electrostatic free energies of proteins, which in turn allows for calculation of association and solvation free energies of protein complexes.

A.2 Brownian dynamics simulations

Brownian dynamics (BD) simulations are used to analyze the diffusional and electrostatically-driven association of proteins. Brownian motion represents the movement of a solute through a viscous medium. In the case of protein diffusion, the Reynolds number describing the flow of fluid around the protein is sufficiently small to indicate that viscous forces dominate over inertial forces, thus Brownian motion can be assumed. In addition to simulating the motion of proteins and observing their trajectories through space, Brownian dynamics can also be used to calculate k_{on} rates for protein-protein association. This method can be used to calculate the association rates for a series of mutants, in order to compare their binding and deduce the influence of electrostatics on protein association.

The governing equation for BD simulations originates from Newton's second law of motion.

$$ma = F + \beta v + S(t) \quad (5).$$

The Langevin equation, shown in Eq. (5), separates the force term into three independent components. The first term on the right represents the forces that arise from potential energy fields acting on the proteins. These forces may consist of both bonded and non-bonded interactions, similarly to molecular dynamics force fields. The second term approximates the drag force exerted on the diffusing protein. This term is used since the protein is modeled using continuum solvation, and the explicit forces between the water molecules and the diffusing protein are not determined. The magnitude of this term is inversely proportional to thermal energy, and thus temperature plays a role in the diffusion of the protein. The last term represents the random motion of the protein due to collisions with water molecules. Since continuum solvation is used, this force cannot be explicitly determined and is estimated using a stochastic term.

Brownian dynamics implies that viscous forces dominate over inertial forces. Given this assumption, particle acceleration is assumed to be negligible. From Stokes-Einstein equation, the drag coefficient can be written as shown below in Eq. (6), as a function of temperature (T) and diffusivity (D).

$$\beta = \frac{k_b T}{D} \tag{6}$$

Substituting position for velocity and integrating, the following relation is obtained,

$$r_{i+1} = r_i + \frac{D}{k_b T} F + \sqrt{2D\Delta t} R \tag{7}$$

Eq. (7) predicts the future position (r_{i+1}) of the particle based on the current position of the particle (r_i), the external and drag forces (F) and frictional coefficient (D/k_bT), and the random motion of the particle due to kinetic energy, which is proportional to the square root of diffusivity (D) and time step (Δt).

In order to determine the external force acting on a protein at any given time, force fields used in molecular dynamics (MD) simulations are implemented. A significant difference exists between MD and BD simulations. In MD simulations, all atomic interactions are taken into account. This includes van der Waals interactions, Coulombic interactions, and covalent geometry potential energies. In BD simulations, however, proteins are typically assumed to be rigid. This means that the bonds and angles within each protein molecule remain stationary at all times, and only interactions between one protein molecule and another protein molecule are calculated at each time step. While this represents a significant simplification, the proteins still retain their properties down to atomic resolution, providing accurate calculation of interactions with other proteins. Additionally, this rigid molecule assumption greatly reduces computational time necessary for simulation, thus much larger time scales can be explored compared to a traditional MD simulation. As a result, only non-bonded interactions between protein molecules are examined, as shown below,

$$U(r) = (-E_{min})\left[\left(\frac{R_{min}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{min}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}} \quad (8).$$

Each term in Eq. (8) is described in the discussion on the MD force field (Appendix A.3). This interaction energy term can be translated into force similarly to the case for the MD force field described earlier, which can be implemented in BD simulations.

In order to analyze the results of BD simulations, the number of diffusional trajectories that lead to encounters (successful or not), versus how many "escape" from the region of interest, needs to be determined. The target protein remains stationary during the simulation, and is positioned at the center of two larger spheres. The inner sphere, known as the b-radius, represents the location of the center of mass of the diffusing species at the start of each trajectory. This radius is chosen such that forces between the particles are symmetric outside the b-radius, but influence the trajectories within the radius. The outer sphere, or q-radius, is chosen such that the flux of the diffusing species across this boundary does not favor a certain direction (i.e. motion is purely diffusive).

In order to quantify the results of BD simulations, it is necessary to calculate k_{on} rates. Rates are calculated, using data obtained from BD simulations, as shown in the equations below,

$$k_{on} = k_D(b)P \quad (9),$$

$$k_D(b) = \left(\int_b^{\infty} dr\left[\frac{exp(U(r)/k_bT)}{4\pi r^2 D}\right]\right)^{-1} \quad (10),$$

$$P=\frac{\left(\frac{\beta}{1-(1-\beta)\Omega}\right)\alpha}{1-(1-\alpha)\left(\Delta+\left(\frac{\beta}{1-(1-\beta)\Omega}\right)(1-\Delta)\right)} \qquad (11).$$

In these equations, $U(r)$ is the potential energy of interaction between the two proteins, D is the diffusivity, P is the overall probability that the diffusing species will successfully react with the target protein, β is the probability of collision, α is the probability of successful collision (meaning the diffusing species makes sufficient contacts and will not diffuse away from the target protein), Δ is the probability that an unsuccessful collision will re-collide with the target protein, and Ω is the probability that a particle at the q-radius will diffuse back within the b-radius. Thus, BD simulations can yield quantitative information regarding the association kinetics of two proteins in a physiological environment.

A.3 Molecular dynamics simulations

Molecular dynamics (MD) simulations are widely used to computationally model the dynamics of protein structure at atomic resolution. In order to better understand the functional characteristics of a protein, it is important to examine the molecular motion on a wide range of time scales. The motion of protein molecules can be approximated using Newton's Laws of motion, and all equations utilized in molecular mechanics are derived from the basic principles of classical physics. The derivation of molecular dynamics begins with the definition of a potential energy function, which describes all physicochemical interactions within a protein molecule. In a full MD simulation, water molecules surround the protein (explicit solvation), and the interactions between the protein and the solvent are directly taken into account. Many approximations are made to facilitate efficient calculations. Energetically preferred topologies and empirical parameters are included in the force field. In order to reduce the number of calculations in the simulation, periodic boundary conditions are used to limit the number of solvent molecules in the protein environment, while maintaining the accuracy of the calculation.

In MD simulations, atomic interaction energies are calculated according to the potential energy function below,

$$U(r) = \sum_{bonds} k\left(r_{ij} - r_0\right)^2 + \sum_{angles} k_\theta\left(\theta - \theta_0\right)^2 + \sum_{angles} k_\psi\left(1 + \cos\left(n\psi + \phi\right)\right)$$
$$+ \sum_{impropers} k_\varphi\left(\varphi - \varphi_0\right)^2 + + \sum_{nonbonded} \left\{ \varepsilon_{ij}\left[\left(\frac{R_{\min}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{\varepsilon_i \varepsilon r_{ij}} \right\} \qquad (12).$$

This equation incorporates both bonded and nonbonded interactions. Bonded interactions are based on Hooke's Law, in which the value of a particular parameter oscillates around an equilibrium value. The k parameters in the bond, angle, torsion, and improper terms represent force constants. The r_0 and θ_0 values represent equilibrium positions of each bond length and angle, while r_{ij} and θ describe the time-dependent values of these parameters. The value for torsion angles is described by a trigonometric function, where n, ψ, and φ describe the function multiplicity, dihedral angle, and phase shift, respectively. The force constants, along with the equilibrium parameter values, have been determined empirically, and are intrinsic to the specific force field used. Nonbonded interactions include van der Waals and electrostatic interactions, which are governed by "through space" interatomic distance. The van der Waals interactions are represented by the 6-12 Lennard-Jones potential, which describes the energy of interaction as a function of the distance between atoms (r_{ij}), distance of the bottom of the energy well (R_{min}), and the depth of the energy well (ε_{ij}). These interactions are only significant at relatively short distance, since the interaction energy decays proportional to $1/r^6$ or $1/r^{12}$. Based on the potential function, van der Waals interactions theoretically exist between all atoms pairs separated by at least three covalent bonds within the protein. Since many of these interactions are extremely weak, a cutoff distance is used to alleviate calculations of negligible contribution to the overall energy of the system. Finally, Coulomb's law describes the energetic contributions of electrostatic interactions. Like the Lennard-Jones potential, Coulombic interactions in theory have no distance limit. The energy of these interactions decreases proportional to $1/r$ (much slower than van der Waals interactions), but a distance cutoff is used to save computational resources and time.

In order to translate the potential energy function into molecular motions, Newton's second law of motion is employed. Potential energy is incorporated via the following relation,

$$F_i = -\frac{dU}{dr_i} \tag{13}$$.

Since force is equal to the product of mass and acceleration, and acceleration is equal to the second derivative of position with respect to time, the position of each atom in the protein can be described as a function of molecular force. The position of each atom at finite time interval Δt can be described by a Taylor series expansion,

$$x(t+\Delta t) = x(t) + \frac{dx(t)}{dt}\Delta t + \frac{d^2x(t)}{dt^2}\frac{(\Delta t)^2}{2} + \ldots \tag{14}$$.

The first term represents the position at time *t*, the second term arises from velocity due to kinetic energy, and the third term represents atomic acceleration, which is related to potential energy via Newton's second law. Numerical algorithms, such as the Verlet and leapfrog algroithms, are used in molecular dynamics software to integrate equations of motion, in order to calculate the relative positions of all protein and solvent atoms at small time intervals.

APPENDIX B: SUPPORTING DATA FOR CHAPTER 2

Note: A subset of the supporting data for Chapter 2 is shown below. For additional supporting data, including tables containing raw free energy data, statistical plots, and software input files, please visit the original article at http://onlinelibrary.wiley.com/doi/10.1002/bip.21644/full.

Table B.1 Coarse/fine grid dimension and resolution of selected protein complexes.

	Grid[a]	Barnase-Barstar	Im9-E9 DNase	RNAi-Ang	hGH-hGHbp	gp120-CD4
LBOX	CG Size (Å)	125 × 125 × 130	175 × 175 × 175	--	185 × 170 × 190	160 × 140 × 180
	CG Res (Å)	1.0 × 1.0 × 1.0	1.4 × 1.4 × 1.4	--	1.4 × 1.3 × 1.5	1.3 × 1.1 × 1.4
	FG Size (Å)	125 × 125 × 130	175 × 175 × 175	--	185 × 170 × 190	160 × 140 × 180
	FG Res (Å)	1.0 × 1.0 × 1.0	1.4 × 1.4 × 1.4	--	1.4 × 1.3 × 1.5	1.3 × 1.1 × 1.4
SBOX	CG Size (Å)	85 × 85 × 90	90 × 108 × 105	180 × 173 × 180	135 × 113 × 135	135 × 173 × 143
	CG Res (Å)	0.7 × 0.7 × 0.7	0.7 × 0.7 × 0.8	1.4 × 1.4 × 1.4	1.1 × 0.9 × 1.1	1.1 × 1.4 × 1.1
	FG Size (Å)	55 × 55 × 60	60 × 72 × 70	120 × 115 × 120	90 × 75 × 90	90 × 115 × 95
	FG Res (Å)	0.4 × 0.4 × 0.5	0.5 × 0.6 × 0.5	0.9 × 0.9 × 0.9	0.7 × 0.6 × 0.7	0.7 × 0.9 × 0.7

[a] CG = coarse grid, FG= fine grid; column gives information on grid size and resolution for each complex.

Table B.2 Correlation between Calculated and Experimental Free Energy (unabridged).

		ΔΔG (Solvation)				ΔG (Solution)				ΔG (Coulombic)			
		ε_P=2	ε_P=10	ε_P=20	ε_P=40	ε_P=2	ε_P=10	ε_P=20	ε_P=40	ε_P=2	ε_P=10	ε_P=20	ε_P=40
LPBE LBOX 0 mM	1	-0.82	-0.83	-0.83	-0.83	0.21	0.71	0.79	0.82	0.83	0.83	0.83	0.83
	2	-0.72	-0.70	-0.68	-0.66	0.79	0.85	0.83	0.78	0.73	0.73	0.73	0.73
	3	--	--	--	--	--	--	--	--	--	--	--	--
	4	-0.16	-0.16	-0.16	-0.16	-0.16	-0.05	0.03	0.09	0.13	0.13	0.13	0.13
	5	0.15	0.15	0.15	0.15	0.15	-0.01	-0.07	-0.11	-0.13	-0.13	-0.13	-0.13
LPBE LBOX 150 mM	1	-0.82	-0.83	-0.83	-0.82	0.10	0.59	0.74	0.81	--	--	--	--
	2	-0.72	-0.69	-0.65	-0.58	0.85	0.96	0.96	0.95	--	--	--	--
	3	--	--	--	--	--	--	--	--	--	--	--	--
	4	-0.16	-0.15	-0.14	-0.11	-0.16	-0.08	0.00	0.12	--	--	--	--
	5	0.15	0.16	0.17	0.18	0.24	0.26	0.24	0.18	--	--	--	--
LPBE SBOX 0 mM	1	-0.83	-0.83	-0.83	-0.83	0.19	0.71	0.79	0.82	0.83	0.83	0.83	0.83
	2	-0.74	-0.71	-0.69	-0.66	0.17	0.81	0.81	0.78	0.73	0.73	0.73	0.73
	3	-0.49	-0.49	-0.48	-0.48	0.22	0.42	0.46	0.47	0.48	0.48	0.48	0.48
	4	-0.16	-0.17	-0.17	-0.17	-0.17	-0.07	0.01	0.09	0.13	0.13	0.13	0.13
	5	0.15	0.15	0.15	0.15	0.16	0.00	-0.07	-0.11	-0.13	-0.13	-0.13	-0.13
LPBE SBOX 150 mM	1	-0.83	-0.83	-0.83	-0.83	0.06	0.59	0.74	0.81	--	--	--	--
	2	-0.73	-0.70	-0.66	-0.59	-0.50	0.92	0.94	0.95	--	--	--	--
	3	-0.49	-0.48	-0.48	-0.46	0.06	0.29	0.38	0.46	--	--	--	--
	4	-0.16	-0.15	-0.14	-0.11	-0.18	-0.10	-0.02	0.11	--	--	--	--
	5	0.15	0.16	0.17	0.18	0.24	0.27	0.25	0.18	--	--	--	--
NPBE LBOX 0 mM	1	-0.82	-0.83	-0.83	-0.83	0.21	0.71	0.79	0.82	0.83	0.83	0.83	0.83
	2	-0.72	-0.70	-0.68	-0.66	0.79	0.85	0.83	0.78	0.73	0.73	0.73	0.73
	3	--	--	--	--	--	--	--	--	--	--	--	--
	4	-0.16	-0.16	-0.16	-0.16	-0.16	-0.05	0.03	0.09	0.13	0.13	0.13	0.13
	5	0.15	0.15	0.15	0.15	0.15	-0.01	-0.07	-0.11	-0.13	-0.13	-0.13	-0.13
NPBE LBOX 150 mM	1	-0.83	-0.83	-0.83	-0.82	-0.03	0.58	0.74	0.81	--	--	--	--
	2	-0.72	-0.69	-0.65	-0.59	0.83	0.96	0.96	0.96	--	--	--	--
	3	--	--	--	--	--	--	--	--	--	--	--	--
	4	-0.15	-0.15	-0.14	-0.11	-0.17	-0.08	0.00	0.12	--	--	--	--
	5	0.15	0.17	0.17	0.19	0.25	0.30	0.30	0.23	--	--	--	--
NPBE SBOX 0 mM	1	-0.83	-0.83	-0.83	-0.83	0.19	0.71	0.79	0.82	0.83	0.83	0.83	0.83
	2	-0.74	-0.71	-0.69	-0.66	0.17	0.81	0.81	0.78	0.73	0.73	0.73	0.73
	3	-0.49	-0.49	-0.48	-0.48	0.22	0.42	0.46	0.47	0.48	0.48	0.48	0.48
	4	-0.16	-0.17	-0.17	-0.17	-0.17	-0.07	0.01	0.09	0.13	0.13	0.13	0.13
	5	0.15	0.15	0.15	0.15	0.16	0.00	-0.07	-0.11	-0.13	-0.13	-0.13	-0.13
NPBE SBOX 150 mM	1	-0.83	-0.83	-0.83	-0.82	-0.06	0.58	0.74	0.81	--	--	--	--
	2	-0.73	-0.70	-0.66	-0.59	-0.57	0.93	0.95	0.95	--	--	--	--
	3	-0.48	-0.48	-0.48	-0.47	0.02	0.28	0.37	0.44	--	--	--	--
	4	-0.16	-0.15	-0.14	-0.11	-0.18	-0.11	-0.02	0.11	--	--	--	--
	5	0.16	0.17	0.18	0.19	0.25	0.30	0.30	0.24	--	--	--	--

$^{a}\Delta\Delta G^{solvation}$, $\Delta G^{Coulombic}$, and $\Delta G^{solution}$ values were calculated according to Eqs. 4, 5, and 6. All free energy values were calculated using four different dielectric coefficients, and under eight calculation conditions. LPBE: linearized Poisson-Boltzmann equation, NPBE: nonlinear Poisson-Boltzmann equation, LBOX: uniform coarse grid and fine grid fit to electrostatic potential at ±1 k_BT/e, SBOX: coarse grid equals protein dimensions multiplied by 1.5, fine grid fit to protein dimensions, 0 mM: no ions present, 150 mM: physiological ionic strength, The numbers on the left refer to the following complexes: 1: Barnase-Barstar, 2: Im9-E9 DNase, 3: RNase Inhibitor-Angiogenin, 4: hGH-hGHbp, 5: gp120-CD4. RNase Inhibitor-Angiogenin free energies were not calculated using LBOX conditions, as the complex was too large and excessively charged to achieve sufficient grid resolution.

Table B.3 Statistical ANOVA p-values for calculated and experimental free energy data.

		ΔΔG (Solvation)				ΔG (Solution)				ΔG (Coulombic)			
		$\varepsilon_P=2$	$\varepsilon_P=10$	$\varepsilon_P=20$	$\varepsilon_P=40$	$\varepsilon_P=2$	$\varepsilon_P=10$	$\varepsilon_P=20$	$\varepsilon_P=40$	$\varepsilon_P=2$	$\varepsilon_P=10$	$\varepsilon_P=20$	$\varepsilon_P=40$
LPBE SBOX 0 mM	1	0.00	0.00	0.00	0.00	0.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	2	0.02	0.02	0.03	0.04	0.64	0.00	0.00	0.01	0.02	0.02	0.02	0.02
	3	0.08	0.08	0.08	0.08	0.45	0.14	0.10	0.09	0.08	0.08	0.08	0.08
	4	0.25	0.24	0.24	0.24	0.22	0.61	0.93	0.54	0.37	0.37	0.37	0.37
	5	0.46	0.46	0.46	0.45	0.43	1.00	0.74	0.61	0.53	0.53	0.53	0.53
LPBE SBOX 150 mM	1	0.00	0.00	0.00	0.00	0.80	0.00	0.00	0.00	--	--	--	--
	2	0.02	0.02	0.04	0.07	0.14	0.00	0.00	0.00	--	--	--	--
	3	0.08	0.08	0.08	0.09	0.83	0.31	0.18	0.10	--	--	--	--
	4	0.26	0.28	0.32	0.42	0.21	0.48	0.89	0.44	--	--	--	--
	5	0.45	0.42	0.40	0.37	0.24	0.19	0.22	0.37	--	--	--	--

*Values of 0.00 indicate $p < 0.005$.

Table B.4 Dependence of free energy correlation on grid resolution (Barnase-Barstar).

Grid Res (Å)	Ion Conc (mM)	ΔΔG (Solvation)				ΔG (Solution)				ΔG (Coulombic)			
		$\varepsilon_P=2$	$\varepsilon_P=10$	$\varepsilon_P=20$	$\varepsilon_P=40$	$\varepsilon_P=2$	$\varepsilon_P=10$	$\varepsilon_P=20$	$\varepsilon_P=40$	$\varepsilon_P=2$	$\varepsilon_P=10$	$\varepsilon_P=20$	$\varepsilon_P=40$
1.5	0	-0.82	-0.82	-0.82	-0.82	0.34	0.71	0.78	0.81	0.82	0.82	0.82	0.82
1.0	0	-0.82	-0.82	-0.82	-0.82	0.16	0.68	0.77	0.80	0.82	0.82	0.82	0.82
0.5	0	-0.82	-0.82	-0.82	-0.82	0.18	0.70	0.78	0.81	0.82	0.82	0.82	0.82
1.5	150	-0.82	-0.82	-0.82	-0.82	0.24	0.62	0.73	0.79	--	--	--	--
1.0	150	-0.82	-0.82	-0.82	-0.82	0.05	0.55	0.71	0.79	--	--	--	--
0.5	150	-0.82	-0.82	-0.82	-0.81	0.06	0.58	0.73	0.79	--	--	--	--

Grid Res (Å)	Ion Conc (mM)	ΔΔG (Solvation)				ΔG (Solution)				ΔG (Coulombic)			
		$\varepsilon_P=2$	$\varepsilon_P=10$	$\varepsilon_P=20$	$\varepsilon_P=40$	$\varepsilon_P=2$	$\varepsilon_P=10$	$\varepsilon_P=20$	$\varepsilon_P=40$	$\varepsilon_P=2$	$\varepsilon_P=10$	$\varepsilon_P=20$	$\varepsilon_P=40$
1.5	0	-0.80	-0.80	-0.80	-0.81	0.30	0.67	0.74	0.78	0.79	0.79	0.79	0.79
1.0	0	-0.80	-0.80	-0.80	-0.81	0.12	0.63	0.73	0.78	0.79	0.79	0.79	0.79
0.5	0	-0.79	-0.80	-0.80	-0.80	0.15	0.66	0.75	0.78	0.79	0.79	0.79	0.79
1.5	150	-0.80	-0.80	-0.80	-0.80	0.19	0.56	0.68	0.75	--	--	--	--
1.0	150	-0.80	-0.80	-0.80	-0.80	0.01	0.49	0.66	0.75	--	--	--	--
0.5	150	-0.79	-0.80	-0.80	-0.80	0.03	0.52	0.68	0.75	--	--	--	--

*Top table includes correlations with complete experimental dataset; bottom table includes correlations for mutants of interfacial (buried) residues only (non-interfacial mutants D54A and E80A were excluded).

Table B.5 Dependence of free energy correlation on grid resolution (Im9-E9 DNase).

Grid Res (Å)	Ion Conc (mM)	ΔΔG (Solvation)				ΔG (Solution)				ΔG (Coulombic)			
		$\varepsilon_P=2$	$\varepsilon_P=10$	$\varepsilon_P=20$	$\varepsilon_P=40$	$\varepsilon_P=2$	$\varepsilon_P=10$	$\varepsilon_P=20$	$\varepsilon_P=40$	$\varepsilon_P=2$	$\varepsilon_P=10$	$\varepsilon_P=20$	$\varepsilon_P=40$
1.5	0	-0.72	-0.70	-0.68	-0.66	0.84	0.86	0.83	0.78	0.73	0.73	0.73	0.73
1.0	0	-0.74	-0.71	-0.70	-0.67	0.10	0.79	0.80	0.77	0.73	0.73	0.73	0.73
0.5	0	-0.74	-0.71	-0.69	-0.66	0.11	0.81	0.81	0.78	0.73	0.73	0.73	0.73
1.5	150	-0.72	-0.68	-0.65	-0.59	0.89	0.96	0.96	0.95	--	--	--	--
1.0	150	-0.74	-0.70	-0.66	-0.60	-0.62	0.89	0.93	0.94	--	--	--	--
0.5	150	-0.74	-0.70	-0.66	-0.59	-0.49	0.92	0.94	0.95	--	--	--	--

Grid Res (Å)	Ion Conc (mM)	ΔΔG (Solvation)				ΔG (Solution)				ΔG (Coulombic)			
		$\varepsilon_P=2$	$\varepsilon_P=10$	$\varepsilon_P=20$	$\varepsilon_P=40$	$\varepsilon_P=2$	$\varepsilon_P=10$	$\varepsilon_P=20$	$\varepsilon_P=40$	$\varepsilon_P=2$	$\varepsilon_P=10$	$\varepsilon_P=20$	$\varepsilon_P=40$
1.5	0	-0.56	-0.55	-0.55	-0.54	0.52	0.58	0.58	0.57	0.56	0.56	0.56	0.56
1.0	0	-0.56	-0.55	-0.55	-0.54	0.53	0.58	0.58	0.57	0.56	0.56	0.56	0.56
0.5	0	-0.56	-0.55	-0.55	-0.54	0.49	0.57	0.58	0.57	0.56	0.56	0.56	0.56
1.5	150	-0.56	-0.55	-0.54	-0.53	0.46	0.62	0.65	0.66	--	--	--	--
1.0	150	-0.56	-0.55	-0.54	-0.53	0.49	0.62	0.65	0.66	--	--	--	--
0.5	150	-0.56	-0.55	-0.54	-0.53	0.37	0.59	0.64	0.66	--	--	--	--

*Top table includes correlations with complete experimental dataset; bottom table includes correlations for mutants of non-interfacial (solvent exposed) residues only (interfacial mutants E30A and E41A were excluded).

Table B.6 Dependence of free energy correlation on grid resolution (RNase Inhibitor-Angiogenin).

Grid Res (Å)	Ion Conc (mM)	ΔΔG (Solvation)				ΔG (Solution)				ΔG (Coulombic)			
		$\varepsilon_P=2$	$\varepsilon_P=10$	$\varepsilon_P=20$	$\varepsilon_P=40$	$\varepsilon_P=2$	$\varepsilon_P=10$	$\varepsilon_P=20$	$\varepsilon_P=40$	$\varepsilon_P=2$	$\varepsilon_P=10$	$\varepsilon_P=20$	$\varepsilon_P=40$
1.5	0	-0.48	-0.48	-0.48	-0.48	0.35	0.44	0.47	0.48	0.48	0.48	0.48	0.48
1.0	0	-0.49	-0.49	-0.49	-0.48	0.20	0.42	0.46	0.48	0.48	0.48	0.48	0.48
0.5	0	-0.49	-0.48	-0.48	-0.47	0.29	0.44	0.47	0.48	0.48	0.48	0.48	0.48
1.5	150	-0.48	-0.48	-0.48	-0.46	0.25	0.34	0.41	0.47	--	--	--	--
1.0	150	-0.49	-0.49	-0.48	-0.47	0.05	0.28	0.38	0.46	--	--	--	--
0.5	150	-0.49	-0.48	-0.48	-0.47	0.16	0.32	0.39	0.45	--	--	--	--

Grid Res (Å)	Ion Conc (mM)	ΔΔG (Solvation)				ΔG (Solution)				ΔG (Coulombic)			
		$\varepsilon_P=2$	$\varepsilon_P=10$	$\varepsilon_P=20$	$\varepsilon_P=40$	$\varepsilon_P=2$	$\varepsilon_P=10$	$\varepsilon_P=20$	$\varepsilon_P=40$	$\varepsilon_P=2$	$\varepsilon_P=10$	$\varepsilon_P=20$	$\varepsilon_P=40$
1.5	0	-0.60	-0.60	-0.61	-0.61	0.42	0.53	0.56	0.58	0.59	0.59	0.59	0.59
1.0	0	-0.60	-0.60	-0.60	-0.60	0.34	0.52	0.56	0.58	0.59	0.59	0.59	0.59
0.5	0	-0.60	-0.60	-0.60	-0.60	0.38	0.53	0.57	0.59	0.59	0.59	0.59	0.59
1.5	150	-0.60	-0.60	-0.61	-0.61	0.29	0.35	0.41	0.48	--	--	--	--
1.0	150	-0.60	-0.60	-0.61	-0.61	0.15	0.32	0.40	0.48	--	--	--	--
0.5	150	-0.60	-0.60	-0.61	-0.61	0.23	0.34	0.40	0.46	--	--	--	--

*Top table includes correlations with complete experimental dataset; bottom table includes correlations for mutants of interfacial (buried) residues only (non-interfacial mutants E287A, R33A, R66A, and R70A were excluded).

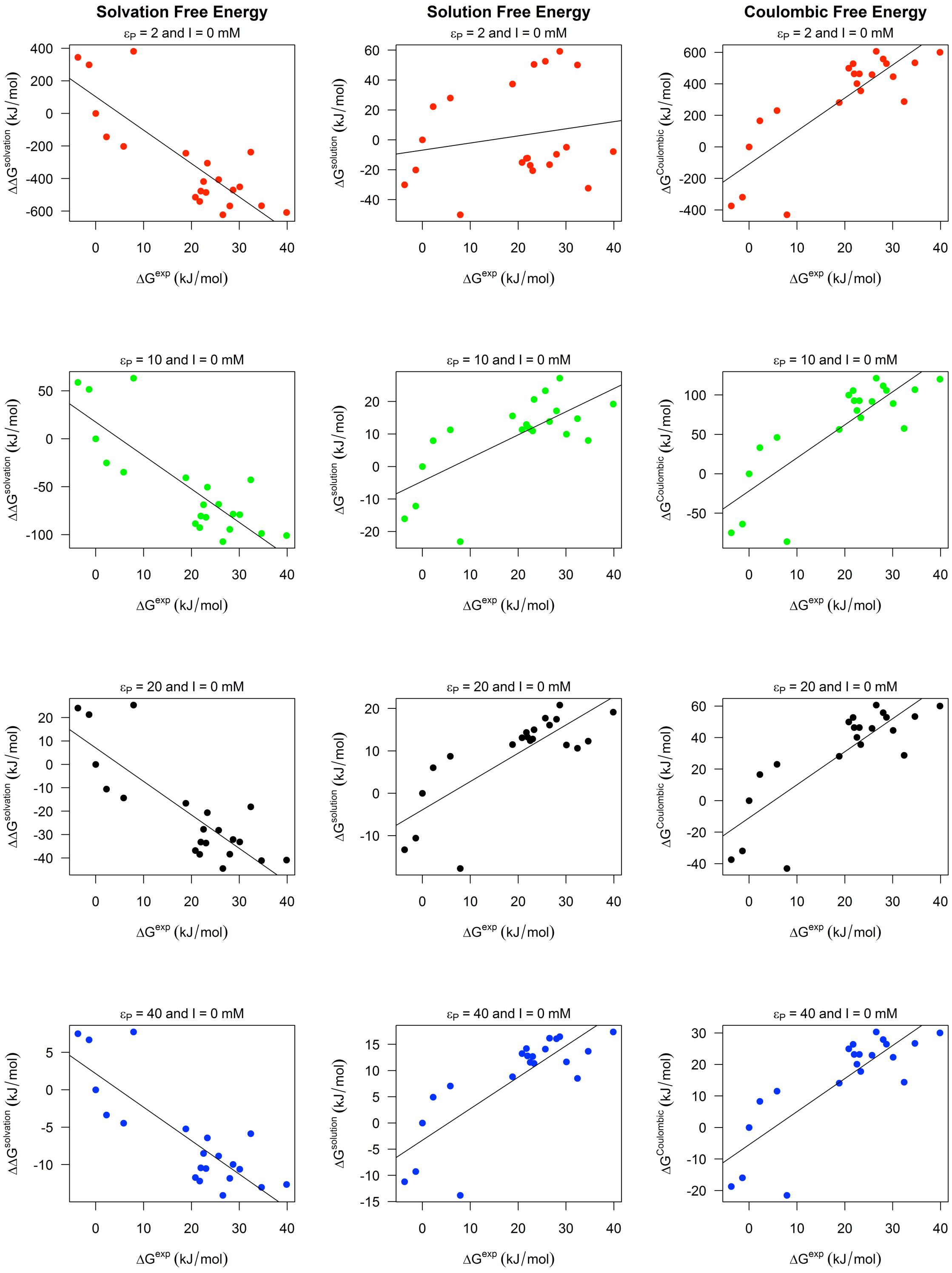


Figure B.1 Scatter plots of calculated and experimental data for Barnase-Barstar (0 mM).

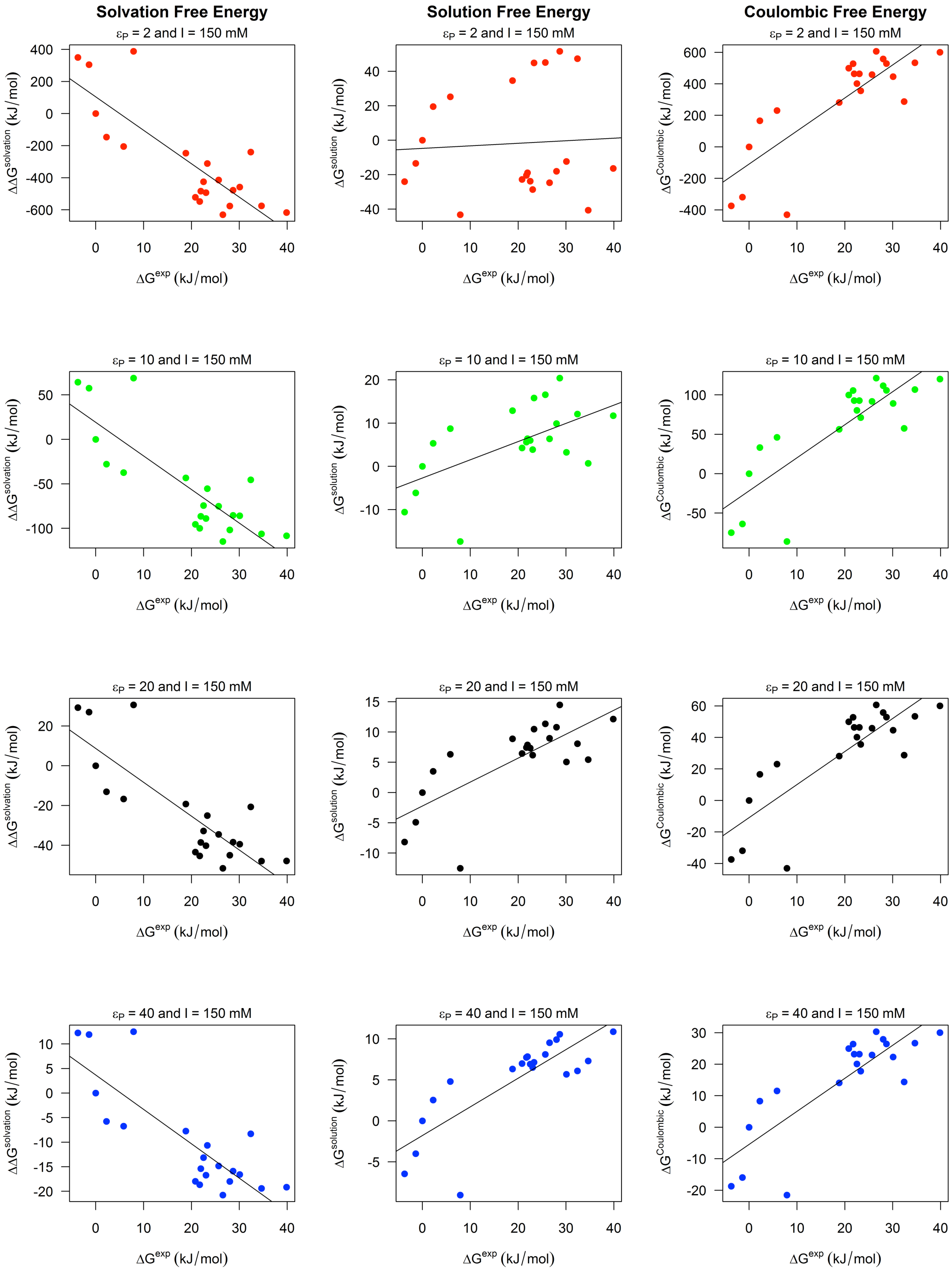


Figure B.2 Scatter plots of calculated and experimental data for Barnase-Barstar (150 mM).

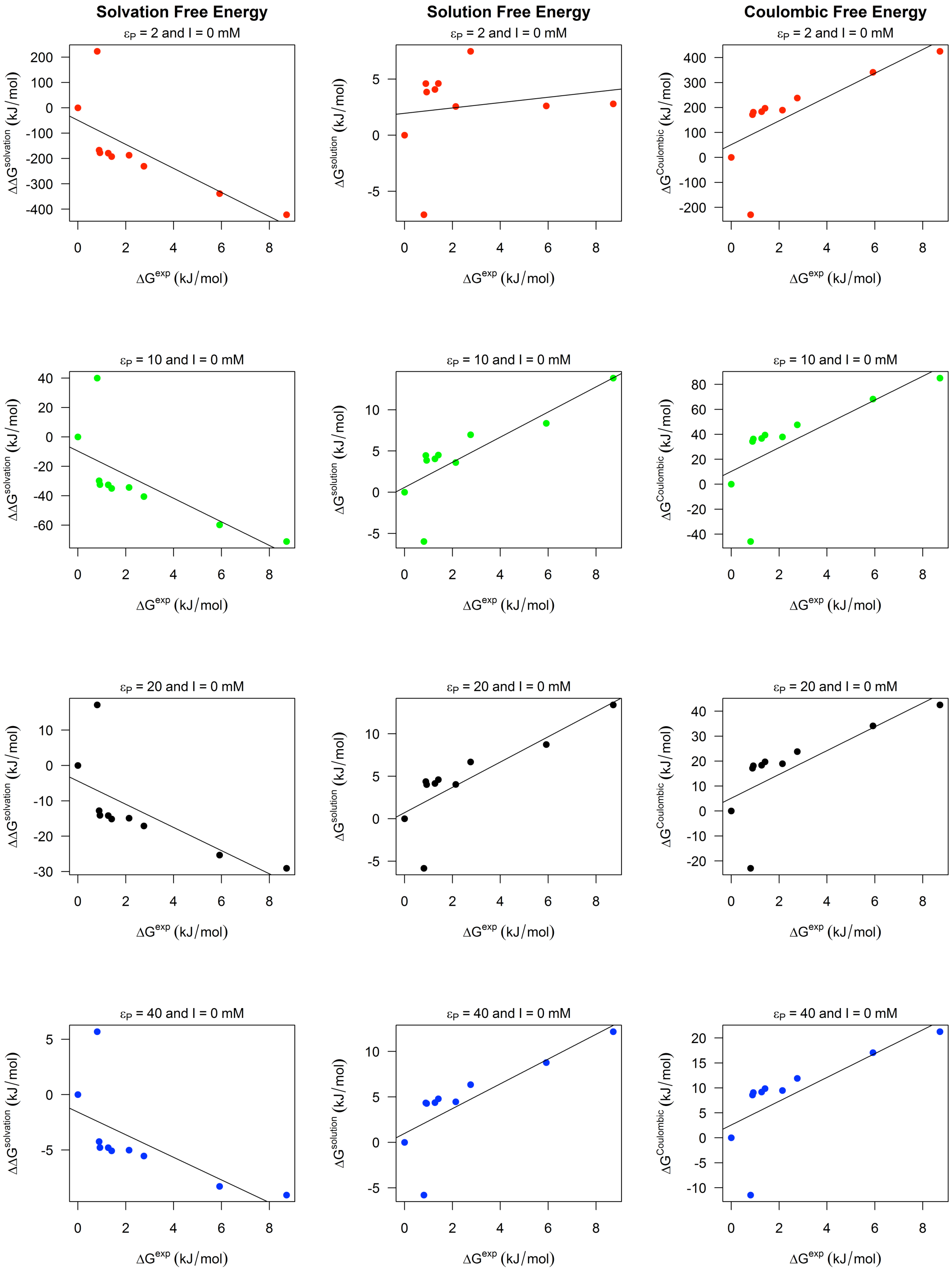


Figure B.3 Scatter plots of calculated and experimental data for Im9-E9 DNase (0 mM).

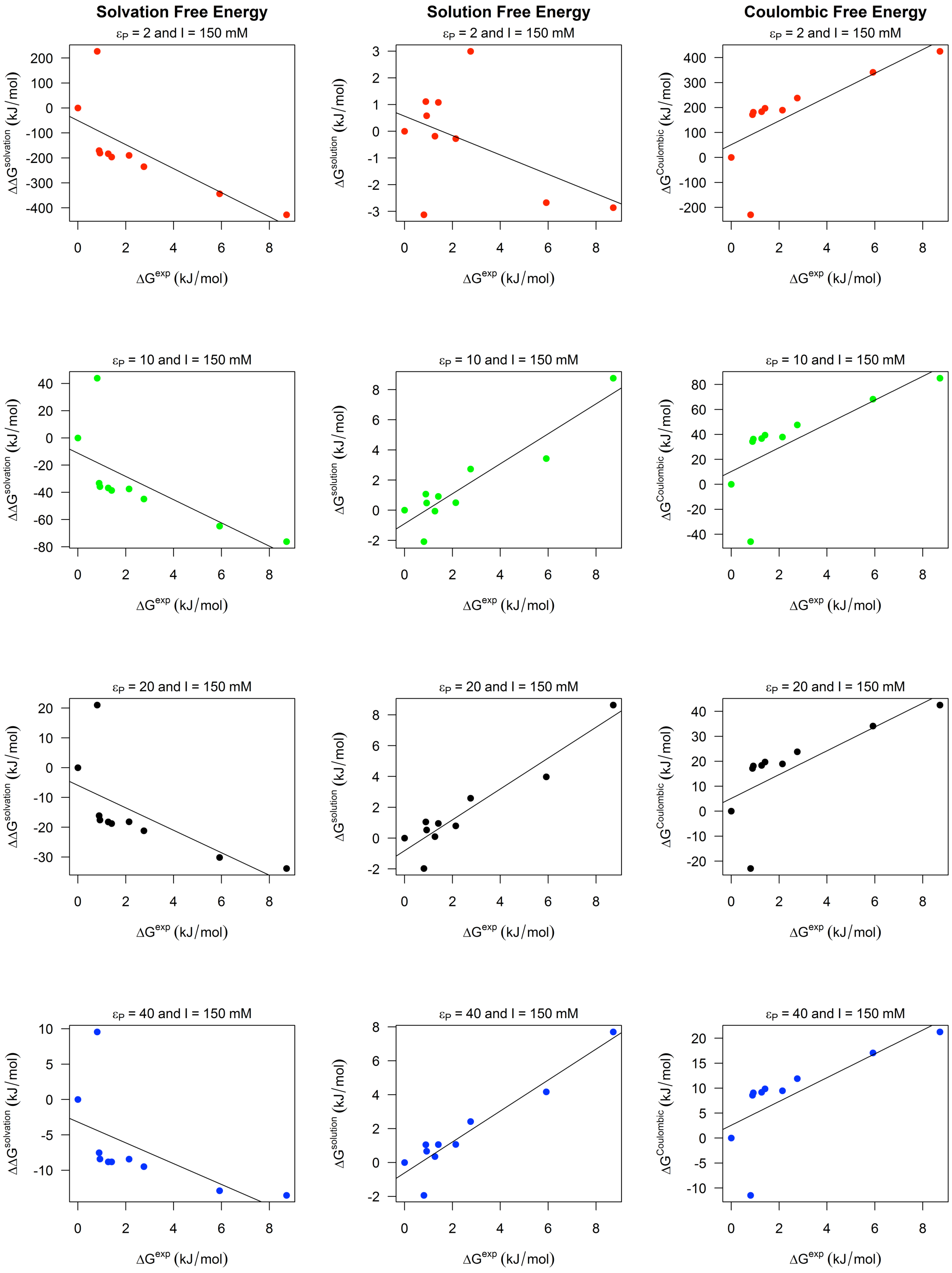


Figure B.4 Scatter plots of calculated and experimental data for Im9-E9 DNase (150 mM).

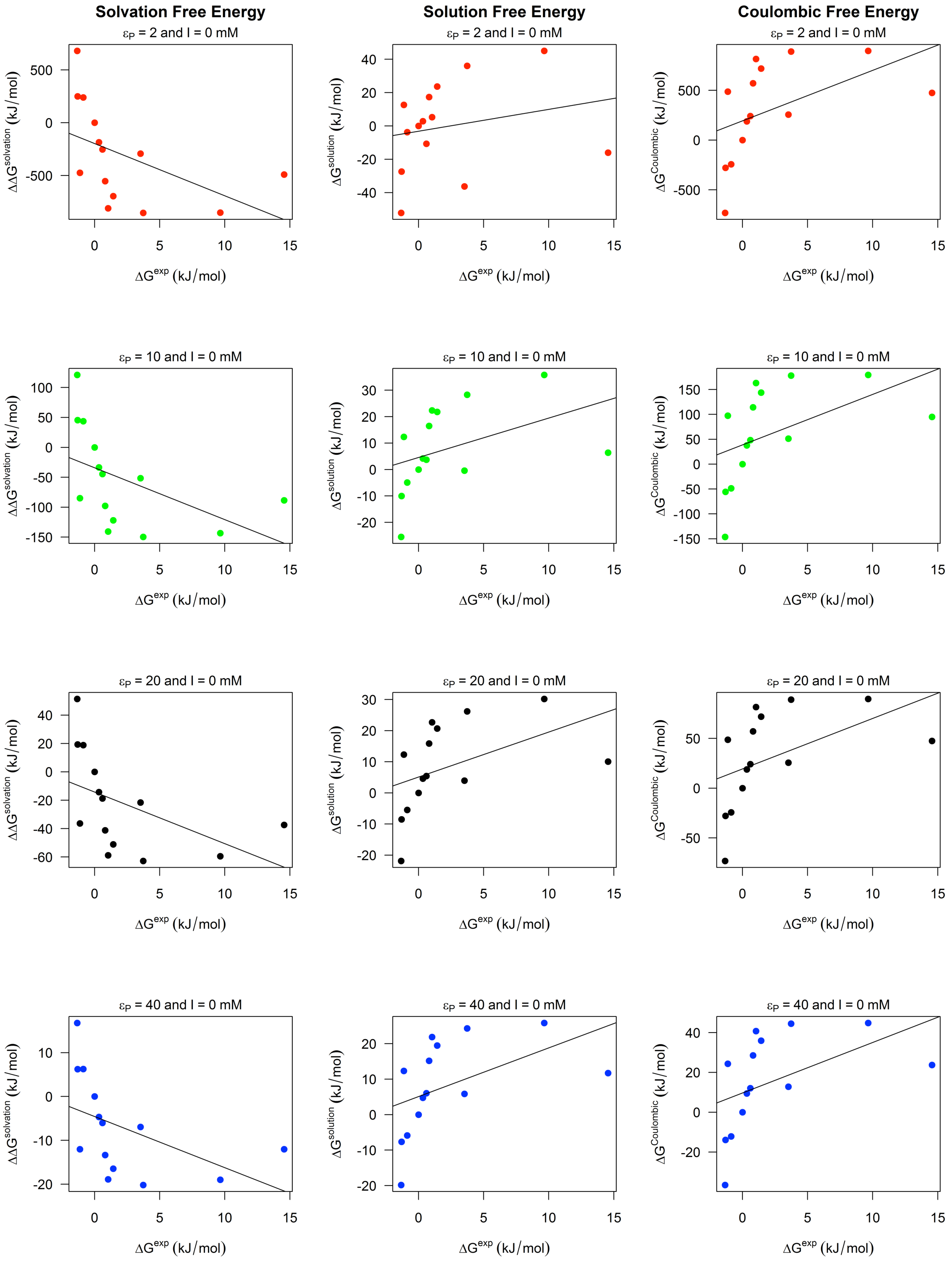


Figure B.5 Scatter plots of calculated and experimental data for RNase Inhibitor-Angiogenin (0 mM).

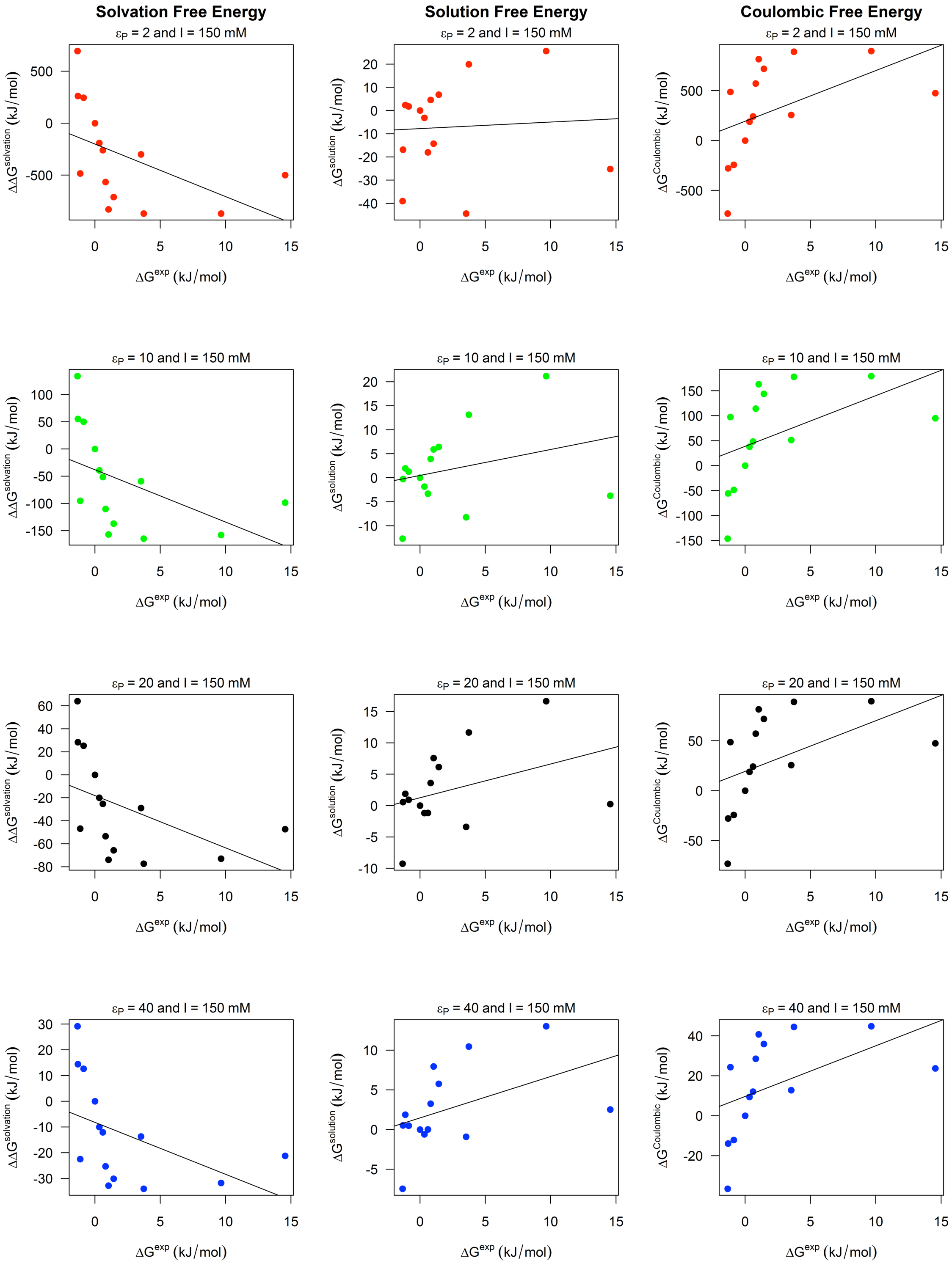


Figure B.6 Scatter plots of calculated and experimental data for RNase Inhibitor-Angiogenin (150 mM).

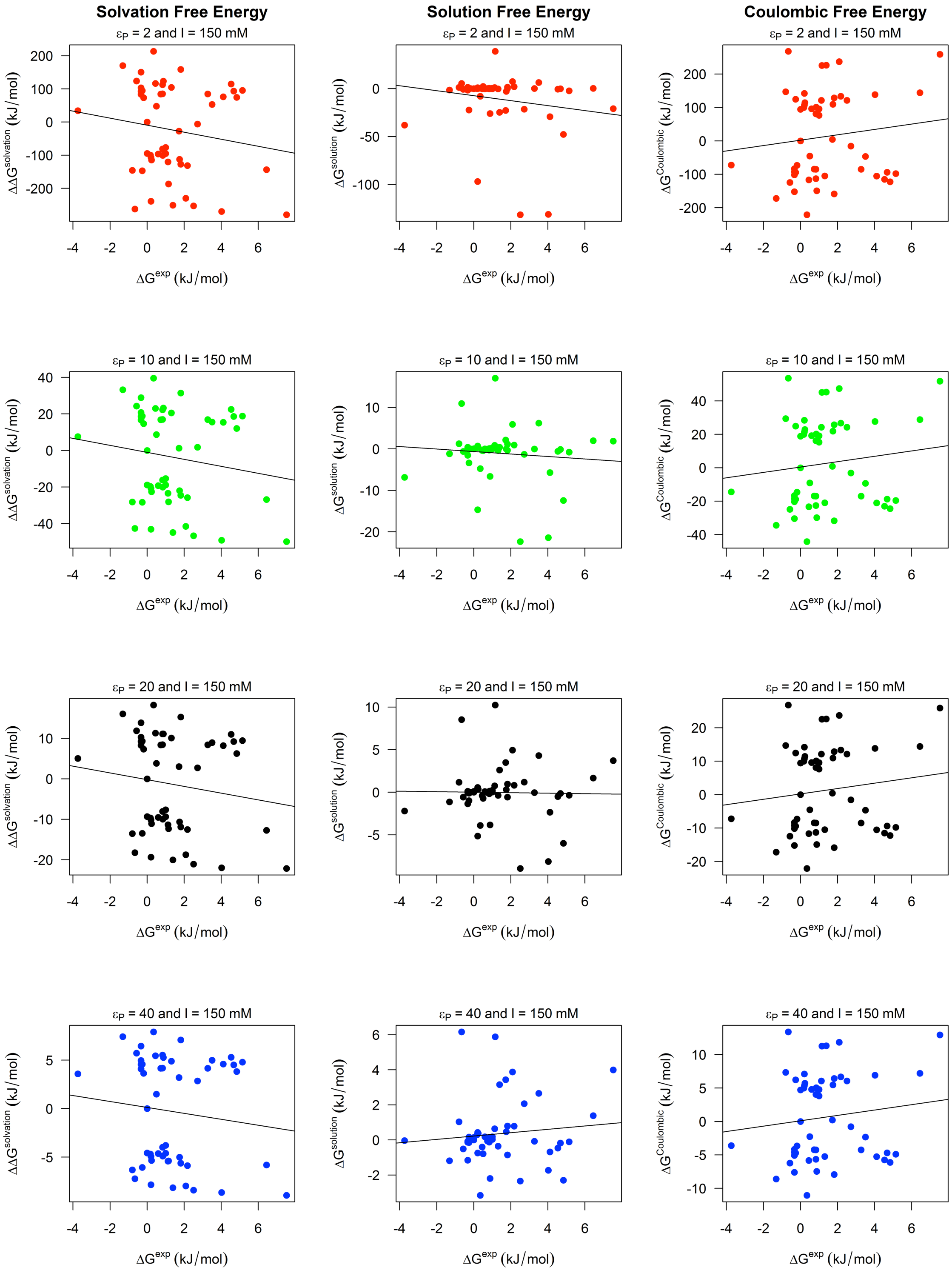


Figure B.7 Scatter plots of calculated and experimental data for hGH-hGHbp (0 mM).

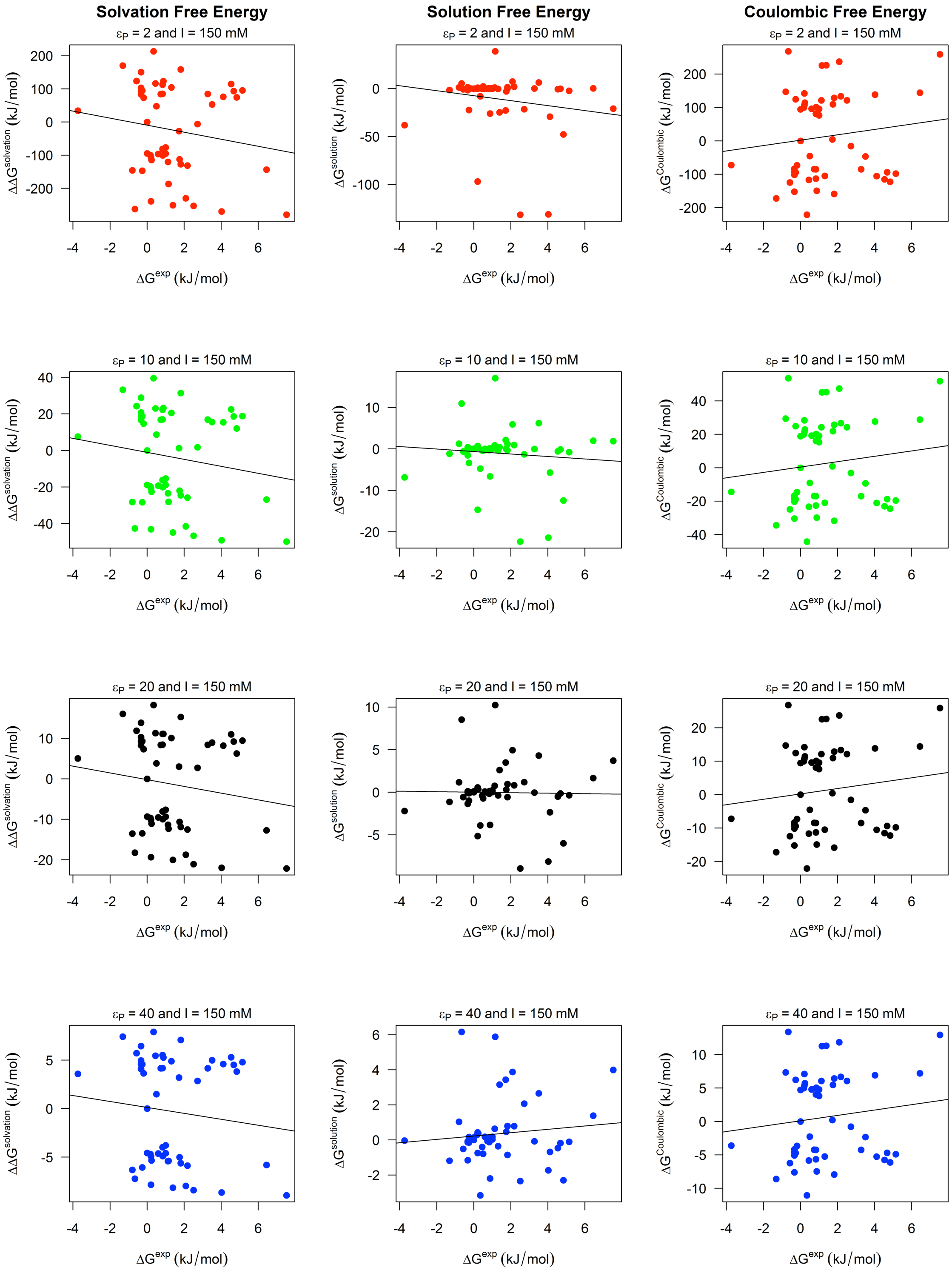


Figure B.8 Scatter plots of calculated and experimental data for hGH-hGHbp (150 mM).

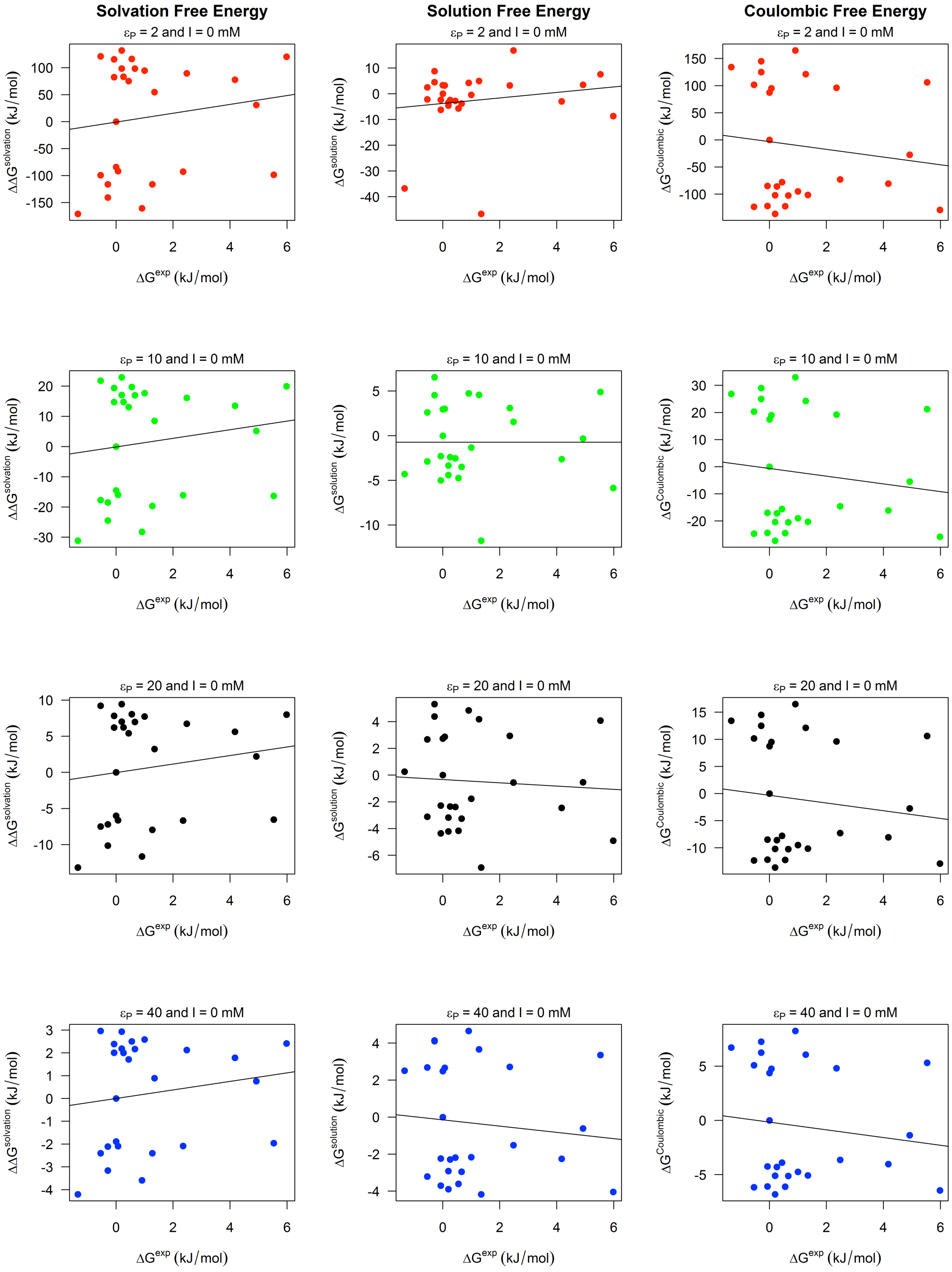


Figure B.9 Scatter plots of calculated and experimental data for gp120-CD4 (0 mM).

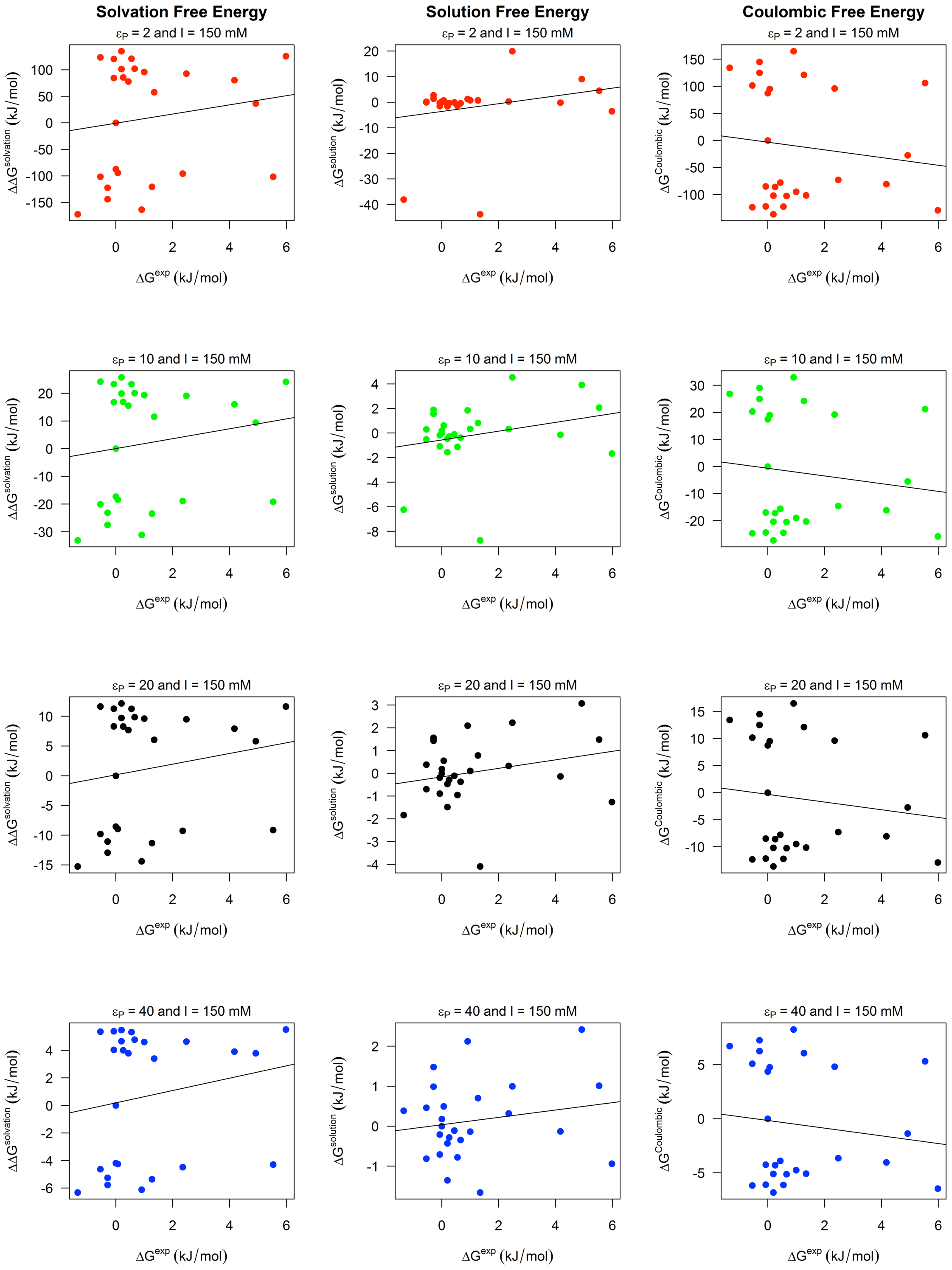


Figure B.10 Scatter plots of calculated and experimental data for gp120-CD4 (150 mM).

APPENDIX C: SUPPORTING MATERIAL FOR CHAPTER 3

Table C.1 Grid dimensions and lengths for electrostatics calculations.

Protein	Grid Dimensions	Coarse Grid Length	Fine Grid Length
C3d	161 × 161 × 161	162.4 × 203.3 × 163.0	74.1 × 87.8 × 74.3
Efb-C	129 × 161 × 97	139.0 × 159.7 × 86.9	66.3 × 73.2 × 49.0
C3d-Efb-C	161 × 161 × 161	177.9 × 203.3 × 183.1	79.3 × 87.8 × 81.0
C3d	161 × 161 × 161	158.1 × 193.7 × 161.9	72.7 × 84.6 × 74.0
Ecb	97 × 193 × 97	104.1 × 215.7 × 103.5	54.7 × 91.9 × 54.5
C3d-Ecb	161 × 193 × 161	170.9 × 215.7 × 161.9	77.0 × 91.9 × 74.0

Table C.2 Intramolecular interactions within C3d (from C3d-Efb-C complex). Unfavorable interactions are shown in bold.

Interaction	Distance (Å)	Interaction	Distance (Å)
K1215-D1216	3.2	**R1134-K1225**	**6.2**
K1084-E1153	3.3	D996-K1036	6.4
K1113-D1115	3.3	**E1120-D1121**	**6.4**
K1244-D1285	3.5	**D996-E998**	**6.4**
K1111-D1121	3.5	E1012-R1072	6.5
K1139-D1140	3.6	K1284-D1285	6.5
E998-K1001	3.6	E1138-R1185	6.5
D1174-R1201	3.6	**D1029-E1030**	**6.6**
E1047-K1050	3.7	E1040-K1041	6.7
E1177-R1201	3.7	**R999-R1260**	**6.9**
E1110-K1111	3.8	K1111-E1120	7.0
E1128-R1134	3.8	**E1110-D1121**	**7.0**
D1029-R1042	3.9	D1115-K1171	7.2
R1219-E1221	4.2	K1215-E1221	7.3
D1216-R1219	4.4	**E1032-E1035**	**7.3**
R1185-D1222	4.5	E998-K1041	7.3
K1001-K1041	**4.8**	D996-R999	7.4
D1245-D1247	**5.0**	R1254-E1258	7.6
K1215-R1219	**5.1**	R993-E1258	7.6
E1159-E1160	**5.3**	E1177-K1203	7.6
E1030-K1284	5.4	E1258-R1260	7.7
D1174-E1177	**5.6**	R1060-D1096	7.8
K1155-E1159	5.6	**D1216-E1221**	**8.0**

Table C.3 Intramolecular interactions within Efb-C (from C3d-Efb-C complex). Unfavorable interactions are shown in bold.

Interaction	Distance (Å)
E120-K123	3.2
K107-E143	3.2
E122-R154	3.8
D102-K106	4.9
E153-D156	**6.8**
R119-R154	**7.3**
K106-K110	**7.5**
K147-R154	**7.6**
R119-E120	8.0

Table C.4 Intramolecular interactions within C3d (from C3d-Ecb complex). Unfavorable interactions are shown in bold.

Interaction	Distance (Å)	Interaction	Distance (Å)
K1084-E1153	3.0	**D1174-E1177**	**6.3**
K1215-E1221	3.3	**R1134-K1225**	**6.3**
K1111-D1121	3.6	R993-E1153	6.6
R1219-E1221	3.7	E1110-K1111	6.9
E1258-R1260	3.9	E1032-K1284	7.0
K1155-E1159	3.9	E998-R999	7.1
D1029-R1042	4.1	K1215-D1216	7.1
R1185-D1222	4.2	**D1029-E1030**	**7.1**
D1174-R1201	4.3	E1138-R1185	7.1
R993-R1084	**4.5**	K1113-D1115	7.1
E1177-R1201	4.7	R993-E1231	7.2
E1030-K1284	4.8	**E1159-E1160**	**7.5**
E1047-K1050	5.4	K1217-D1247	7.6
K1203-D1245	5.6	E1012-R1072	7.6
E1120-D1121	**5.8**	**D1156-E1159**	**7.6**
K1215-R1219	**5.9**	E1120-K1139	7.7
E1035-K1036	6.0	D1216-K1217	7.9
R1060-D1096	6.3	K1284-D1285	8.0
E1128-R1134	6.3		

Table C.5 Intramolecular interactions within Ecb (from C3d-Ecb complex). Unfavorable interactions are shown in bold.

Interaction	Distance (Å)
E91-K94	3.6
E57-K76	5.2
K60-K76	**5.7**
E57-K60	6.1
E63-R67	6.4
K93-R97	**6.5**
R97-D100	6.5
K93-D100	7.2
E63-K66	7.8
K53-E57	7.9

Clustering and Free Energy of C3d Mutants at 0 mM (from C3d-Efb-C complex)

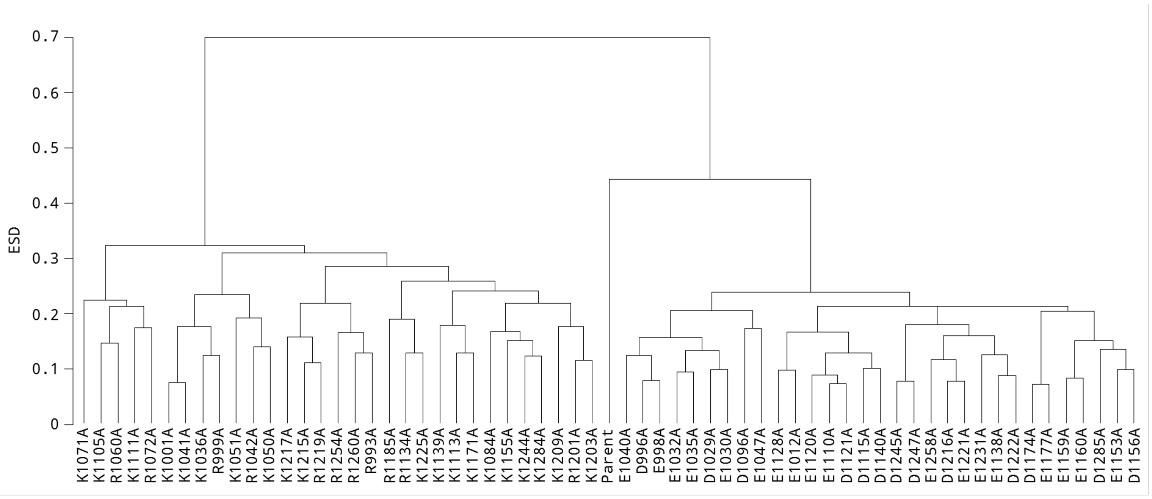


Figure C.1 Clustering of C3d mutants (from C3d-Efb-C complex) at 0 mM ionic strength.

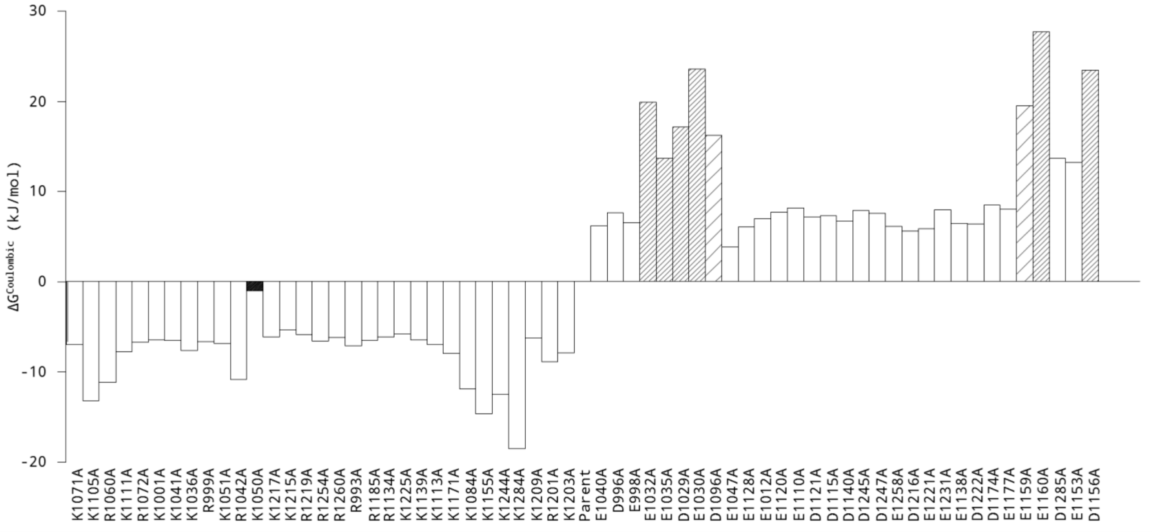


Figure C.2 $\Delta G^{Coulombic}$ free energy values for C3d mutants (from C3d-Efb-C complex) at 0 mM ionic strength. White bars indicate participation in no intermolecular contacts with Ecb within 8 Å, while lightly hashed, densely hashed, and black bars represent mutants of residues participating in interactions within 5-8 Å, 3.5-5 Å, and < 3.5 Å, respectively.

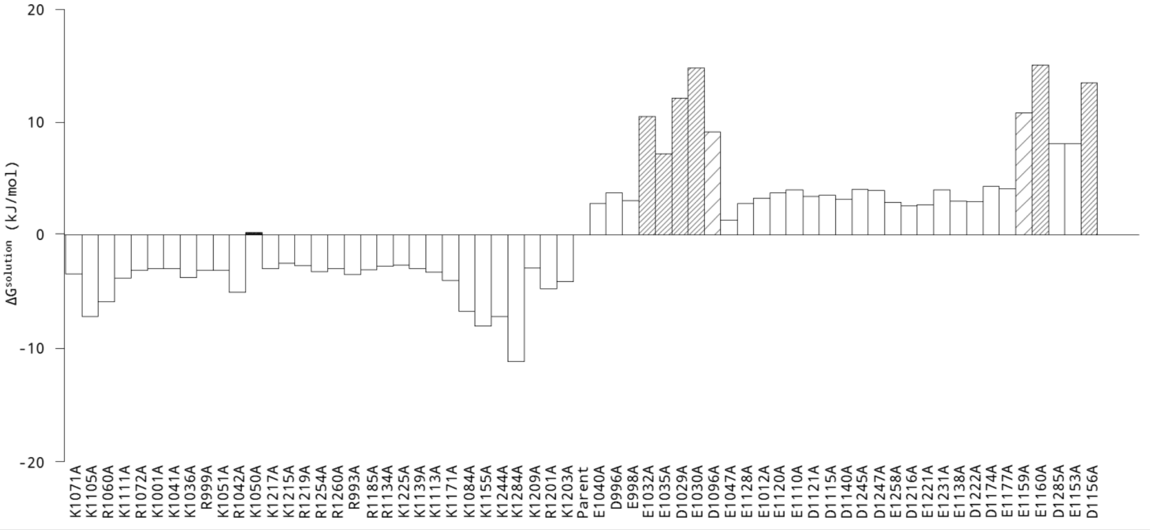


Figure C.3 $\Delta G^{solution}$ free energy values for C3d mutants (from C3d-Efb-C complex) at 0 mM ionic strength. White bars indicate participation in no intermolecular contacts with Ecb within 8 Å, while lightly hashed, densely hashed, and black bars represent mutants of residues participating in interactions within 5-8 Å, 3.5-5 Å, and < 3.5 Å, respectively.

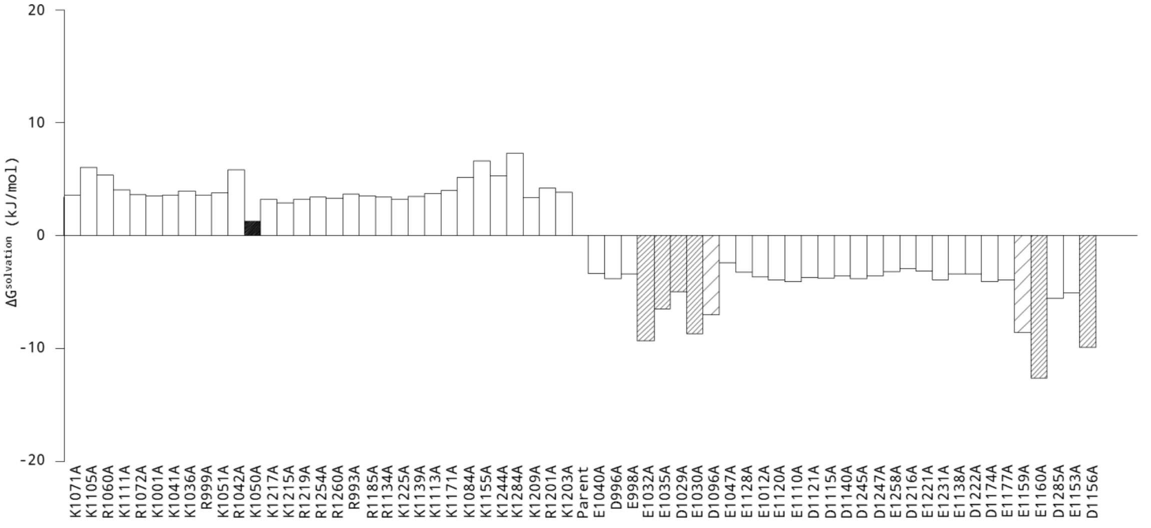


Figure C.4 $\Delta\Delta G^{solvation}$ free energy values for C3d mutants (from C3d-Efb-C complex) at 0 mM ionic strength. White bars indicate participation in no intermolecular contacts with Ecb within 8 Å, while lightly hashed, densely hashed, and black bars represent mutants of residues participating in interactions within 5-8 Å, 3.5-5 Å, and < 3.5 Å, respectively.

Clustering and Free Energy of C3d Mutants at 150 mM (from C3d-Efb-C complex)

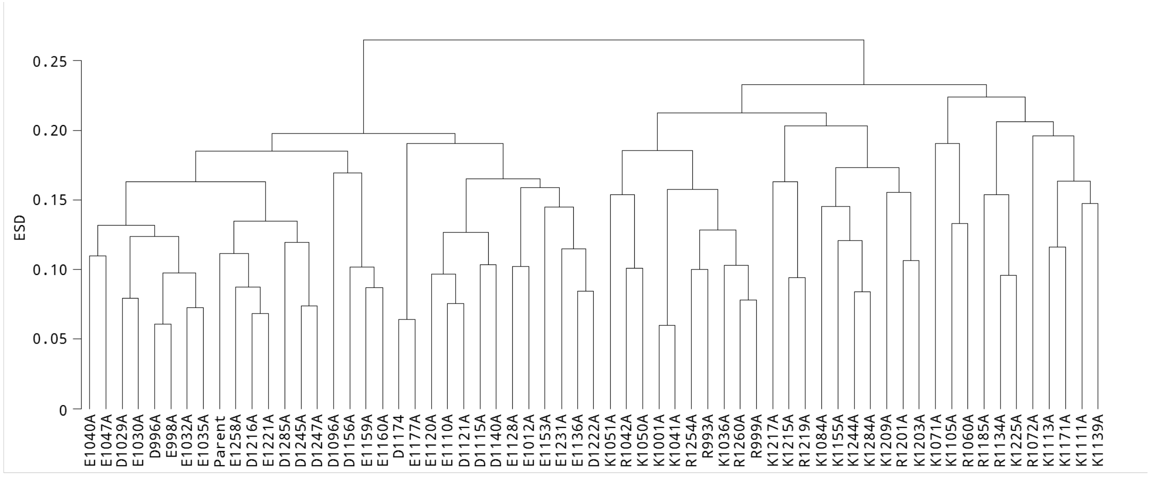


Figure C.5 Clustering of C3d mutants (from C3d-Efb-C complex) at 150 mM ionic strength.

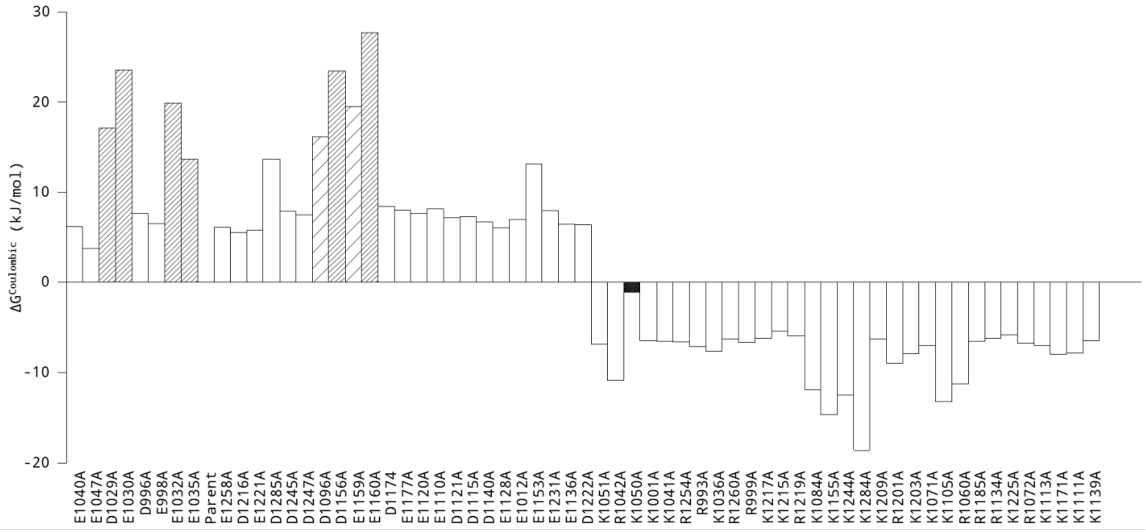


Figure C.6 $\Delta G^{Coulombic}$ free energy values for C3d mutants (from C3d-Efb-C complex) at 150 mM ionic strength. White bars indicate participation in no intermolecular contacts with Ecb within 8 Å, while lightly hashed, densely hashed, and black bars represent mutants of residues participating in interactions within 5-8 Å, 3.5-5 Å, and < 3.5 Å, respectively.

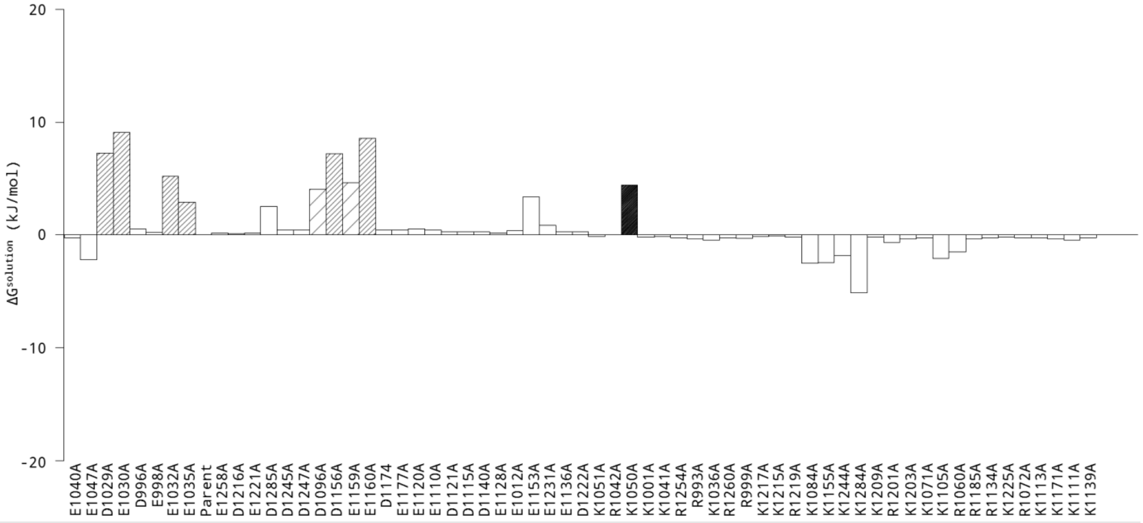


Figure C.7 $\Delta G^{solution}$ free energy values for C3d mutants (from C3d-Efb-C complex) at 150 mM ionic strength. White bars indicate participation in no intermolecular contacts with Ecb within 8 Å, while lightly hashed, densely hashed, and black bars represent mutants of residues participating in interactions within 5-8 Å, 3.5-5 Å, and < 3.5 Å, respectively.

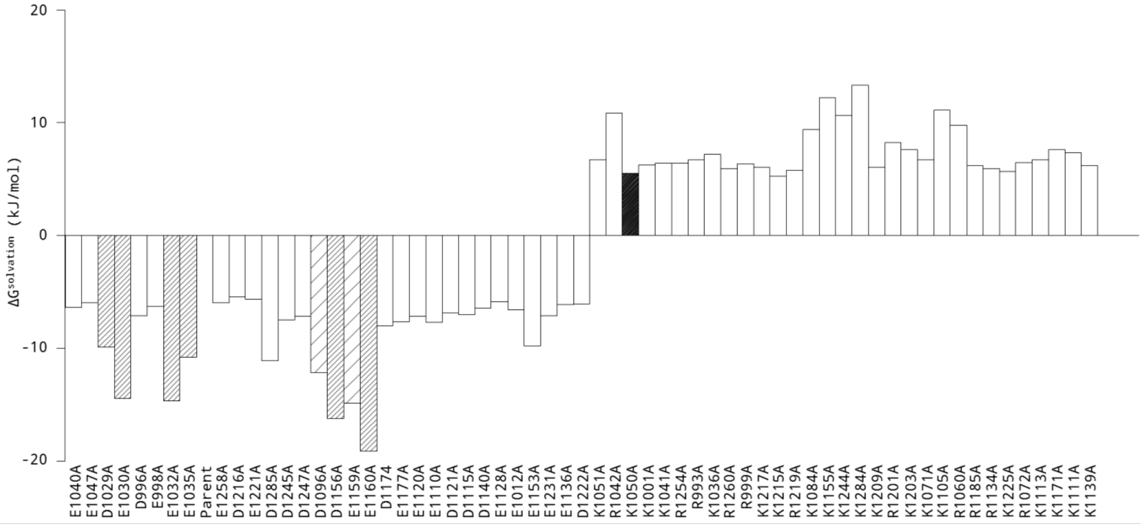


Figure C.8 $\Delta\Delta G^{solvation}$ free energy values for C3d mutants (from C3d-Efb-C complex) at 150 mM ionic strength. White bars indicate participation in no intermolecular contacts with Ecb within 8 Å, while lightly hashed, densely hashed, and black bars represent mutants of residues participating in interactions within 5-8 Å, 3.5-5 Å, and < 3.5 Å, respectively.

Clustering and Free Energy of C3d Mutants at 0 mM (from C3d-Ecb complex)

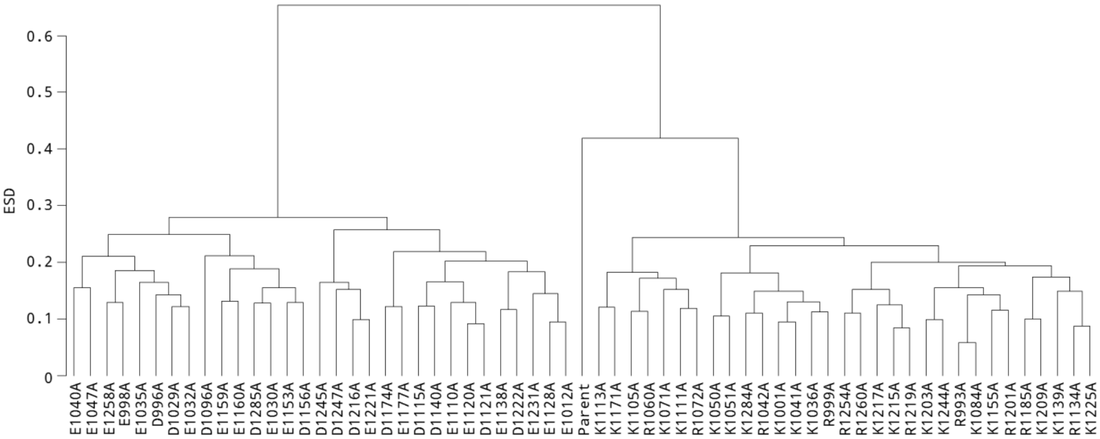


Figure C.9 Clustering of C3d mutants (from C3d-Ecb complex) at 0 mM ionic strength.

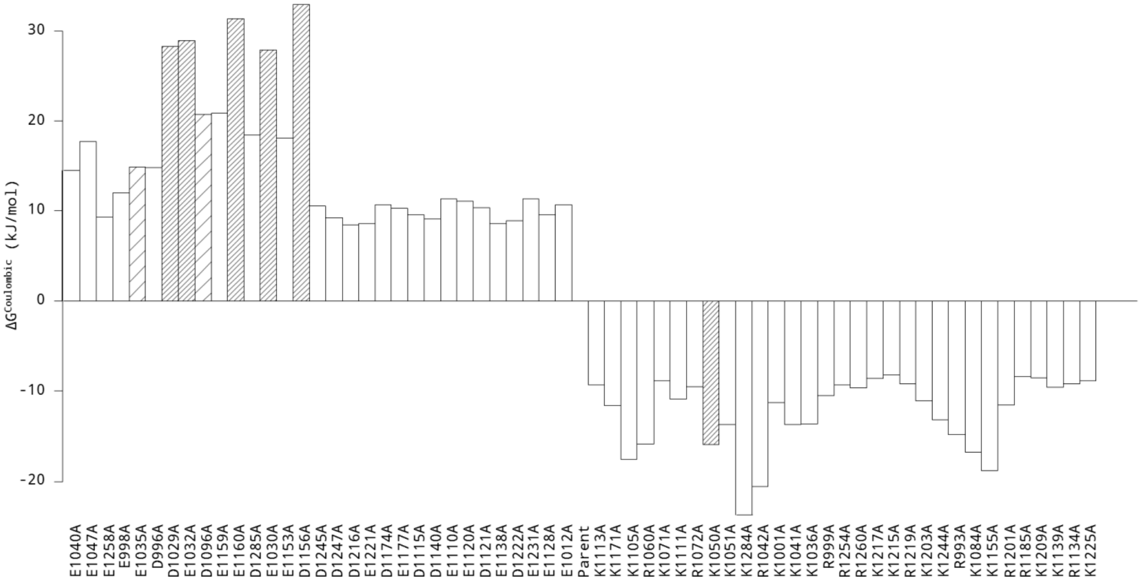


Figure C.10 $\Delta G^{Coulombic}$ free energy values for C3d mutants (from C3d-Ecb complex) at 0 mM ionic strength. White bars indicate participation in no intermolecular contacts with Ecb within 8 Å, while lightly hashed and densely hashed bars represent mutants of residues participating in interactions within 5-8 Å and 3.5-5 Å, respectively.

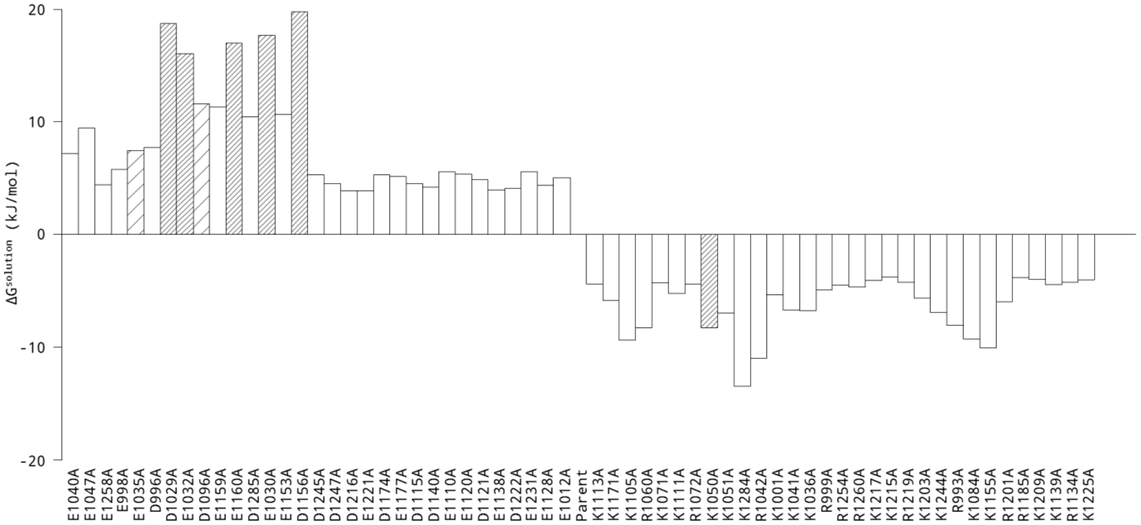


Figure C.11 $\Delta G^{solution}$ free energy values for C3d mutants (from C3d-Ecb complex) at 0 mM ionic strength. White bars indicate participation in no intermolecular contacts with Ecb within 8 Å, while lightly hashed and densely hashed bars represent mutants of residues participating in interactions within 5-8 Å and 3.5-5 Å, respectively.

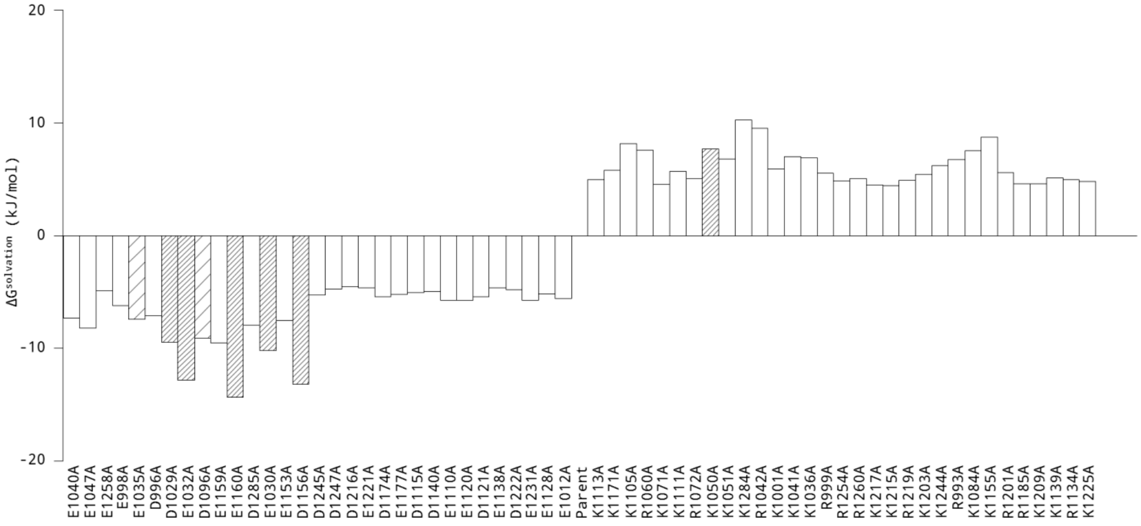


Figure C.12 $\Delta\Delta G^{solvation}$ free energy values for C3d mutants (from C3d-Ecb complex) at 0 mM ionic strength. White bars indicate participation in no intermolecular contacts with Ecb within 8 Å, while lightly hashed and densely hashed bars represent mutants of residues participating in interactions within 5-8 Å and 3.5-5 Å, respectively.

Clustering and Free Energy of C3d Mutants at 150 mM (from C3d-Ecb complex)

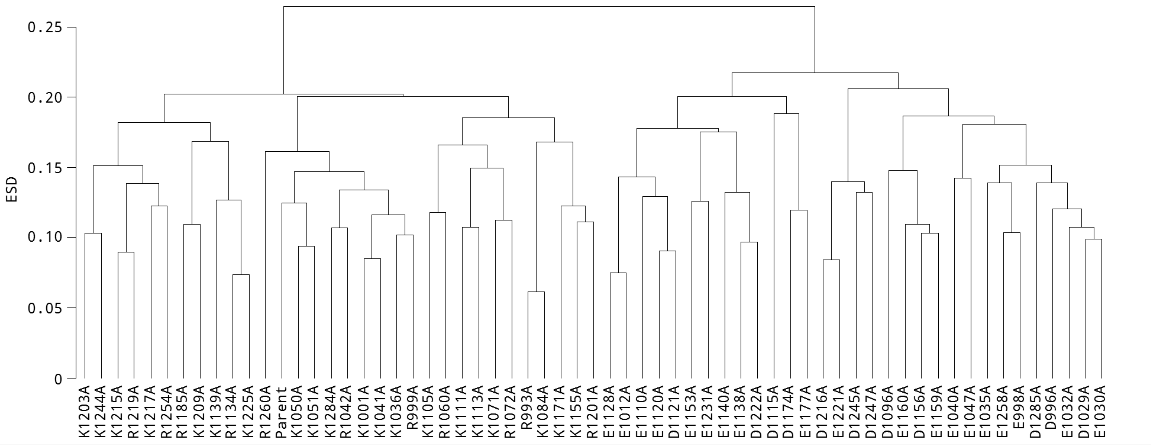


Figure C.13 Clustering of C3d mutants (from C3d-Ecb complex) at 150 mM ionic strength.

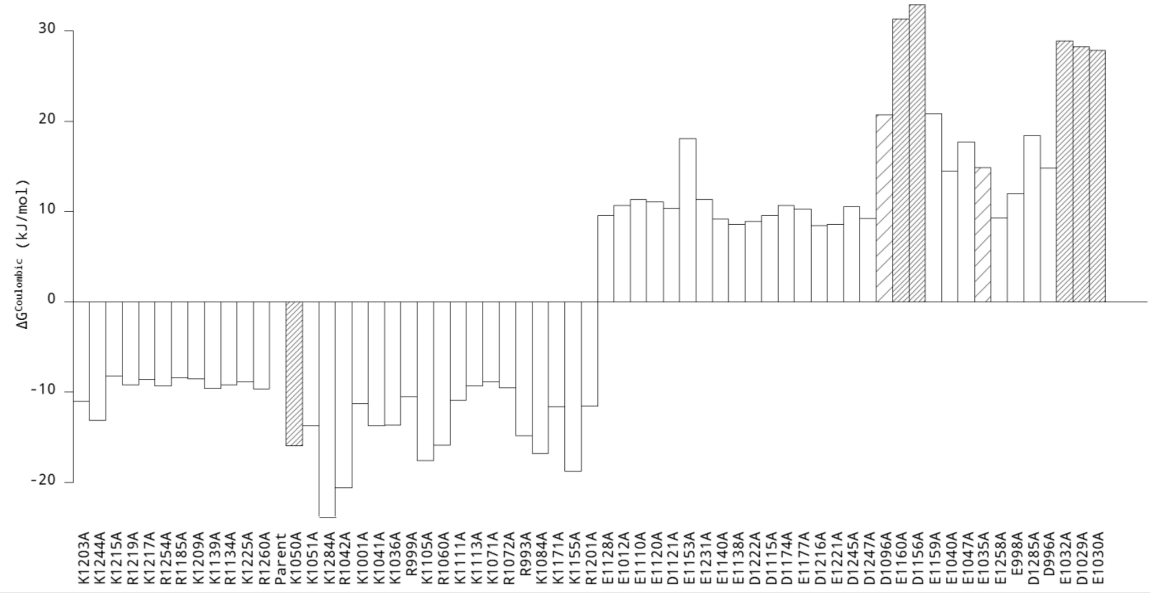


Figure C.14 $\Delta G^{Coulombic}$ free energy values for C3d mutants (from C3d-Ecb complex) at 150 mM ionic strength. White bars indicate participation in no intermolecular contacts with Ecb within 8 Å, while lightly hashed and densely hashed bars represent mutants of residues participating in interactions within 5-8 Å and 3.5-5 Å, respectively.

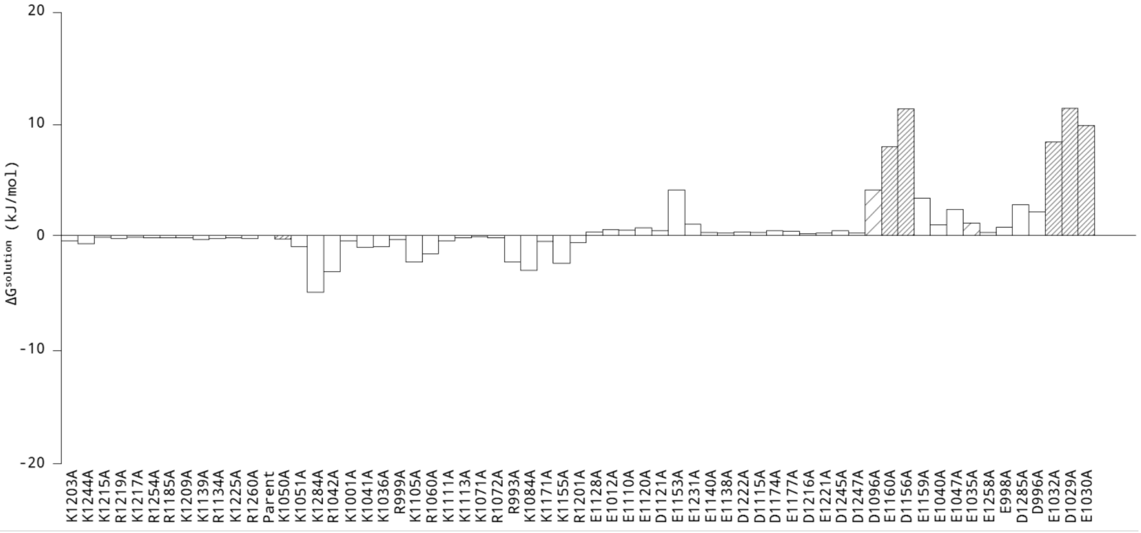


Figure C.15 $\Delta G^{solution}$ free energy values for C3d mutants (from C3d-Ecb complex) at 150 mM ionic strength. White bars indicate participation in no intermolecular contacts with Ecb within 8 Å, while lightly hashed and densely hashed bars represent mutants of residues participating in interactions within 5-8 Å and 3.5-5 Å, respectively.

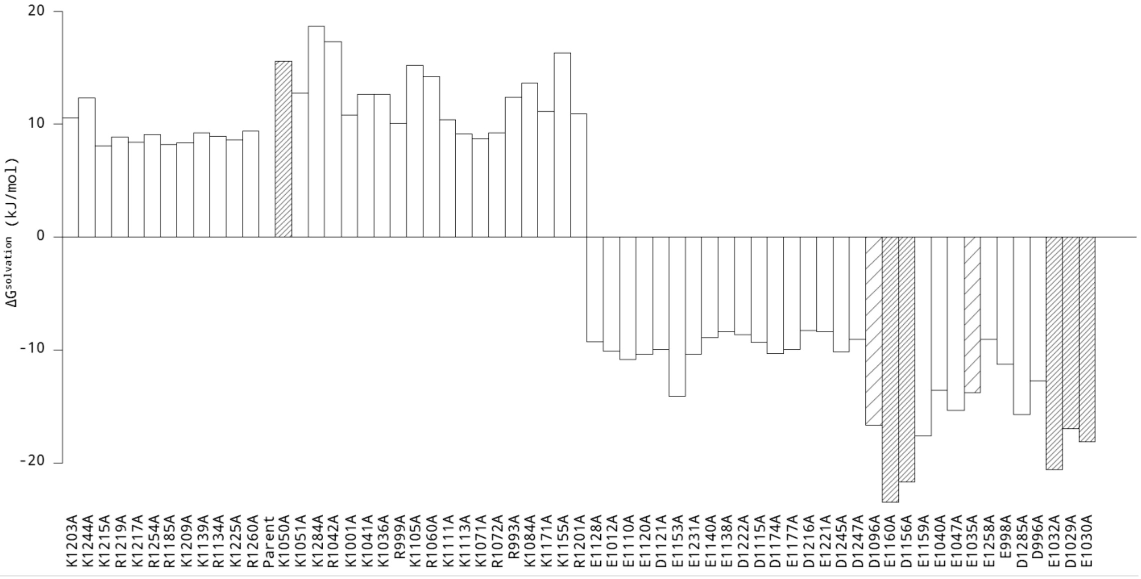


Figure C.16 $\Delta\Delta G^{solvation}$ free energy values for C3d mutants (from C3d-Ecb complex) at 150 mM ionic strength. White bars indicate participation in no intermolecular contacts with Ecb within 8 Å, while lightly hashed and densely hashed bars represent mutants of residues participating in interactions within 5-8 Å and 3.5-5 Å, respectively.

APPENDIX D: SUPPORTING MATERIAL FOR CHAPTER 4

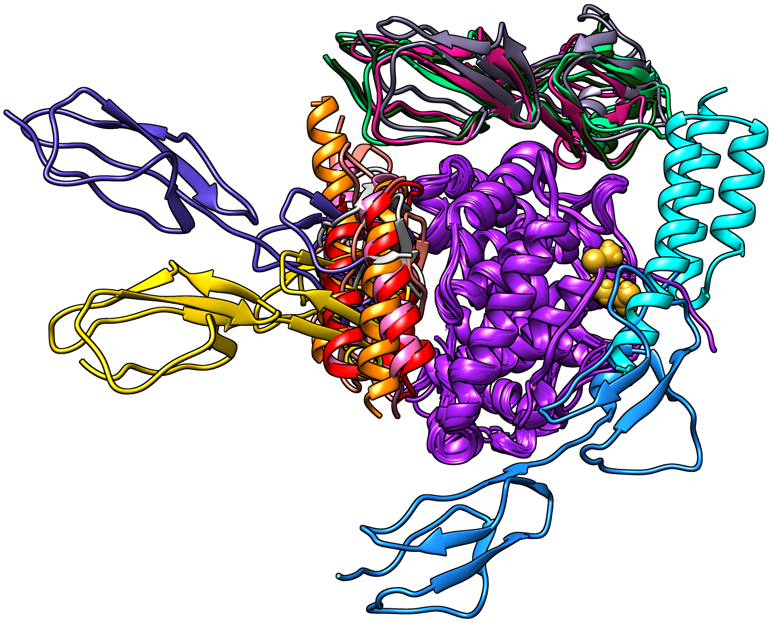


Figure D.1 Structures of known ligands in complex with C3d. This figure shows a ribbon representation of C3d (purple) in complex with Sbi-IV (Complex 1, orange; Complex 2, cyan (1)), Efb-C (pink (2)), Ecb (red (3)), CR2 (salmon (4)), Factor H Domain 4 (light gray (5)), and Factor H Domains 19-20 (Kajander et al. Chain A-C (6), magenta; Kajander et al. Chain B-C, yellow; Morgan et al. Chain A-D (7), blue; Morgan et al. Chain A-E, green; Morgan et al. Chain B-D, indigo; Morgan et al. Chain B-F, dark green; Morgan et al. Chain C-D, light purple). The location of the thioester bond in the convex surface of C3d is shown as gold spheres. The location of the concave surface of C3d is located at the opposite site of the convex surface.

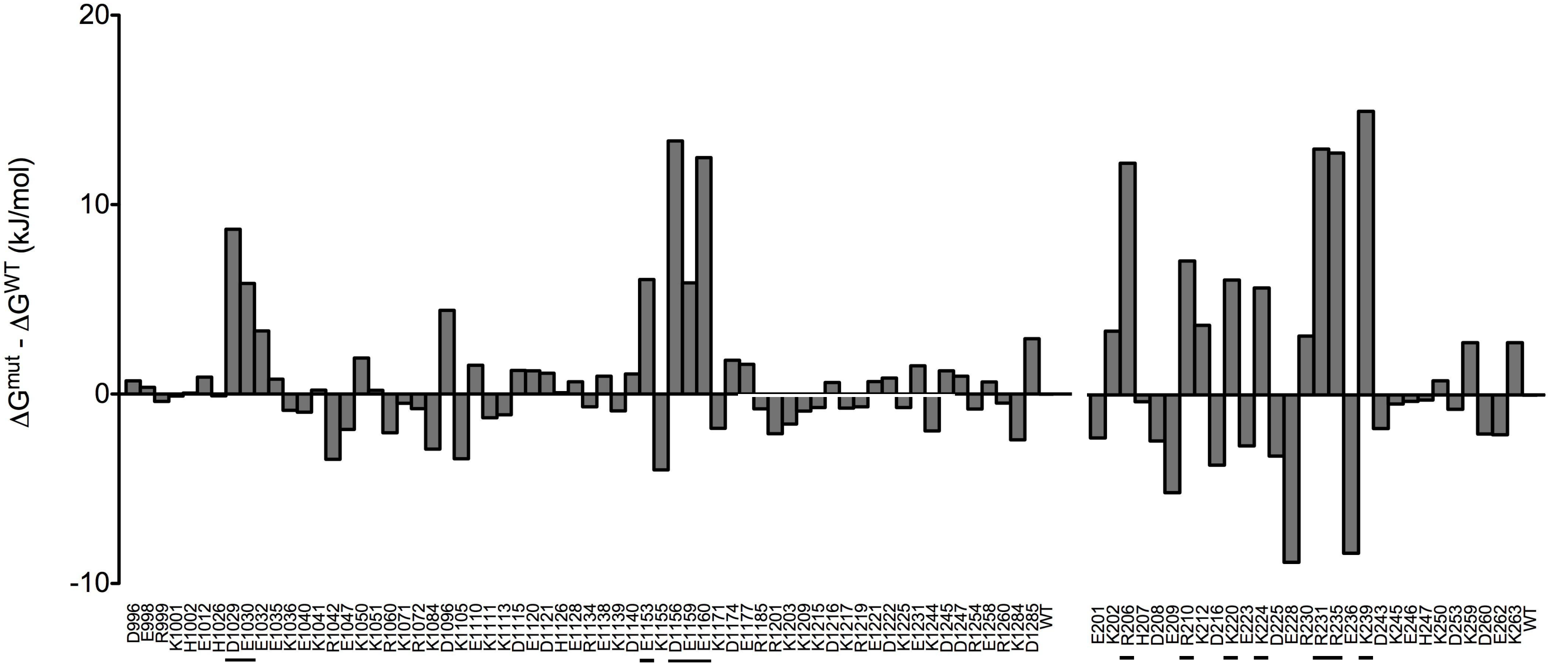


Figure D.2 Electrostatic free energies of C3d-Sbi-IV (Complex 1) association. Free energies were calculated using a thermodynamic cycle including protein association and solvation, using Poisson-Boltzmann electrostatics calculations for Complex 1. Residue labels refer to the corresponding energy of single alanine mutations of those residues, and magnitudes refer to energies relative to the wild-type complex. Only ionizable residues are included. Mutants of C3d are shown on the left, and mutants of Sbi-IV on the right.

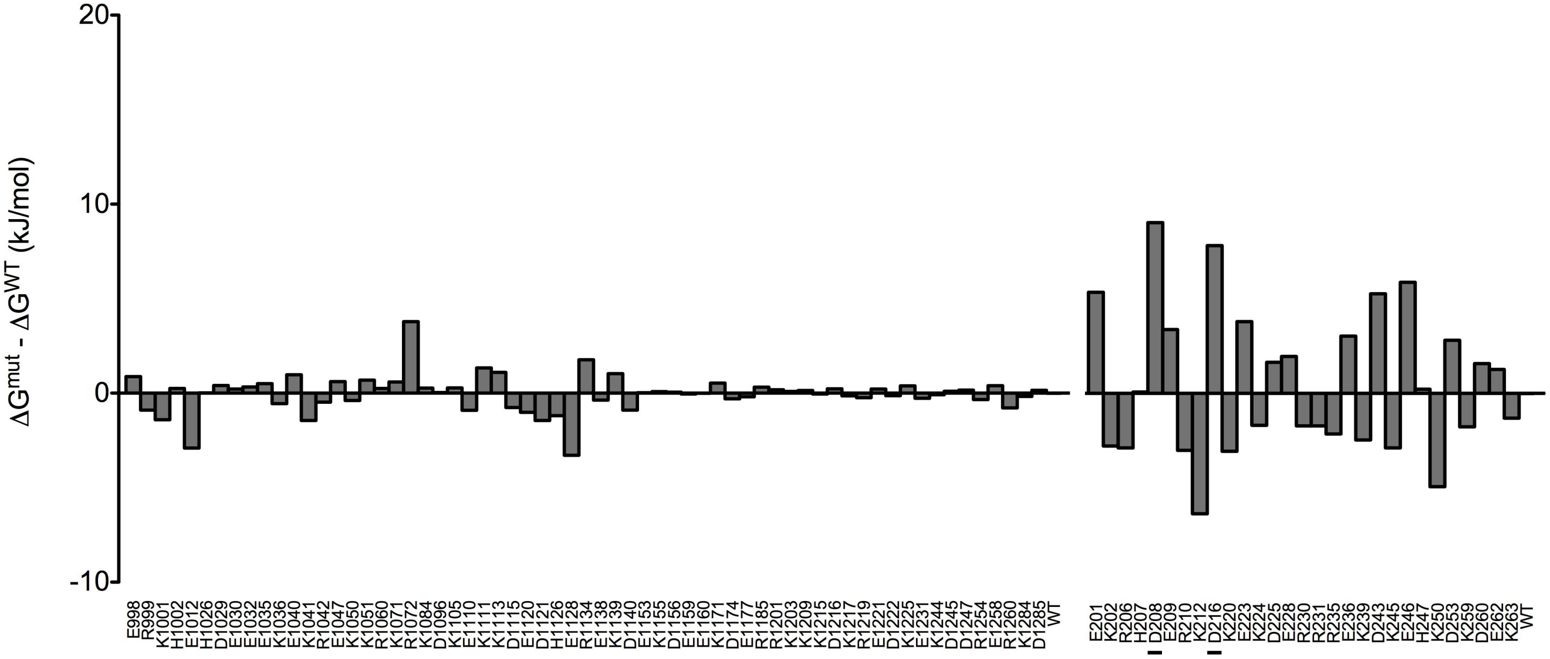


Figure D.3 Electrostatic free energies of C3d-Sbi-IV (Complex 2) association. Free energies were calculated using a thermodynamic cycle including protein association and solvation, using Poisson-Boltzmann electrostatics calculations for Complex 2. Residue labels refer to the corresponding energy of single alanine mutations of those residues, and magnitudes refer to energies relative to the wild-type complex. Only ionizable residues are included. Mutants of C3d are shown on the left, and mutants of Sbi-IV on the right.

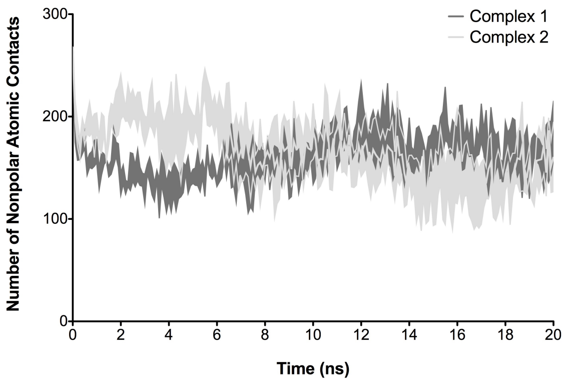


Figure D.4 Number of nonpolar atomic contacts during MD trajectories. The plot shows the mean +/- SEM of the number of nonpolar atomic contacts in Complex 1 (dark gray) and Complex 2 (light gray) of C3d-Sbi-IV during three independent 20 ns MD trajectories.

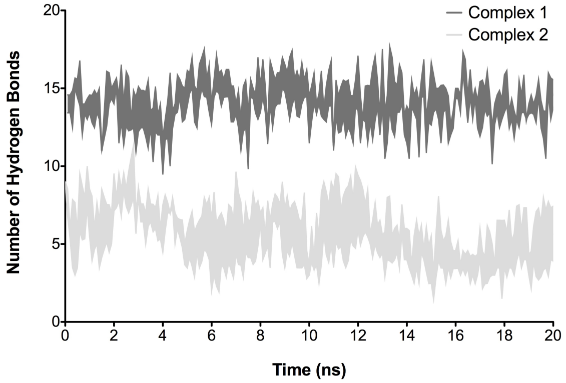


Figure D.5 Number of hydrogen bonds during MD trajectories. The plot shows the mean +/- SEM of the number of hydrogen bonds in Complex 1 (dark gray) and Complex 2 (light gray) of C3d-Sbi-IV during three independent 20 ns MD trajectories.

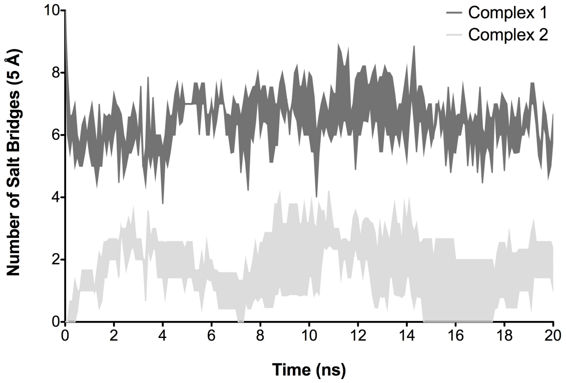


Figure D.6 Number of salt bridges (measured as charge-charge distance of < 5 Å) during MD trajectories. The plot shows the mean +/- SEM of the number of hydrogen bonds in Complex 1 (dark gray) and Complex 2 (light gray) of C3d-Sbi-IV during three independent 20 ns MD trajectories.

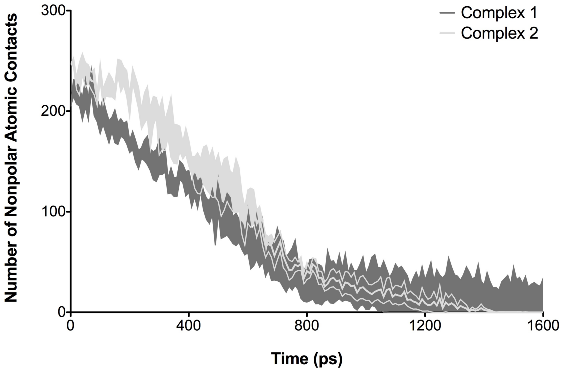


Figure D.7 Number of nonpolar atomic contacts during SMD trajectories. The plot shows the mean +/- SEM of the number of nonpolar atomic contacts in Complex 1 (dark gray) and Complex 2 (light gray) of C3d-Sbi-IV during five independent 1.6 ns SMD trajectories.

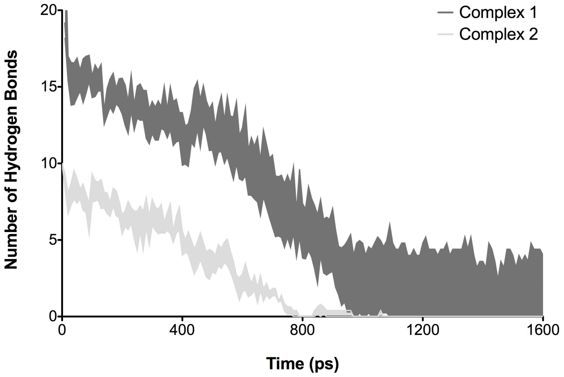


Figure D.8 Number of hydrogen bonds during SMD trajectories. The plot shows the mean +/- SEM of the number of hydrogen bonds in Complex 1 (dark gray) and Complex 2 (light gray) of C3d-Sbi-IV during five independent 1.6 ns SMD trajectories.

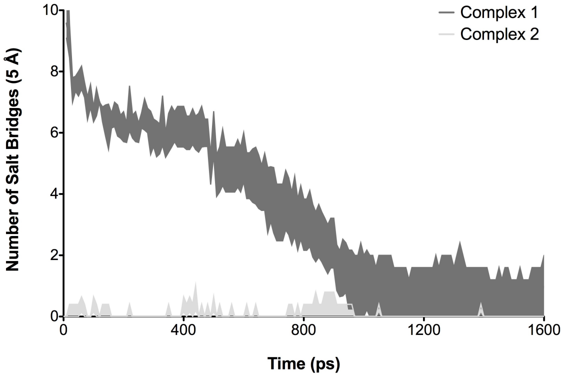


Figure D.9 Number of salt bridges (measured as charge-charge distance of < 5 Å) during SMD trajectories. The plot shows the mean +/- SEM of the number of hydrogen bonds in Complex 1 (dark gray) and Complex 2 (light gray) of C3d-Sbi-IV during five independent 1.6 ns SMD trajectories.

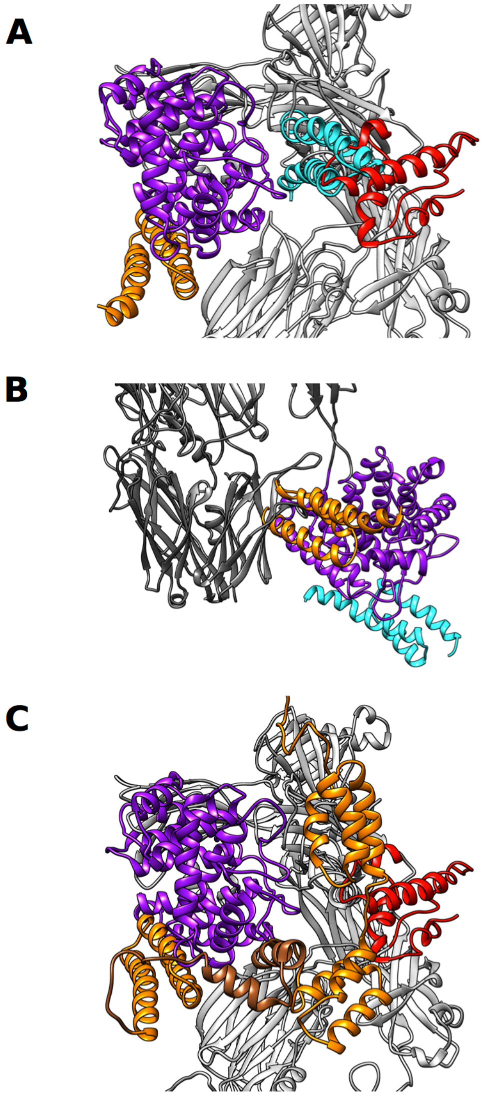


Figure D.10 Sbi binding to C3 and C3b. (A) Ribbon representation of Sbi-IV bound to C3 (gray) in Complex 1 (orange) and Complex 2 (cyan) binding modes. The thioester domain of C3 (C3d) is colored in purple, and the anaphylotoxin domain (C3a) is colored in red. Steric clashes exist between Sbi-IV of Complex 2 and the anaphylotoxin and MG8 domains of C3. (B) Sbi-IV binding to C3b (dark gray). Steric clashes exist between Sbi-IV of Complex 1 and MG1 domain of C3b. (C) Model of full-length extracellular Sbi (domains I-IV) bound to C3, with Sbi-IV bound in the Complex 1 binding mode. Sbi is colored in orange, with domain III colored in brown. Note the flexible loop region between domains III and IV, and the potential of domain III to reach C3a (in red).

References

1. Clark, E.A., S. Crennell, A. Upadhyay, A.V. Zozulya, J.D. Mackay, et al. 2010. A structural basis for Staphylococcal complement subversion: X-ray structure of the complement-binding domain of Staphylococcus aureus protein Sbi in complex with ligand C3d. Mol. Immunol. 48: 452–462.

2. Hammel, M., G. Sfyroera, D. Ricklin, P. Magotti, J.D. Lambris, et al. 2007. A structural basis for complement inhibition by Staphylococcus aureus. Nat. Immunol. 8: 430–437.

3. Hammel, M., G. Sfyroera, S. Pyrpassopoulos, D. Ricklin, K.X. Ramyar, et al. 2007. Characterization of Ehp, a secreted complement inhibitory protein from Staphylococcus aureus. J. Biol. Chem. 282: 30051–30061.

4. van den Elsen, J.M.H., and D.E. Isenman. 2011. Complement receptor complex structure. Science. 332: 608–611.

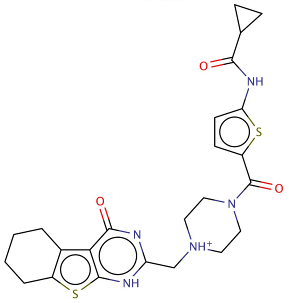
5. Wu, J., Y.-Q. Wu, D. Ricklin, B.J.C. Janssen, J.D. Lambris, et al. 2009. Structure of complement fragment C3b–factor H and implications for host protection by complement regulators. Nat. Immunol. 10: 729–734.

6. Kajander, T., M.J. Lehtinen, S. Hyvärinen, A. Bhattacharjee, E. Leung, et al. 2011. Dual interaction of factor H with C3d and glycosaminoglycans in host-nonhost discrimination by complement. Proc. Natl. Acad. Sci. U.S.A. 108: 2897–2902.

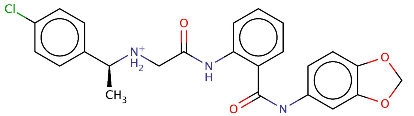
7. Morgan, H.P., C.Q. Schmidt, M. Guariento, B.S. Blaum, D. Gillespie, et al. 2011. Structural basis for engagement by complement factor H of C3b on a self surface. Nat. Struct. Mol. Biol. 18: 463–470.

APPENDIX E: SUPPORTING MATERIAL FOR CHAPTER 5

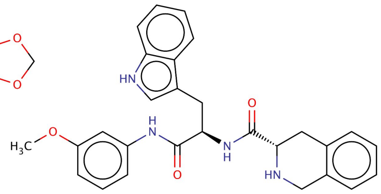
A1



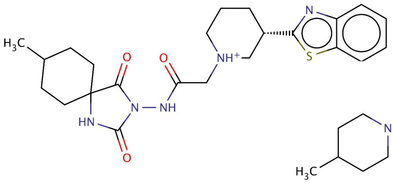
A2



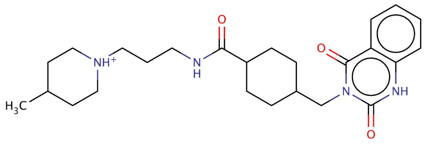
A3



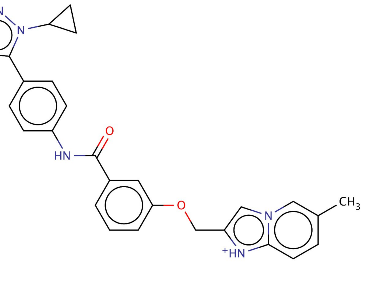
A4



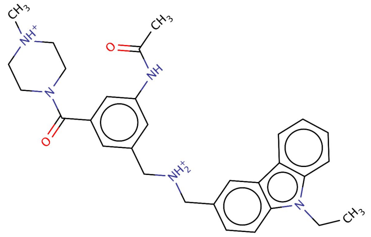
A5



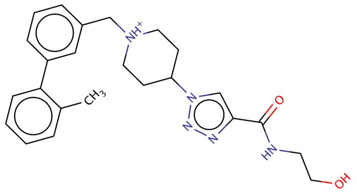
A6



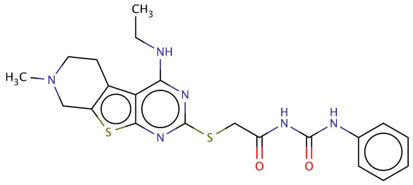
A7



A8



A9



A10

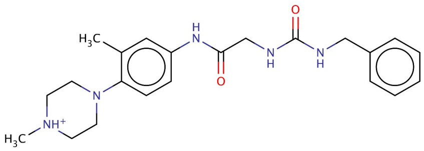


Figure E.1 Sturctures of molecules from Virtual Screen A selected for experimental testing. Molecule A1 was the observed C3d-binding molecule (see Figure 5.7B)

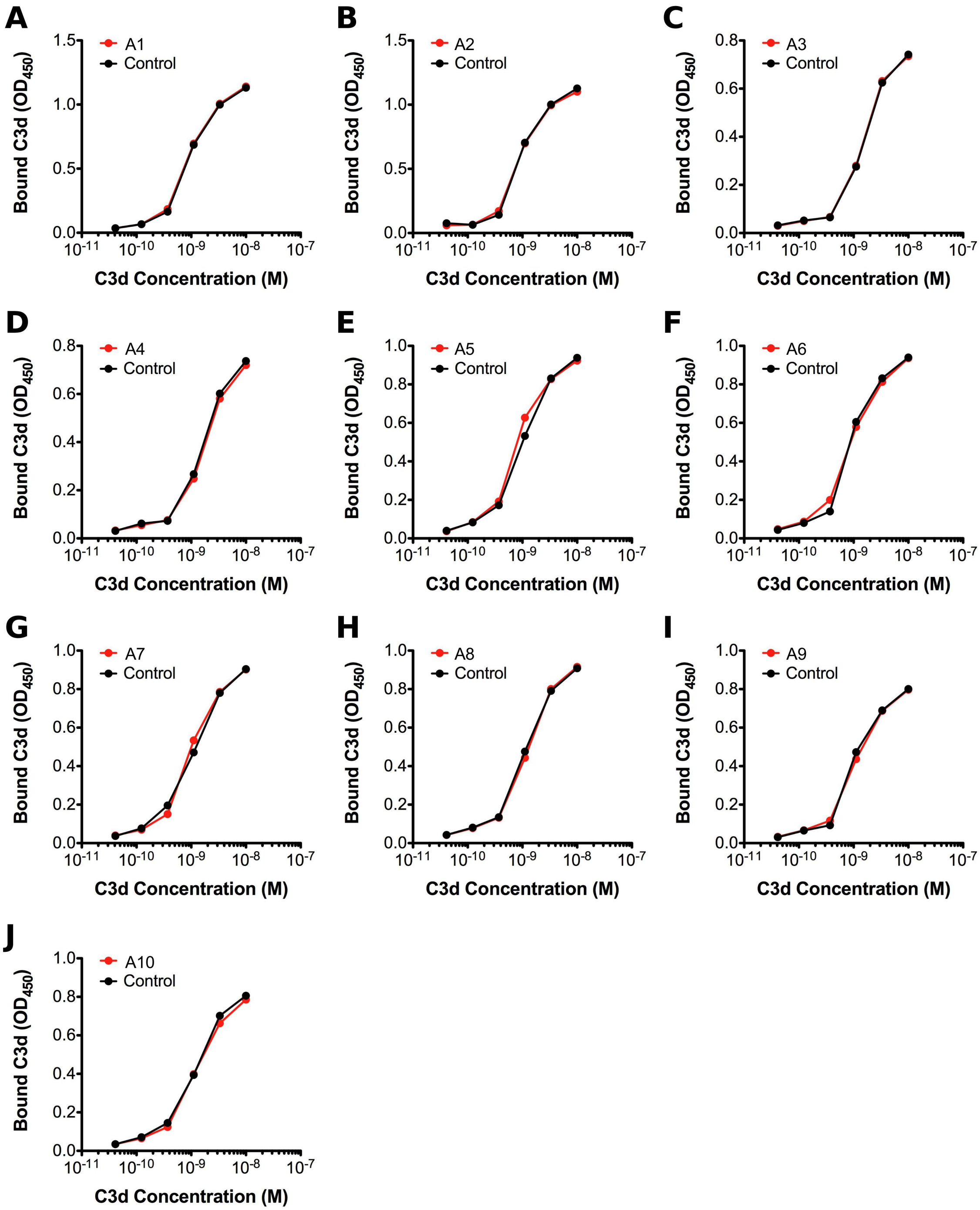


Figure E.2 Competitive binding ELISAs for small molecules. Ten small molecules selected from the initial virtual screen were preincubated with C3d and then added to Efb-C coated ELISA plates (red line). C3d alone was used as a control (black).

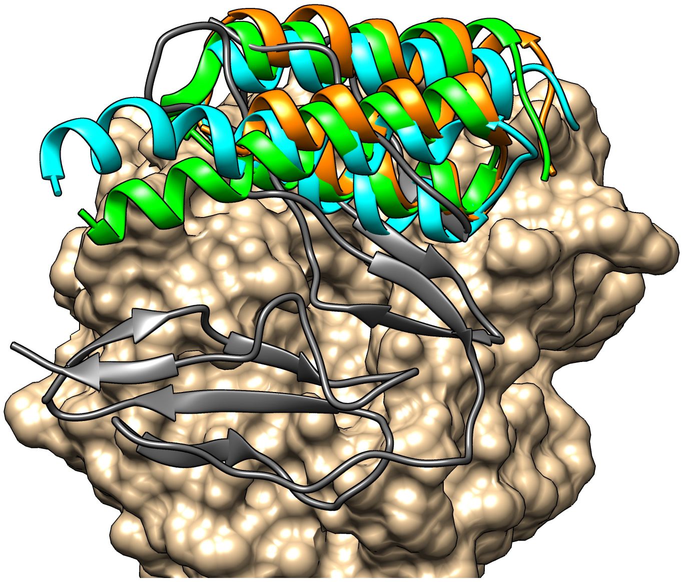
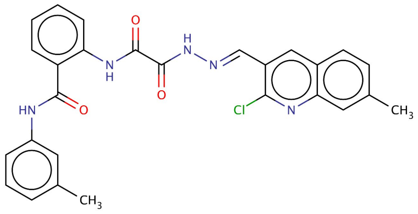
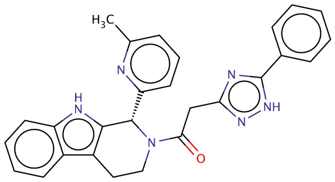


Figure E.3 Molecular graphics representation of C3d (tan) bound to CR2 (gray), Efb-C (green), Ecb (orange), and Sbi-IV (cyan). Notice the overlap in binding sites, making this site amenable for drug design.

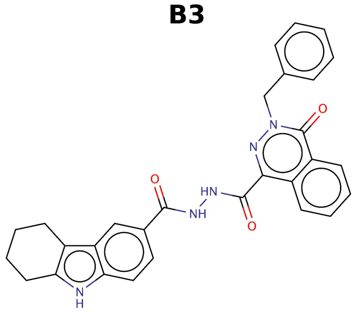
B1



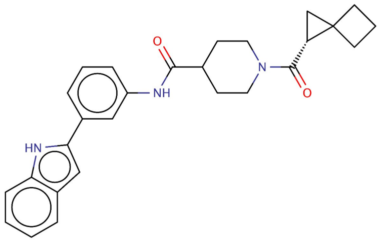
B2



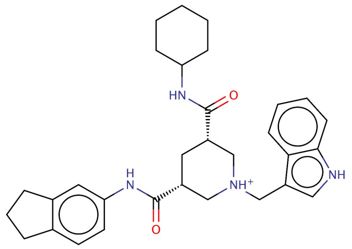
B3



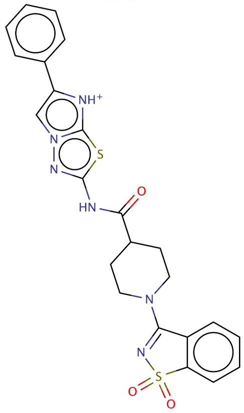
B4



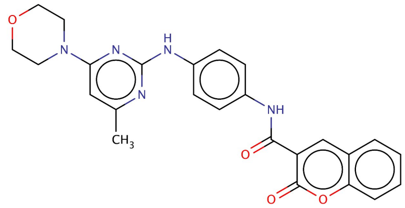
B5



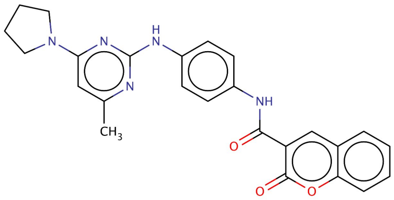
B6



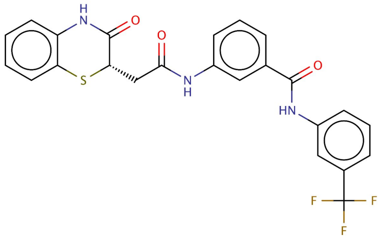
B7



B8



B9



B10

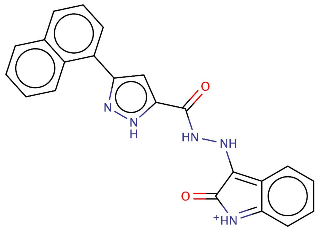
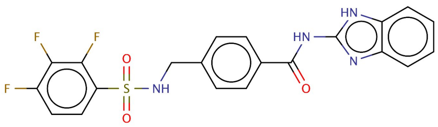
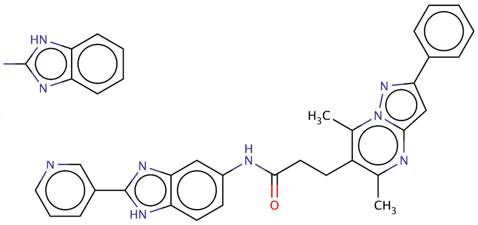


Figure E.4 Sturctures of molecules from Virtual Screen B selected for experimental testing.

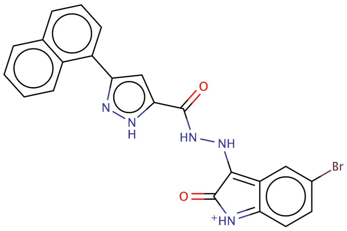
B11



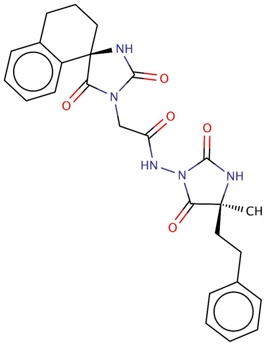
B12



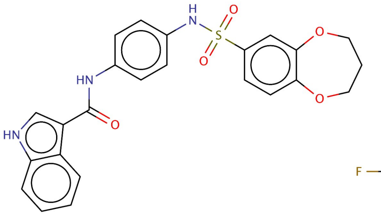
B13



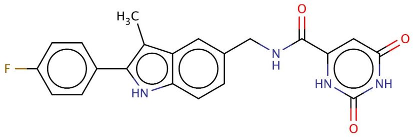
B14



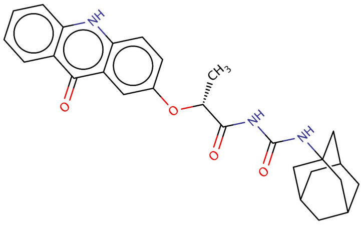
B15



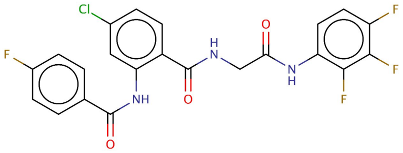
B16



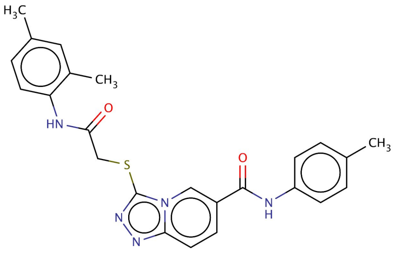
B17



B18



B19



B20

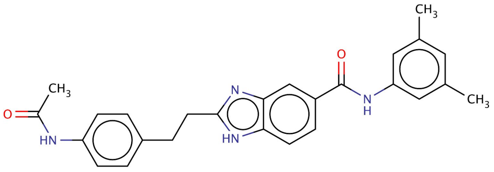
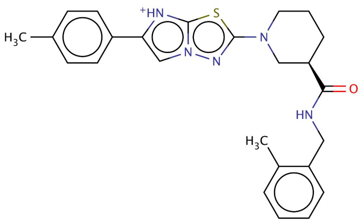
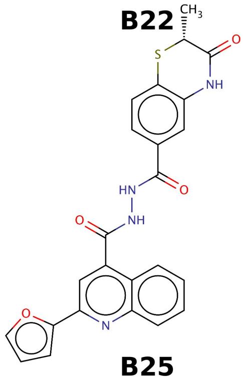


Figure E.4 Sturctures of molecules from Virtual Screen B selected for experimental testing (continued).

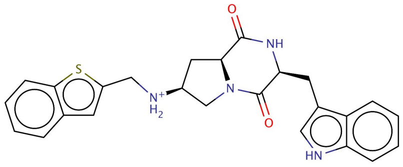
B21


H3C
+HN
S
N
N
N
O
HN
H3C

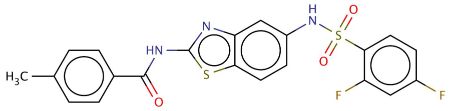
B22


CH3
S
O
NH
HN
O
O
NH
O
N

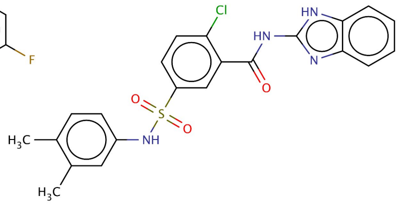
B23


S
N+
H2
O
NH
N
O
HN

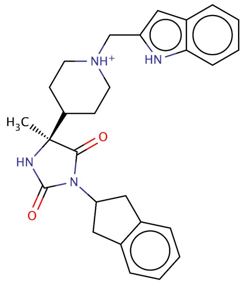
B24


H3C
HN
O
N
S
H
N
S
O
O
F
F

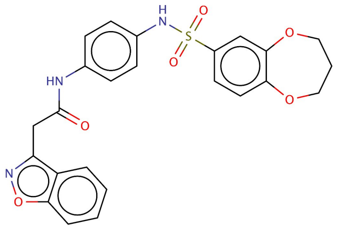
B25


Cl
HN
HN
N
O
O
S
O
NH
H3C
H3C

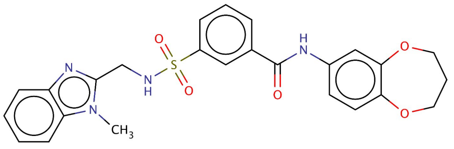
B26


NH+
HN
H3C
O
HN
N
O

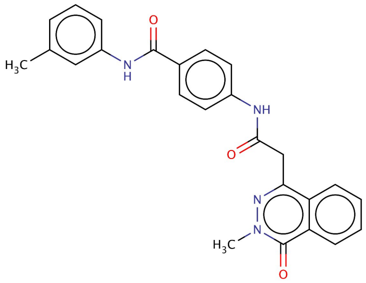
B27


H
N
S
O
O
O
O
HN
O
N
O

B28


N
N
CH3
N
H
S
O
O
H
N
O
O
O

B29


H3C
N
H
O
NH
O
N
N
H3C
O

B30

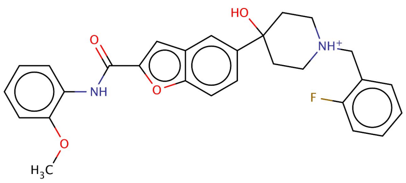
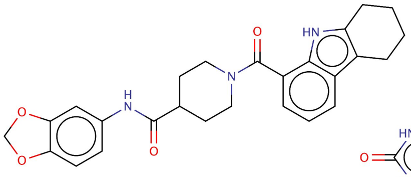
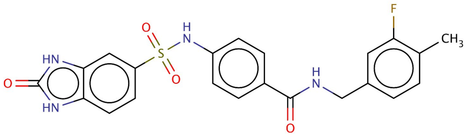

HO
NH+
O
NH
O
O
H3C
F

Figure E.4 Sturctures of molecules from Virtual Screen B selected for experimental testing (continued).

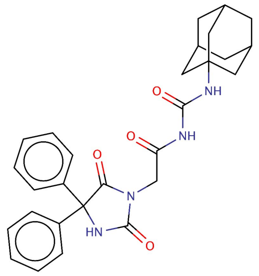
B31



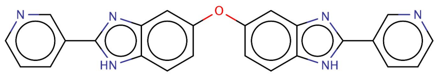
B32



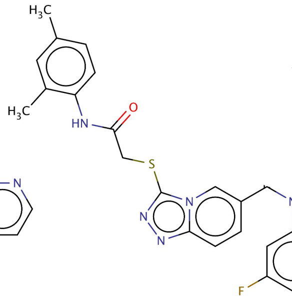
B33



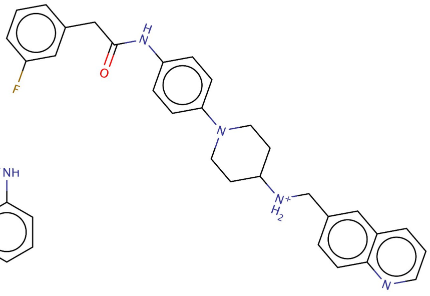
B34



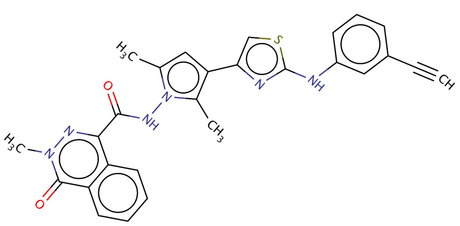
B35



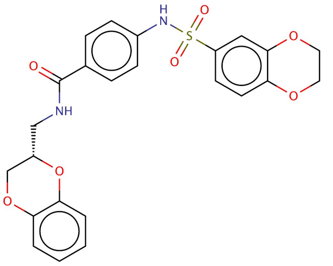
B36



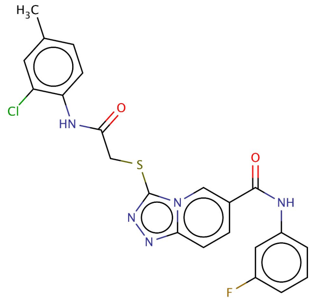
B37



B38



B39



B40

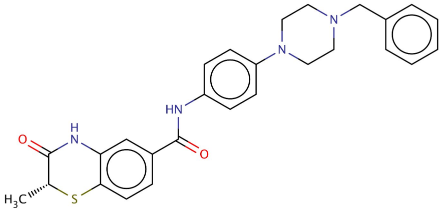
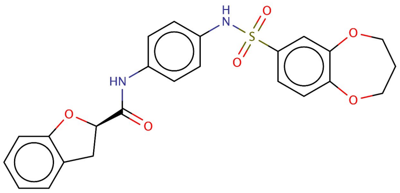
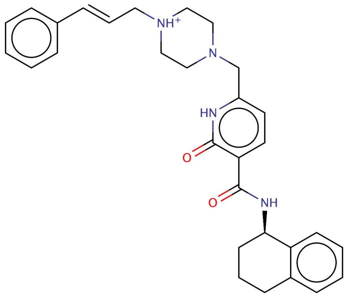


Figure E.4 Sturctures of molecules from Virtual Screen B selected for experimental testing (continued).

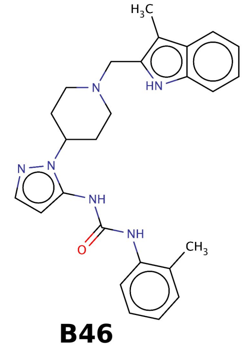
B41


H
N
O
S
O
O
O
HN
O
O

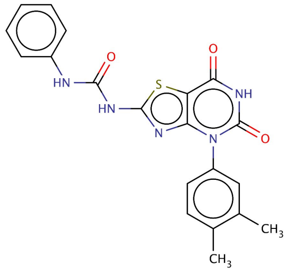
B42



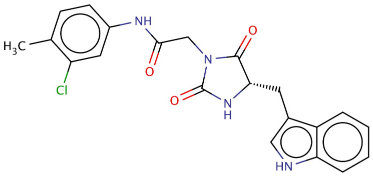
B43


H3C
N
HN
N
N
NH
O
NH
CH3

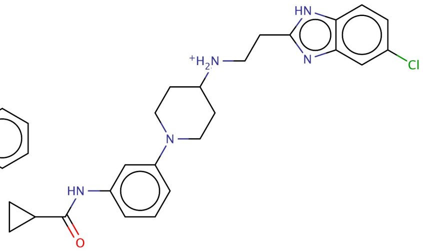
B44


O
O
HN
S
NH
HN
N
N
O
CH3
CH3

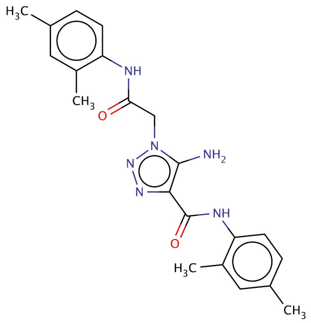
B45


NH
O
H3C
N
O
Cl
O
O
N
H
HN

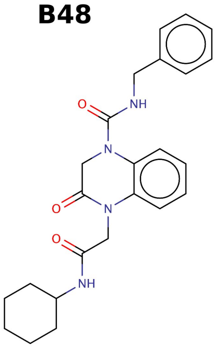
B46


HN
+H2N
N
Cl
N
HN
O

B47


H3C
NH
CH3
O
N
NH2
N
N
NH
O
H3C
CH3

B48



B49

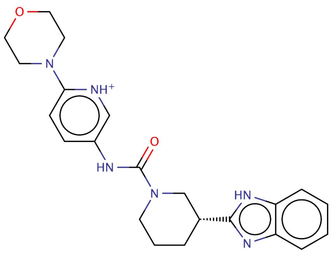

O
N
NH+
O
HN
N
HN
N

Figure E.4 Sturctures of molecules from Virtual Screen B selected for experimental testing (continued).

APPENDIX F: EXPERIMENTAL TESTING OF A NEW GENERATION OF COMPSTATIN-FAMILY PEPTIDES

F.1 Introduction

Compstatin is a 13-residue cyclic peptide that inhibits activation of the complement alternative pathway by binding to and inhibiting cleavage of C3 (1, 2). Originally discovered through phage display experiments, its sequence has been optimized using computational and experimental methods (reviewed in (3-10)), leading to identification of the most potent compstatin analog to date consisting entirely of natural amino acids (11-13). This analog consists of two amino acid substitutions, including Trp at position 4 and Ala at position 9 (W4A9). Structural determination showed the mechanism of compstatin (W4A9) binding to C3 at atomic resolution, and in conjunction with computational analyses, revealed that the mechanism of binding between compstatin peptides and C3 is dominated by hydrophobic interactions, as well as hydrogen bonds (14, 15). Incorporation of methylated-Trp at position 4 led to a more active analog (meW4A9) (16), which is currently in Phase II clinical trials for treatment of age-related macular degeneration (4, 5, 17-20). Despite potent complement inhibition, this peptide is prone to aggregation in aqueous solution, posing challenges to therapeutic development (17, 19, 20). Thus, it is of interest to identify potent compstatin analogs that maintain key hydrophobic contacts for strong binding to C3, while manipulating other residues to achieve improved solubility characteristics.

Our recent collaborative work has employed *de novo* sequence optimization techniques and rational design to predict sequence modifications that may improve binding of compstatin to C3 (21, 22). In addition, MD simulations were performed on selected sequences, to evaluate compstatin-C3 interactions in more detail (15, 23). Recent computational and rational design studies of compstatin peptides suggested that a di-serine N-terminal extension and arginine replacement at the N-terminus produce additional contacts that improve the stability of the complex (15). In this study, we evaluated three compstatin analogs with N-terminal modifications using *in vitro* complement assays.

Table F.1 Compstatin peptide sequences.

Peptide	**Sequence**[a]	**Molecular Mass**[b]
	-10 123 4567890123[c]	
I	Ac- **R**CV WQDWGAHRCT-NH$_2$	1656.0
II	Ac- **R**CV**meW**QDWGAHRCT-NH$_2$	1669.8
III	Ac-**RS**-ICV**meW**QDWGAHRCT-NH$_2$	1870.6
IV	Ac- ICV WQDWGAHRCT-NH$_2$	1614.0
Parent	ICV VQDWGHHRCT-NH$_2$	1549.8

[a] Bold face characters denote new amino acid combinations. Tryptophans methylated at the indole amide position are shown as meW. The peptides are cyclized with a disulfide bridge between C2 and C12. Ac denotes acetylation and NH$_2$ denotes amidation.

[b] From mass spectrometry.

[c] Numbering indicates sequence position, starting from position -1 and ending at position 13.

F.2 Materials and Methods

F.2.1 Peptide synthesis

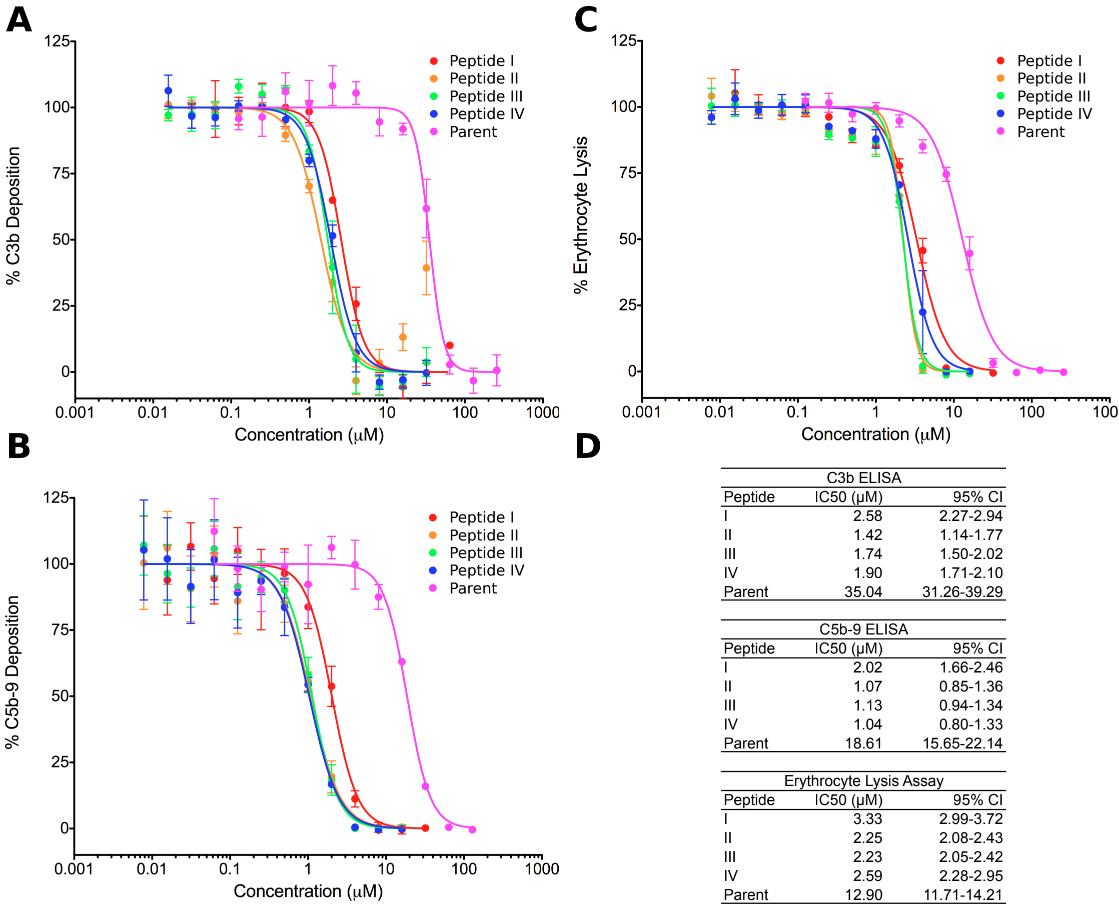
Compstatin family peptides were custom synthesized and characterized using liquid chromatography – mass spectrometry by Abgent Inc. The peptides were acetylated at the N-terminus and amidated at the C-terminus, with the exception of parent compstatin control, which was not acetylated at the N-terminus. All peptides were cyclized through a disulfide bridge between Cys2 and Cys12. Peptides containing methylated tryptophan (methylated-Trp) were synthesized using Fmoc-1-methyl-DL-tryptophan, and the resulting mixture of peptides containing D-methylated-Trp and L-methylated-Trp were subsequently purified to obtain peptides containing only the L-enantiomer. The peptide sequences are shown in Table F.1.

F.2.2 C3b and C5b-9 ELISAs

Inhibition of the complement alternative pathway (AP) by compstatin family peptides was assayed by ELISA. Peptides were dissolved in phosphate buffered saline (PBS, 150 mM). Initial peptide concentrations were calculated using the Beer–Lambert Law with an extinction coefficient of 5,500 $M^{-1}\cdot cm^{-1}$ for Trp and 5,470 $M^{-1}\cdot cm^{-1}$ for methylated-Trp present in the sequence, at 280 nm. Nunc Maxisorp 96-well plates were coated with 1 μg lipopolysaccharides (LPS) from *Salmonella enteriditis* for ~16 h at room temperature. Plates were washed three times with PBS/0.05% Tween-20 (PBS-T) between each step. Plates were blocked with 4% bovine serum albumin (BSA) in PBS-T for 1 h at 37°C. Serial peptide dilutions were performed in 96-well plates, using gelatin veronal-buffered saline with 5 mM $MgCl_2$ and 10 mM EGTA (GVBS-MgEGTA). Normal human serum (Complement Technology Inc.) was diluted in GVBS-MgEGTA and mixed with compstatin dilutions to a final concentration of 30%. Serum diluted in GVBS-MgEGTA and GVBS (containing 20 mM EDTA) were used as positive and negative controls for complement activation, respectively. Dilutions were preincubated for 15 minutes at room temperature, transferred to ELISA plates, and incubated for 1 h at 37°C. Generation of C3b and C5b-9 were assayed using horseradish peroxidase (HRP)-conjugated anti-C3 (MP Biomedicals) and anti-C5b-9 aE11 (Abcam), respectively. Plates were washed and incubated with anti-C3-HRP (1:5000 in 1% BSA/PBS-T) or anti-C5b-9 (1:1000 in 1% BSA/PBS-T) for 1 h at 37°C. For C5b-9 detection, primary antibody incubation was followed by incubation with anti-mouse-HRP (BioRad) for 1 h at 37°C (1:5000 in 1% BSA/PBS-T). Bound C3b and C5b-9 were quantified using a 3,3',5,5'-tetramethylbenzidine substrate solution containing urea hydrogen peroxide in 0.11 M sodium acetate buffer, followed by a 1 N H_2SO_4 acid stop. Plates were measured spectrophotometrically at 450 nm. Percent inhibition of C3b and C5b-9 deposition was plotted against peptide concentration and data was fitted to a logistic dose response curve with Prism (GraphPad) to determine IC_{50} values. Detailed protocols of C3b and C5b-9 ELISAs are given in Appendix G.1.

F.2.3 Hemolytic Assays

Inhibition of complement was also measured via lysis of erythrocytes. Rabbit erythrocytes (Complement Technology Inc., Tyler, TX, USA) were washed in PBS and resuspended in veronal-buffered saline with 5 mM $MgCl_2$ and 10 mM EGTA (VBS-MgEGTA). Peptide and serum dilutions were prepared as described above, and 1×10^7 erythrocytes were added to each serum/peptide mixture. Erythrocytes diluted in sterile deionized water and in VBS-MgEGTA were used as positive and negative controls for lysis, respectively. Plates were incubated for 20 minutes at 37°C, and centrifuged at 2500×g for 10 minutes. Supernatants were diluted 1:2 and absorbance was measured at 405 nm. Detailed protocols of C3b and C5b-9 ELISAs are given in Appendix G.2.



C3b ELISA

Peptide	IC50 (μM)	95% CI
I	2.58	2.27-2.94
II	1.42	1.14-1.77
III	1.74	1.50-2.02
IV	1.90	1.71-2.10
Parent	35.04	31.26-39.29

C5b-9 ELISA

Peptide	IC50 (μM)	95% CI
I	2.02	1.66-2.46
II	1.07	0.85-1.36
III	1.13	0.94-1.34
IV	1.04	0.80-1.33
Parent	18.61	15.65-22.14

Erythrocyte Lysis Assay

Peptide	IC50 (μM)	95% CI
I	3.33	2.99-3.72
II	2.25	2.08-2.43
III	2.23	2.05-2.42
IV	2.59	2.28-2.95
Parent	12.90	11.71-14.21

Figure F.1 ELISA and hemolytic assay data for compstatin peptides. (A) C3b ELISA data, representing the inhibition of cleavage of C3 to C3a and C3b by compstatin peptides, quantified as inhibition of the formation of C3b. (B) C5b-9 ELISA data, representing inhibition of the formation of the C5b-9 terminal complex of complement activation. (C) Hemolytic assay data, representing inhibition of rabbit erythrocyte hemolysis by the C5b-9 terminal complex activation. (D) IC50 values and 95% confidence intervals extracted from A, B, and C. The data points and error bars correspond to means and S.E.M. from three independent experiments.

F.3 Results and Discussion

This study was motivated by the impetus to design and identify new compstatin analogs with improved binding, complement inhibition, and solubility characteristics. The scope of the design was two-fold; first, to promote binding to the target protein C3 through key amino acid side chain interactions, and second, to enhance solubility compared to previously known compstatin analogs, by improving the peptide polarity/hydrophobicity ratio. The previously reported most potent peptide to date, similar to Peptide IV but with methylated-Trp4, is known to possess precipitation characteristics and exhibits intravitreal deposit formation (17, 19, 20), thus it is of interest to identify potent compstatin analogs with improved solubility.

The design of the new compstatin peptides was inspired by our recent studies that include (i) the application of two-stage computational frameworks for protein design that are based on sequence selection, fold specificity, and approximate binding affinity calculations (21); (ii) rational design based on structural/physicochemical analysis – activity relations (22); and (iii) structural and binding analysis using molecular dynamics simulations (15). These studies identified new active compstatin family peptides with novel features, such as introduction of tryptophans at positions 1 and 13, arginine at position 1, and an N-terminal di-serine extension. Some of these analogs suffer from solubility/aggregation problems, owing to increased hydrophobicity content conferred by the presence of up to four tryptophans. Nevertheless, these studies revealed three new sites of optimization in the compstatin peptide sequence at positions 1, 0, and -1. We reasoned that incorporation of polar amino acids at positions 1, 0, and -1, while maintaining the amino acids that are involved in hydrophobic contacts with C3, would aid in balancing the polarity/hydrophobicity ratio and in improving solubility. The choice of amino acids for positions 1, 0, and -1 was based on our previous molecular dynamics simulations (15) and supported by new molecular dynamics simulations.

Table F.1 shows the sequences of the peptides studied. With the exception of Parent, the sequences of Table F.1 contain either Trp or methylated-Trp at position 4 and Ala at position 9. The new sequences include N-terminal replacements of Ile with Arg (Peptides I and II), as well as an N-terminal extension (Peptide III). These sequences are modeled after the sequence of the most potent peptides comprised entirely of natural amino acids [control Peptide IV; (13)] and the most potent peptide to-date, which is similar to Peptide IV but with methylated-Trp at position 4, with methylation being attached at the indole nitrogen (16). Two of the new sequences in Table F.1 incorporate methylated-Trp at position 4 (Peptides II and III), in an effort to maintain high potency, and all contain polar amino acids at positions -1, 0, and 1. In addition, Table F.1 also shows two compstatin peptides (Peptide IV and Parent) that were used as controls.

Since compstatin family peptides are known to selectively inhibit the complement alternative pathway (AP), we performed ELISAs and erythrocyte lysis assays to examine AP inhibition. In all assays, MgEGTA was used to chelate Ca^{2+} ions required for classical/lectin pathway complement activation, while providing Mg^{2+} necessary for AP activation. ELISAs allow for direct detection of C3b and C5b-9 formed during AP activation, while erythrocyte lysis assays provide a functional assessment of AP-mediated hemolysis in serum. Figure F.1 shows the C3b and C5b-9 ELISA and erythrocyte lysis data for Peptides I, II, III, IV, and Parent. Parent exhibited significantly less complement inhibition than all other peptides in all three assays. Newly-designed peptides (Peptides I, II, and III) showed similar IC_{50} values to Peptide IV, with Peptide I having a slightly higher IC_{50} than the others in all cases. The relative potencies of the peptides, based on the measured IC_{50} values, are consistent across all three assays (Peptide II ~ Peptide III ~ Peptide IV > Peptide I > Parent). Based on the assay conditions used in this study, peptides inhibited C5b-9 formation with moderately better efficacy than formation of C3b, while slightly higher concentrations were necessary to inhibit AP-mediated hemolysis. Irrespective of the magnitudes of complement inhibition in the various assays used in this (and other) studies, the relative efficacies of Peptides I, II, III, IV, and Parent remained consistent.

During the course of these experiments, we also assessed the relative solubility of each peptide. Solubility was determined via concentration measurement (absorbance at 280 nm) of saturated peptide solutions, with Peptides I, II, and Parent having approximate aqueous solubility of > 1 mM, Peptide III > 500 μM, and Peptide IV < 500 μM. Given the improved solubility properties of Peptides II and III, they may be good alternatives to the currently most potent peptide (possessing methylated-Trp4), which exhibits aggregation at high concentrations in aqueous environments, and provide templates for further optimization of highly soluble and potent compstatin analogs.

F.4 References

1. Sahu, A., B. Kay, and J.D. Lambris. 1996. Inhibition of human complement by a C3-binding peptide isolated from a phage-displayed random peptide library. J. Immunol. 157: 884-891.

2. Morikis, D., N. Assa-Munt, A. Sahu, and J.D. Lambris. 1998. Solution structure of Compstatin, a potent complement inhibitor. Protein Sci. 7: 619–627.

3. Morikis, D., and J.D. Lambris. 2005. Structure, dynamics, activity, and function of compstatin anfd design of more potent analogues. In: Structural Biology of the Complement System. Boca Raton, FL, CRC Press USA, pp. 317-340.

4. Ricklin, D., and J.D. Lambris. 2007. Complement-targeted therapeutics. Nat Biotechnol. 25: 1265–1275.

5. Ricklin, D., and J.D. Lambris. 2008. Compstatin: A complement inhibitor on its way to clinical application. In: Current Topics in Complement II. New York, NY: Springer US. pp. 262–281.

6. Qu, H., D. Ricklin, and J.D. Lambris. 2009. Recent developments in low molecular weight complement inhibitors. Mol. Immunol. 47: 185–195.

7. Morikis, D., and J.D. Lambris. 2002. Structural aspects and design of low-molecular-mass complement inhibitors. Biochem. Soc. Trans. 30: 1026–1036.

8. Morikis, D., A.M. Soulika, B. Mallik, J.L. Klepeis, C.A. Floudas, et al. 2004. Improvement of the anti-C3 activity of compstatin using rational and combinatorial approaches. Biochem. Soc. Trans. 32: 28–32.

9. Holland, M.C.H., D. Morikis, and J.D. Lambris. 2004. Synthetic small-molecule complement inhibitors. Curr. Opin. Investig. Drugs. 5: 1164–1173.

10. Morikis, D., C. Floudas, and J. Lambris. 2005. Structure-based integrative computational and experimental approach for the optimization of drug design. In: ICCS 2005, Lecture Notes in Computer Science: Computational Science. Berlin-Heidelberg, Springer-Verlag, Germany, pp. 680–688.

11. Klepeis, J.L., C.A. Floudas, D. Morikis, C.G. Tsokos, E. Argyropoulos, et al. 2003. Integrated computational and experimental approach for lead optimization and design of compstatin variants with improved activity. J. Am. Chem. Soc. 125: 8422–8423.

12. Klepeis, J.L., C.A. Floudas, D. Morikis, C.G. Tsokos, and J.D. Lambris. 2004. Design of peptide analogues with improved activity using a novel de novo protein design approach. Ind. Eng. Chem. Res. 43: 3817–3826.

13. Mallik, B., M. Katragadda, L.A. Spruce, C. Carafides, C.G. Tsokos, et al. 2005. Design and NMR characterization of active analogues of compstatin containing non-natural amino acids. J. Med. Chem. 48: 274–286.

14. Janssen, B.J.C., E.F. Halff, J.D. Lambris, and P. Gros. 2007. Structure of compstatin in complex with complement component C3c reveals a new mechanism of complement inhibition. J. Biol. Chem. 282: 29241–29247.

15. Tamamis, P., A. López de Victoria, R.D. Gorham, M.L. Bellows-Peterson, P. Pierou, et al. 2012. Molecular dynamics in drug design: new generations of compstatin analogs. Chem. Biol. Drug Des. 79: 703–718.

16. Katragadda, M., P. Magotti, and G. Sfyroera. 2006. Hydrophobic effect and hydrogen bonds account for the improved activity of a complement inhibitor, compstatin. J. Med. Chem. 49: 4616-4622.

17. Chi, Z.-L., T. Yoshida, J.D. Lambris, and T. Iwata. 2010. Suppression of drusen formation by compstatin, a peptide inhibitor of complement C3 activation, on cynomolgus monkey with early-onset macular degeneration. Adv. Exp. Med. Biol. 703: 127–135.

18. Zarbin, M.A., and P.J. Rosenfeld. 2010. Pathway-based therapies for age-related macular degeneration: an integrated survey of emerging treatment alternatives. Retina. 30: 1350–1367.

19. Yehoshua, Z., P.J. Rosenfeld, and T.A. Albini. 2011. Current clinical trials in dry AMD and the definition of appropriate clinical outcome measures. Semin. Ophthalmol. 26: 167–180.

20. Qu, H., D. Ricklin, H. Bai, H. Chen, E.S. Reis, et al. 2013. New analogs of the clinical complement inhibitor compstatin with subnanomolar affinity and enhanced pharmacokinetic properties. Immunobiology. 218: 496–505.

21. Bellows, M.L., H.K. Fung, M.S. Taylor, C.A. Floudas, A.L. de Victoria, et al. 2010. New compstatin variants through two de novo protein design frameworks. Biophys. J. 98: 2337–2346.

22. López de Victoria, A., R.D. Gorham Jr, M.L. Bellows-Peterson, J. Ling, D.D. Lo, et al. 2011. A new generation of potent complement inhibitors of the Compstatin family. Chem. Biol. Drug Des. 77: 431–440.

23. Tamamis, P., D. Morikis, C.A. Floudas, and G. Archontis. 2010. Species specificity of the complement inhibitor compstatin investigated by all-atom molecular dynamics simulations. Proteins. 78: 2655–2667.

APPENDIX G: PROTOCOLS FOR FUNCTIONAL ASSAYS OF COMPLEMENT INHIBITION

G.1 C3b and C5b-9 alternative pathway complement ELISAs

Stock buffer preparation:

Sodium carbonate buffer (0.1M, pH 9.6) – 300ml final volume, store at 4°C
- $NaHCO_3$ (0.1M)(84g/mol)(250ml) = 2.10g in 250ml MQ
- Na_2CO_3 (0.1M)(106g/mol)(100ml) = 1.06g in 100ml MQ
- Add $NaHCO_3$ to 100ml Na_2CO_3 until pH 9.6

Lipopolysaccharides (LPS) from *Salmonella enterica* serotype *enteritidis* (1mg/ml) – 1ml, store at -20°C
- Dissolve 1mg in vial directly in 1ml Milli-Q water (MQ)

10x PBS - 500ml final volume, store at room temperature
- 2.28g NaH_2PO_4 (monobasic)
- 11.5g Na_2HPO_4 (dibasic)
- 43.84g NaCl
- Dilute in 450ml MQ
- Adjust to pH 7.4
- Fill to 500ml MQ

VBS (5x) – 250ml final volume, store at 4°C
- 10.63g NaCl
- 468.75mg sodium barbital
- 468.75mg barbituric acid
- Dilute in 200ml MQ
- Adjust to pH 7.4
- Fill to 250ml with MQ

Mg-EGTA (10x) – 250ml final volume, store at 4°C
- $MgCl_2$ (0.05M)(95.21g/mol)(0.25ml) = 1.19g
- 50 ml EGTA (0.5M)
- Fill to 250ml with MQ

Gelatin – 25ml final volume, store in aliquots at 4°C
- Gelatin (10%)(25ml) = 2.5g gelatin
- Dissolve in 25ml MQ

Sodium acetate (1.1M, pH 6) – 100ml final volume, store at 4°C
- CH_3COONa (1.1M)(82.03g/mol)(100ml) = 9.02g
- Dilute in 80ml MQ
- Adjust to pH 6
- Fill to 100ml with MQ

TMB (6mg/ml in DMSO) – 20ml final volume, store at room temperature (in darkness and in glass vial)
- 3,3',5,5'-tetramethylbenzidine (6mg/ml)(20ml) = 120mg
- Dissolve in 20ml DMSO

Urea hydrogen peroxide (20mg/ml) – 20ml final volume, store at 4°C
- Urea hydrogen peroxide (20mg/ml)(20ml) = 400mg
- Dissolve in 20ml MQ

H_2SO_4 (1N) – 250ml final volume, store at room temperature
- 1N H_2SO_4 = 0.5M H_2SO_4
- (0.5M)(98g/mol)(0.25L)/(0.95)(1.84g/L) = 7.0ml H_2SO_4 + 243ml MQ

Additional supplies: normal human serum (Complement Technology, Inc.), Nunc Maxisorp 96-well plates

Daily buffer preparation:

PBS-T (1x, 0.05% Tween 20) – 500ml (can store until used completely at room temperature)
- 50ml 10x PBS
- 450ml MQ
- 250ul Tween 20

LPS (20µg/ml for assay use)
- Dilute 1:50 in sodium carbonate buffer
- For each ELISA plate, use 120ul stock in 6ml sodium carbonate buffer

BSA (4% in PBS-T)
- Dilute 1g BSA in 25ml PBS-T

GVBS/Mg-EGTA
- 4ml VBS (5x)
- 2ml MgEGTA solution
- 200ul gelatin (10% solution)
- Fill to 20ml with MQ

Detection Solution (per plate) – prepare immediately prior to detection of ELISA plates
- 5.4ml MQ
- 600ul sodium acetate buffer
- 100ul TMB solution
- 50ul urea hydrogen peroxide solution

Select appropriate serum concentration to measure complement inhibition by test molecules:
- We need to select a serum concentration to use for each ELISA type (C3b and C5b-9), so we will perform serum dilutions to examine how highly each concentration absorbs after detection, and how well that absorption signal is inhibited by a known complement inhibitor control.
- After running this ELISA, make note of the lowest serum concentration at which definitive inhibition has occurred [OD_{450} (serum) – OD_{450} (serum+inhibitor)]

Day 1:

Coat plates with LPS

Prepare 20µg/ml LPS in 0.1M Na_2CO_3 buffer, pH 9.6
Add 50µl of LPS (20µg/ml) to each well using multichannel pipette
Incubate plate overnight at room temperature (16-18h)

Day 2:

Blocking:
- Wash LPS-coated plates 3x in PBS-T
- Add 80µl/well 4% BSA in PBS-T
- Incubate at 37°C for 1h

Serum dilutions:
- Prepare a 2X stock of the inhibitor molecule (control)
  - Dilute inhibitor in GVBS-MgEGTA to two times the desired final concentration
  - Should choose a concentration that will surely fully inhibit alternative pathway
- Test serum concentrations from 0-50%
  - Prepare 100, 80, 60, 40, 30, 20, 10, and 0% serum (in GVBS-MgEGTA)

- In a predilution plate:
    - Mix equal parts of 2X inhibitor (or buffer) with 2X serum dilutions (i.e. 30µl + 30µl)
    - Incubate for 15 mins at room temperature

Serum incubation:

- Wash blocked plates 3x with PBS-T
- Transfer 50µl predilutions to each plate
- Incubate at 37°C for 1h

Antibody Incubation

- Wash plates 3x with PBS-T
- Prepare antibody:
    - Anti-C3-HRP (1:1000 in 1% BSA/PBS-T)
    - Anti-C5b-9 (1:1000 in 1% BSA/PBS-T)
- Add 50µl of antibody to appropriate wells
- Incubate at 37°C for 1h

Detection

- Prepare detection solution (as described above)
- Add 50µl to all wells
- Read OD_{450} after a few minutes (depends on the rate of reaction)

Test inhibitor candidates for inhibition of C3b and C5b-9 formation

- Follow identical protocol as above, but perform serial dilutions of inhibitor rather than serum

Inhibitor dilutions:

- Prepare a 2X stock of serum
    - Dilute serum in GVBS-MgEGTA to two times the desired final concentration
- Test inhibitor at different concentrations
    - If you know the approximate inhibitory range, prepare 2-fold dilutions around the midpoint of the inhibitory concentrations
    - If you are unsure, start with a high concentration (~200µM) and perform 3-fold dilutions
    - All dilutions should be prepared in GVBS-MgEGTA
- In a predilution plate:
    - Mix equal parts of 2X inhibitor dilutions with 2X serum (i.e. 30µl + 30µl)
    - Incubate for 15 mins at room temperature

Follow remainder of protocol as described for the initial ELISA

G.2 Erythrocyte Lysis Assay (also known as hemolysis assay or AP_{50})

Stock buffer preparation:

10x PBS - 500ml final volume, store at room temperature
- 2.28g NaH_2PO_4 (monobasic)
- 11.5g Na_2HPO_4 (dibasic)
- 43.84g NaCl
- Dilute in 450ml MQ
- Adjust to pH 7.4
- Fill to 500ml MQ

VBS (5x) – 250ml final volume, store at 4°C
- 10.63g NaCl
- 468.75mg sodium barbital
- 468.75mg barbituric acid
- Dilute in 200ml MQ
- Adjust to pH 7.4
- Fill to 250ml with MQ

Mg-EGTA (10x) – 250ml final volume, store at 4°C
- $MgCl_2$ (0.05M)(95.21g/mol)(0.25ml) = 1.19g
- 50 ml EGTA (0.5M)
- Fill to 250ml with MQ

Daily buffer preparation:

PBS (1x) – 500ml (can store until used completely at room temperature)
- 50ml 10x PBS
- 450ml MQ
- 250ul Tween 20

VBS/Mg-EGTA
- 8ml VBS (5x)
- 4ml MgEGTA solution
- 28ml MQ

Additional supplies: normal human serum and rabbit erythrocytes (Complement Technology, Inc.), Costar 96-well round-bottom plates, Nunc Maxisorp 96-well plates

---

Select appropriate serum concentration to measure complement inhibition by test molecules:
- We need to select a serum concentration to use for the lysis assay, so we will perform serum dilutions to examine how well lysis is inhibited by a known complement inhibitor control.
- After running this ELISA, make note of the lowest serum concentration at which definitive inhibition has occurred [OD_{405} (serum) – OD_{405} (serum+inhibitor)]

Erythrocyte preparation

- Calculate the number of cells needed for the whole experiment
    - Typically use 50µl of 1 x 10^7 cells/well (stock cells from CompTech are 5 x 10^8 cells/ml)
- Wash cells 3x in PBS
    - Aliquot cells in 1ml microcentrifuge tubes
    - Centrifuge cells for 5 min, 2000rpm, 4deg
    - Remove supernatant and resuspend cells in 1ml PBS
    - Repeat 2x more
    - After final wash, resuspend pellets in 1ml VBS/MgEGTA
- Measure cell concentration
    - Dilute 50ul cells in 950ul MQ - will lyse cells
    - Use MQ as blank, and then read OD_{405} of sample
    - If OD_{405} is > 1.23, dilute cells in more VBS-MgEGTA until OD_{405} ~ 1.23

Serum dilutions:

- Prepare a 3X stock of the inhibitor molecule (control)
    - Dilute inhibitor in VBS-MgEGTA to two times the desired final concentration
    - Should choose a concentration that will surely fully inhibit alternative pathway
- Test serum concentrations from 0-30%
    - Prepare 90, 75, 60, 45, 30, 15, 7.5, and 0% serum (in VBS-MgEGTA)
- In a predilution plate (Costar round-bottom plate):
    - Mix 50µl of 3X inhibitor (or buffer) with 50µl 3X serum dilutions
    - Include a few wells with 100µl MQ or 100µl VBS-MgEGTA as positive and negative controls for erythrocyte lysis
    - Incubate for 15 mins at room temperature

Serum incubation:

- Add 50µl rabbit erythrocytes (2 x 10^8 cells/ml) to all wells
- Incubate at 37°C for 20 mins

Lysis determination:

- Centrifuge plate at 2500rpm for 10 mins
- Transfer 50µl supernatant from each well to Nunc Maxisorp ELISA plate (do not disturb pellet)
- Add 100ul MQ to each well in ELISA plate
- Read OD_{405}

Test inhibitor candidates for inhibition of C3b and C5b-9 formation

- Follow identical protocol as above, but perform serial dilutions of inhibitor rather than serum

Inhibitor dilutions:

- Prepare a 3X stock of serum
    - Dilute serum in VBS-MgEGTA to three times the desired final concentration
- Test inhibitor at different concentrations
    - If you know the approximate inhibitory range, prepare 2-fold dilutions around the midpoint of the inhibitory concentrations
    - If you are unsure, start with a high concentration (~200µM) and perform 3-fold dilutions
    - All dilutions should be prepared in VBS-MgEGTA

Follow remainder of protocol as described for the initial erythrocyte lysis assay

G.3 Suppliers and catalog numbers for reagents

Reagent	**Supplier**	**Catalog Number**
Sodium bicarbonate ($NaHCO_3$)	Sigma-Aldrich	S6297
Sodium carbonate (Na_2CO_3)	Sigma-Aldrich	S7795
LPS from *Salmonella enteritidis*	Sigma-Aldrich	L7770
NaH_2PO_4	Fisher	BP329-500
Na_2HPO_4	Fisher	BP332-500
NaCl	Acros Organics (Fisher)	AC207790010
Sodium barbital	Sigma-Aldrich	B0500-25G
Barbituric acid	Sigma-Aldrich	185698-25G
$MgCl_2$	Acros Organics (Fisher)	AC22321-1000
EGTA	Fisher	Q2783-100
Gelatin	Sigma-Aldrich	G9391-100G
Sodium acetate (CH_3COONa)	Sigma-Aldrich	S8750
3,3’,5,5’-tetramethylbenzidine (TMB)	Sigma-Aldrich	860336
Urea hydrogen peroxide	Sigma-Aldrich	289132
H_2SO_4	Sigma-Aldrich	320501
Tween-20	Fisher	BP337-100
Bovine serum albumin	Sigma-Aldrich	A7906
Rabbit erythrocytes	Complement Technology	B302
Normal human serum	Complement Technology	NHS
Goat anti-C3-HRP	MP Biomedicals (Fisher)	0855237
Mouse anti-C5b-9 (aE11)	Abcam	ab66768
Goat anti-mouse-HRP	BioRad	172-1011
96-well round bottom plates	Costar (Fisher)	07-200-105
Maxisorp 96-well plates	Nunc (Fisher)	12-565-136

APPENDIX H: CURRICULUM VITAE

Contact Information:

Office Address: 900 University Avenue, MSE 225, University of California, Riverside, CA, 92521
Phone Number: (951) 236-9305
E-mail Address: rgorham2@gmail.com

Education:

University of California, Riverside (Riverside, CA) 2008 – 2013
- Ph.D. in Bioengineering

Johns Hopkins University (Baltimore, MD) 2004 – 2008
- B.S. in Chemical and Biomolecular Engineering

Research Experience:

University of California, Riverside (Riverside, CA) 2009 – 2013
Advisor: Dimitrios Morikis, Department of Bioengineering
Title: Ph.D. Candidate
- Evaluated and benchmarked Poisson-Boltzmann electrostatic free energy calculations against experimental alanine mutagenesis data
- Identified potential small molecule inhibitors of bacterial-complement protein interactions
- Designed, optimized, and biophysically characterized potential complement-binding peptides derived from bacterial virulence factors
- Examined molecular basis of complement regulation and evasion
- Skills: molecular modeling, molecular and Brownian dynamics simulations, continuum electrostatics, homology modeling, pharmacophore modeling, virtual screening/docking, R and Python scripting, NMR, circular dichroism, ELISA

University Medical Center Utrecht (Utrecht, Netherlands) 2011 – 2012
Advisor: Suzan Rooijakkers, Department of Medical Microbiology
Title: Visiting Ph.D. Student, Whitaker International Fellow
- Performed experimental evaluation of computationally-designed molecular therapeutic candidates targeted against *Staphylococcus aureus* infection using ELISA
- Explored possible new immune activation mechanism involving coagulation factors and platelets
- Skills: ELISA, SDS-PAGE, western blot, flow cytometry, bacterial culture, protein expression, protein purification, immunoprecipitation

Academia Sinica (Taipei, Taiwan) 2011
Advisor: Jung-Hsin Lin, Research Center for Applied Sciences
Title: Visiting Ph.D. Student, NSF EAPSI Fellow
- Performed virtual high-throughput screening for small molecular inhibitors of interactions between staphylococcal virulence factors and complement proteins
- Skills: molecular dynamics simulations, homology modeling, pharmacophore modeling, virtual screening/docking, R and bash scripting

GE Global Research (Niskayuna, NY) 2007
Advisor: Reginald Smith, Bioanalytics and Protein Sciences Laboratory
Title: Summer Research Intern

- Evaluated high-throughput method for identifying proper refolding conditions for recombinant inclusion body proteins
- Skills: protein expression, protein purification, protein folding, SDS-PAGE, western blot

Johns Hopkins Medical Institute (Baltimore, MD) 2006 – 2007
Advisor: Peter Maloney, Department of Physiology

- Examined role of charged residues in *Oxalobacter formigenes* to determine relative effects on oxalate transport activity *in vitro*
- Skills: protein expression, SDS-PAGE, western blot, membrane reconstitution

Publications:

10. **Gorham Jr RD**, Forest DL, Tamamis P, López de Victoria A, Kraszni M, Kieslich CA, Banna CD, Bellows-Peterson ML, Larive CK, Floudas CA, Archontis G, Johnson LV, Morikis D. Novel compstatin family peptides inhibit complement activation by drusen-like deposits in human retinal pigmented epithelial cell cultures. *Experimental Eye Research* 2013, accepted for publication.
9. Kieslich CA, Tamamis P, **Gorham Jr RD**, López de Victoria A, Sausman N, Archontis G, Morikis D. Exploring protein-ligand and protein-protein interactions in the immune system using molecular dynamics and continuum electrostatics. *Current Physical Chemistry* 2012, 2, 324-343.
8. Tamamis P, López de Victoria A, **Gorham Jr RD**, Pierou P, Bellows M, Floudas CA, Morikis D, Archontis G. Molecular dynamics-based drug design of compstatin new generation analogs. *Chemical Biology and Drug Design* 2012, 79, 703-718.
7. **Gorham Jr RD**, Kieslich CA, Morikis D. Complement inhibition by Staphylococcus aureus: Electrostatics of C3d-EfbC and C3d-Ehp association. *Cellular and Molecular Bioengineering* 2012, 5, 32-43.
6. **Gorham Jr RD**, Kieslich CA, Nichols A, Sausman NU, Foronda M, Morikis D. An evaluation of Poisson-Boltzmann free energy calculations through comparison with experimental mutagenesis data. *Biopolymers* 2011, 95, 746-754.
5. El Assaad AM, Kieslich CA, **Gorham Jr RD**, Morikis D. Electrostatic exploration of the C3d-FH4 interaction using a computational alanine scan. *Molecular Immunology* 2011, 48, 1844-1850.
4. Hakkoymaz H, Kieslich CA, **Gorham Jr RD**, Gunopulos D, Morikis D. Electrostatic similarity determination using multi-resolution analysis. *Molecular Informatics* 2011, 30, 733–746
3. López de Victoria A, **Gorham Jr RD**, Bellows ML, Ling J, Lo DD, Floudas CA, Morikis D. A new generation of potent complement inhibitors of the compstatin family. *Chemical Biology and Drug Design* 2011, 77, 431-440.
2. **Gorham Jr RD**, Kieslich CA, Morikis D. Electrostatic clustering and free energy calculations provide a foundation for protein design and optimization. *Annals of Biomedical Engineering* 2011, 39, 1252-1263.
1. Kieslich CA, **Gorham Jr RD**, Morikis D. 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, 707-716.

Awards and Honors:

- UCR Dissertation Year Fellowship, 2012-2013 academic year
- Utrecht University Short-Stay Fellowship, Summer 2012
- Whitaker International Fellowship, 2011-2012 academic year
- UCR Graduate Research Mentorship Program Fellowship, Fall 2011
- NSF East Asia and Pacific Summer Institute (EAPSI) Fellowship, Summer 2011
- Outstanding Teaching Assistant Award, Bioengineering – 2010-2011 Academic Year
- Outstanding Student Award, Orange County Engineering Council, February 21, 2011
- Associate Faculty Member, Faculty of 1000 Biology – Post-publication peer review of top scientific literature, Fall 2010-present
- UCR Graduate Research Mentorship Program Fellowship, Fall 2010
- Membership Officer, Society for Advancement of Chicanos and Native Americans in Science – Community outreach events and academic mentorship of undergraduate students, 2010-2011
- First Place Outstanding Poster Award – 11th Annual UC Systemwide Bioengineering Symposium, June 17-19, 2010.

Conferences:

- Gorham Jr RD, Rodriguez W, Morikis D. Molecular Analysis of the Interaction Between Staphylococcal Virulence Factor Sbi-IV and Complement C3d. Poster Presentation, 14th Annual UC Systemwide Bioengineering Symposium, June 19-21, 2013.
- **Gorham Jr RD**, Morikis D. Development of peptidic inhibitors against Staphylococcus aureus infection. Poster Presentation, AAPS National Biotechnology Conference, May 16-18, 2011, San Francisco, California, USA.
- **Gorham Jr RD**, López de Victoria A, Tamamis P, Bellows M, Floudas CA, Archontis G, Morikis D. Molecular Dynamics Simulations give insight into Structure-Activity Relations and Species Specificity of Compstatin-derived Peptides. Oral Presentation, 241st ACS National Meeting & Exposition, March 27-31, 2011, Anaheim, California, USA
- **Gorham Jr RD**, Kieslich CA, Nichols A, Sausman NU, Foronda M, Morikis D. Calculation of electrostatic free energy of protein association using Poisson-Boltzmann electrostatics: Validation with experimental kinetic data. Poster Presentation, 11th Annual UC Systemwide Bioengineering Symposium, June 17-19, 2010.
- **Gorham Jr RD**, Kieslich CA, Morikis D. Electrostatic analysis of C3d/Efb-C interaction. Oral Presentation, 10th Annual UC Systemwide Bioengineering Symposium, June 19-21, 2009.

Teaching Experience:

Research Mentor (BioMoDeL Research Group) 2009 – 2013

- Mentored graduate and undergraduate students, supervised research projects, trained students to use tools and protocols developed in our lab, assisted in running simulations and interpreting and analyzing data

Teaching Assistant for BIEN 175 2011

- Bioengineering Senior Design (2 quarters) – Worked with course professor to organize the course, graded homework assignments and reports, provided individual help to student groups

Teaching Assistant for BIEN 010 2008 – 2009

- Overview of Bioengineering (2 quarters) – Worked with other teaching assistants to develop and improve introductory bioengineering course, led lab section weekly, held office hours twice per week, helped students develop presentation skills, graded homework assignments, provided individual help to students

Teaching Assistant for BIEN 120 2009

- Biosystems and Signals Analysis (1 quarter) – Helped prepare course material, held office hours twice per week, led discussion section weekly, graded homework assignments, proctored exams