Title
Mapping the Druggability of Complement C3 and its Derivatives, and Inhibitor Design

Permalink
https://escholarship.org/uc/item/3nd7p20b

Author
Mohan, Rohith

Publication Date
2018

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Mapping the Druggability of Complement C3 and its Derivatives,
and Inhibitor Design

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

Doctor of Philosophy

in

Bioengineering

by

Rohith R. Mohan

December 2018

Dissertation Committee:

    Dr. Dimitrios Morikis, Chairperson
    Dr. Xin Ge
    Dr. Valentine Vullev

The Dissertation of Rohith R. Mohan is approved:

_____

_____

_____
Committee Chairperson

University of California, Riverside

Thanks to my other funding sources, the National Institutes of Health and the BrightFocus foundation, without which my work would not be possible. I would also like to thank the consultants and patrons of GradQuant for a memorable couple of years that both exposed me to the variety of research at UCR and allowed me to help others achieve their research goals. I am grateful to Dr. Yelda Serin for her guidance and am excited to see the new heights GradQuant reaches under her supervision.

Finally, I would like to thank my friends in and outside UCR that helped me through ups and downs, both personal and professional, over the past 5 years. I am also eternally grateful to my parents and my brother for being supportive and motivating me over the years, I would not be here without them.

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

1. Mohan, R. R., Gorham Jr., R. D., and Morikis, D. A theoretical view of the C3d:CR2 binding controversy. *Molecular Immunology*, 64(1):112–122, March 2015

2. Mohan, R. R., Huber, G. A., and Morikis, D. Electrostatic Steering Accelerates C3d:CR2 Association. *J. Phys. Chem. B*, April 2016

3. Mohan, R. R., Cabrera, A. P., Harrison, R. E., Gorham, R. D., Johnson, L. V., Ghosh, K., and Morikis, D. Peptide redesign for inhibition of the complement system: Targeting age-related macular degeneration. *Molecular Vision*, 22:1280–1290, 2016

4. Mohan, R. R., Wilson, M., Gorham, R. D., Harrison, R. E. S., Morikis, V. A., Kieslich, C. A., Orr, A. A., Coley, A. V., Tamamis, P., and Morikis, D. Virtual Screening of Chemical Compounds for Discovery of Complement C3 Ligands. *ACS Omega*, 3(6): 6427–6438, June 2018

5. Harrison, R. E. S., Mohan, R. R., Gorham, R. D., Kieslich, C. A., and Morikis, D. AESOP: A Python Library for Investigating Electrostatics in Protein Interactions. *Biophysical Journal*, 112(9):1761–1766, May 2017

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

DEDICATION

I dedicate this dissertation to my family for their unconditional love and support.

ABSTRACT OF THE DISSERTATION

Mapping the Druggability of Complement C3 and its Derivatives,
and Inhibitor Design

by

Rohith R. Mohan

Doctor of Philosophy, Graduate Program in Bioengineering
University of California, Riverside, December 2018
Dr. Dimitrios Morikis, Chairperson

The complement system, consisting of several plasma proteins and cell-bound receptors, is an important part of innate immunity. Complement component C3 is central to the multiple complement pathways and its cleavage results in the activation of various cascades for the generation of effector proteins and complexes as a response to injury or infection. A lack of regulation in this response can result in adverse effects such as inflammatory and autoimmune diseases. Despite complement's role in several autoimmune and inflammatory diseases, there is a noted dearth of complement-targeted therapeutics on the marker. In this thesis, we outline a variety of studies in which we employ computational and experimental methods to characterize the mechanistic properties and dynamics of C3 and its derivatives in order to map their druggability. We explore the mechanisms driving the C3d:CR2 interaction using electrostatic analysis, molecular dynamics simulations (both steered and explicit solvent), and MM-GBSA calculations. We investigate the role of electrostatic steering in the C3d:CR2 interaction through Brownian Dynamics simulations and confirm the ionic strength-dependence of the interaction as well as gain additional insights into the amino

acids driving the association of the C3d:CR2 complex. We utilize the insights into significant intermolecular interactions gained from the C3d:CR2 complex for the design of CR2-based peptides targeting C3d for biomarker and inhibitor development. In continuing with the pursuit of C3d-binding ligands, we implement a virtual screening workflow for the identification of small molecules with fluorescence properties for potential theranostic applications. In addition, we explore the dynamics of the C3d:CR3 complex to further characterize the structural and physicochemical properties of C3d and identify amino acids crucial to the interaction through molecular dynamics (both explicit and steered) and electrostatic analysis. We perform redesign of a compstatin family peptidic inhibitor targeting C3, and small molecule biomarker discovery at the site of compstatin-C3 binding using virtual screening. Our results provide a fundamental mechanistic understanding of the physicochemical, structural and dynamic properties of C3, and form the foundation for the development of potential diagnostic imaging molecular sensors and therapeutics.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The complement system

The complement system, an important part of innate immunity, consists of several proteins and cell surface receptors and responds to pathogens or injury through opsonization, lysis and chemotaxis. Complement activation can occur through three possible pathways: classical, lectin and alternative. All three pathways converge at complement component C3 leading to multiple cascades for potential responses[1].

C3 is cleaved by an enzyme, C3 convertase, into C3a and C3b. C3a, an anaphylotoxin, binds to C3a receptor C3aR on neutrophils leading to an inflammatory response, such as the release of histamines, and chemotaxis. C3b, an opsonin, covalently binds to pathogen cell surfaces leading to opsonization of the pathogen. Bound C3b is capable of forming C3 and C5 convertases and complement fragment C3d (through a series of cleavage steps with Factor I and a co-factor). Cleavage of C5 by C5 convertase results in the C5a and C5b with C5a functioning similar to C3a while C5b forms a complex with C6, C7, C8,

and multiple copies of C9, leading to cell lysis further downstream. Complement fragment C3d interacts with complement receptor CR2 leading to enhanced B-cell mediated antibody production, thus demonstrating the crucial role of complement as a link between innate and adaptive immunity.

Besides being an opsonin, C3b is also capable of forming C3 convertases serving as a positive feedback amplification loop. The amplification loop is beneficial during infection states as it allows for rapid complement response but without proper and balanced regulation, it can result in detrimental effects. The complement system's failure to perform incorrectly is implicated in many autoimmune and inflammatory diseases. At the same time, complete inhibition of the complement system can be detrimental as well as it prevents proper response to injury and infection. By targeting C3 and its derivatives in the context of the alternative pathway, we can dampen complement response while still allowing for potential activation through the other pathways (classical and lectin).

1.2 Role of complement in autoimmune and inflammatory diseases

The complement system is the first line of defense against pathogens and injury and plays an important role in the link between innate and adaptive immunity. This also makes complement response a bit of a double-edged sword as overactive complement response or a lack of regulated response can result in autoimmune and inflammatory diseases such as atypical hemolytic uremic syndrome (aHUS), rheumatoid arthritis (RA), paroxysmal nocturnal hemoglobinurea (PNH), chronic obstructive pulmonary disease (COPD), and age-

related macular degeneration (AMD)[2–4] while an underactive complement system increases susceptibility to infection. Despite complement being implicated in over 30 different disease, only two complement-targeted therapeutics are currently in the clinic, Cinryze (from Shire)[5] and Soliris (from Alexion)[6]. Both Cinryze and Soliris are protein-based therapeutics, which drives their production cost high, and they suffer from the standard limitations of protein therapeutics, such as batch variability, stability, bioavailability, and administration. Low molecular mass complement inhibitors, such as peptides and chemical compounds can be less costly to scale up for clinical production, can be administered orally or locally (as opposed to intra-venous injections), and will have higher stability and in the case of chemical compounds, bioavailability. Although, dampening complement response is beneficial in this case, it reduces the ability to respond to infection and injury. Targeting the alternative pathway can allow for retention of the ability to activate complement response through the classical and lectin pathways.

In order to develop inhibitors and biomarkers targeting C3 and its cleavage products for prevention and management of the aforementioned diseases, there must first be a thorough understanding of the physicochemical and structural profile of these complement proteins and how it affects their mechanisms of action. In the context of age-related macular degeneration, complement activation has been demonstrated to be a crucial factor in inflammation and tissue damage[7,8]. Drusen formation is the accumulation and deposition of debris at the RPE-Bruch's membrane interface and is the major symptomatic indicator of AMD. Dysregulation of complement system may not initiate drusen formation but it has been shown to contribute to drusen accumulation and further progression of AMD pathol-

ogy[9,10]. As drusen accumulation is directly related to complement overactivation, there is also deposition of terminal complement products, like C3d, in excess in AMD tissues. By characterizing the mechanistic properties and structural dynamics of C3d, a foundation for development of potential inhibitors and biomarkers can be acquired. This is not limited to C3d and is extensible to C3b and C3 as well. It also applicable to the other disorders in which complement response is implicated.

1.3 *In-silico* characterization of mechanistic properties, electrostatics and dynamics in proteins

With recent improvements and updates, computational methods have been shown to complement experimental methods in a valuable manner[11]. With appropriate parameterization, it is possible to accurately reproduce and predict binding and energetics observed in biological phenomena using computational tools. Utilizing a combination of computational and experimental methods allows us to reinforce, validate and efficiently explore solutions involving structure-dynamics-function relations and function inhibition.

The complement system is particularly amenable to computational studies due to the pervasiveness of electrostatic interactions[12–20]. It has been well established that electrostatic interactions play an important role in ramping up complement response by driving conformational changes and complex stability. Poisson-Boltzmann and Coulombic calculations provide an efficient way to characterize electrostatic potentials as they account for the solvation environment of proteins and both long and short-range contributions. With the availability of crystallographic structures of many of the components of the complement

system, it is facile to apply these methods to the available protein structures. We have also implemented our computational framework AESOP (Analysis of Electrostatic Structures Of Proteins) as a Python package in order to facilitate these types of analyses[21]. With functionality for computational alanine scanning, directed mutagenesis and electrostatic similarity and hotspot analysis, AESOP can provide valuable insight into the long and short-range contributions of residues to the electrostatic effects underlying the interaction.

As the collection of structural data at the Protein Data Bank (PDB) continues to grow, so to does our structural understanding of complement proteins[22–28]. However, crystallographic structures are inherently limited by their static nature so in order to evaluate protein interactions in a more detailed manner, molecular dynamics (MD) simulations are necessary. Simulations can provide valuable insight into the dynamic stability of protein complexes at the atomic scale and can also be used to evaluate both conformational transitions of binding, using explicit solvent MD simulations, and unbinding, using steered molecular dynamics (SMD) simulations. Through these methodologies we can extract many insights such a persistent intermolecular interactions, identification of molecular switches through timeseries of interatomic distances, and dynamic cross correlation indicating significant domain motions. Analyzing the data generated by such simulations requires creative approaches to reduce the complexity and extract biological insights and methods such as Principal Component Analysis (PCA), time-lagged Independent Component Analysis (tICA), and Markov Modeling can be valuable in these applications[29,30].

1.4 Targeting complement response: peptides and small molecules

There has been repeated evidence demonstrating the role of unregulated complement response in autoimmune and inflammatory diseases and yet there is a dearth of therapeutics targeting complement response for inhibition or biomarkers in the market. The few complement-based therapeutics that do exist are biopharmaceutical-based which often suffer limitations in bioavailability, cost of production, and medium of delivery. Peptides and small molecules can be viable alternatives as they address many of these limitations while retaining specificity. The application of a two-pronged approach incorporating both peptide design and small molecule discovery can be advantageous for inhibitor and biomarker discovery.

Depending on their bioavailability, peptides serve as a middle ground to biopharmaceuticals and small molecules due to their specificity, relatively low cost of production, and oral delivery capability. Cocrystal structures of complexes such as that of C3c (a cleavage product of C3) and Compstatin and C3d:CR2 can be the basis for peptide design. Rational design approaches can leverage existing structural information to identify potential peptidic designs incorporating secondary structural elements like β-turns to facilitate cyclization. Rational design informed by electrostatic analysis, molecular dynamics simulations and energetics analysis can guide the optimization of these designs as well. Combinatorial approaches powered by software suites like Rosetta[31,32] can be a cost- and time-effective alternative to experimental approaches like phage display, thus allowing for further refinement of the design of peptides.

As an alternative to rational peptidic design, virtual screening of chemical compounds is advantageous in that it is robust, efficient, and has the benefits of being high throughput while being more time and resource efficient than its experimental counterparts due to its in silico nature[33,34]. Small drug-like compounds have the advantage of exhibiting ideal pharmacological properties albeit depending on their specificity. Virtual screening generally comprises of two stages: pharmacophore model screening and molecular docking. A pharmacophore model is simply a framework of spheres representing the spatial distribution of physicochemical properties (aromaticity, hydrophobicity, hydrogen bond donor/acceptor capability, and positive/negative charge.) of an active ligand. Each moiety of the model is considered a pharmacophore feature and each pharmacophore model may contain several features. By integrating MD simulations with our virtual screening workflow, we can extract pharmacophore features from the MD trajectories of the ligand-receptor complex by identifying highly conserved interactions and key residues. During screening, a database of compounds is matched up against the pharmacophore model and if the compound contains matching pharmacophore features within the corresponding tolerance radii then it is considered a positive hit. Following pharmacophore screening, the subsequent positive hits are docked to the receptor of interest using molecular docking. During molecular docking, the compounds occupy a variety of "poses" representing a range of plausible conformations within a predefined region of the receptor. The fit of the compound, in its various poses, with the receptor is scored on the basis of a force-field and empirical data and the compounds are then ranked by predicted binding affinity. In addition to predicted binding affinity, the compounds can also be evaluated for predicted solubility using the partition

coefficient (logP) and can be visually inspected for proper geometric and spatial distribution. Top-ranked compounds can then be purchased for characterization of binding, using microscale thermophoresis, and complement inhibition, using ELISAs and hemolytic assays. By leveraging structural and dynamics information extracted from molecular dynamics simulations of structures of C3d[27,35], C3c[25], and their corresponding binding partners, virtual screening can be a valuable approach to identifying ligands for inhibition and biomarker targeting C3 and its derivatives.

1.5 Overview

The work described here aims to investigate the dynamics and mechanistic properties of C3 and its derivatives in order to map their druggability. Additionally, these insights are leveraged for the design and discovery of peptides and small drug-like molecules, so that they may provide a basis for the therapeutic, diagnostic, and theranostic development. In Chapter 2, we explore the binding mode controversy of C3d:CR2 through a variety of computational methodologies and electrostatic analysis. We identify amino acids participating in significant intermolecular interactions through electrostatic analysis via our computational framework AESOP and molecular dynamics simulations (both explicit solvent and steered). We provide evidence for the physiological relevance of the 2011 crystallographic structure of the C3d:CR2 complex while also identifying further evidence supporting the crystal packing energetics influence in 2001 structure. In Chapter 3, we delve deeper into the role of electrostatic interactions in the C3d:CR2 interaction through a combination of computational mutagenesis and Brownian Dynamics simulations. We confirm the role

of electrostatic steering in accelerating the C3d:CR2 interaction and the ionic strength dependence of the interaction while also confirming the previously identified electrostatic contributions of amino acids at the binding interface. In Chapter 4, we implement a virtual screening workflow to identify C3d-binding ligands with fluorescence capabilities for biomarker applications. Using the insights on the C3d:CR2 interaction from our prior studies and molecular dynamics simulations we develop robust pharmacophore models for screening, perform molecular docking, and identify a subset of small molecules demonstrating promise in *in-silico* analysis. In Chapter 5, we continue our pursuit of C3d-binding ligands for biomarker discovery through a peptidic design approach. Utilizing the wealth of knowledge on the structural dynamics of the C3d:CR2 that we previously compiled, we implement rational and combinatorial design to identify C3d-binding peptides. In Chapter 6, we explore another facet of C3d through the investigation of the C3d:CR3 interaction using molecular dynamics, steered molecular dynamics, and electrostatic analysis. Through our investigation, we identify the mechanisms driving the interaction while also provide a basis for exploring peptide design through the non-overlapping yet adjacent CR2- and CR3-binding sites of C3d. In Chapter 7, we utilize a peptide design approach for targeting C3, specifically in the context of age-related macular degeneration. Through a combination intermolecular interaction analysis and prior work, we utilize peptide redesign for the development of a PEGylated Compstatin analog peptide with increased solubility and inhibitory activity in hemolytic assays and human RPE cell-based assays. Finally, in Chapter 8, we describe a virtual screening workflow for the identification of C3 binding small molecules. Through the development of a diverse and robust set of pharmacophore models, we iden-

tified a subset of small molecules with promising predicted binding affinties, from which we confirmed the binding capability of one small molecule and provide a basis for further development for potential biomarker applications. In addition to providing a thorough understanding of basic immunology at a molecular level, these studies will also provide a basis for the discovery of new complement inhibitors and biomarkers.

1.6   References

[1] Zipfel, P. F. and Skerka, C. Complement: The Alternative Pathway. *eLS*, 2001.

[2] Ricklin, D. and Lambris, J. D. Complement in Immune and Inflammatory Disorders: Pathophysiological Mechanisms. *The Journal of Immunology*, 190(8):3831–3838, April 2013. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.1203487. 00107.

[3] Liszewski, M. K. and Atkinson, J. P. Complement regulators in human disease: Lessons from modern genetics. *Journal of Internal Medicine*, 277(3):294–305, March 2015. ISSN 1365-2796. doi: 10.1111/joim.12338.

[4] Morgan, B. P. and Harris, C. L. Complement, a target for therapy in inflammatory and degenerative diseases. *Nature Reviews Drug Discovery*, 14(12):857–877, October 2015. ISSN 1474-1776, 1474-1784. doi: 10.1038/nrd4657. 00002.

[5] Cocchio, C. and Marzella, N. Cinryze, a Human Plasma-Derived C1 Esterase Inhibitor for Prophylaxis Of Hereditary Angioedema. *P T*, 34(6):293–328, June 2009. ISSN 1052-1372.

[6] Kim, J. S., Lee, J. W., Kim, B. K., Lee, J.-H., and Chung, J. The use of the complement inhibitor eculizumab (Soliris®) for treating Korean patients with paroxysmal nocturnal hemoglobinuria. *Korean J Hematol*, 45(4):269–274, December 2010. ISSN 1738-7949. doi: 10.5045/kjh.2010.45.4.269.

[7] Gehrs, K. M., Anderson, D. H., Johnson, L. V., and Hageman, G. S. Age-related macular degeneration—emerging pathogenetic and therapeutic concepts. *Annals of Medicine*, 38(7):450–471, January 2006. ISSN 0785-3890, 1365-2060. doi: 10.1080/07853890600946724.

[8] Troutbeck, R., Al-Qureshi, S., and Guymer, R. H. Therapeutic targeting of the complement system in age-related macular degeneration: A review: Complement and AMD. *Clinical & Experimental Ophthalmology*, 40(1):18–26, January 2012. ISSN 14426404. doi: 10.1111/j.1442-9071.2011.02581.x.

[9] Anderson, D. H., Radeke, M. J., Gallo, N. B., Chapin, E. A., Johnson, P. T., Curletti, C. R., Hancox, L. S., Hu, J., Ebright, J. N., Malek, G., Hauser, M. A., Rickman, C. B., Bok, D., Hageman, G. S., and Johnson, L. V. The Pivotal Role of the Complement System in Aging and Age-related Macular Degeneration: Hypothesis Re-visited. *Prog Retin Eye Res*, 29(2):95–112, March 2010. ISSN 1350-9462. doi: 10.1016/j.preteyeres.2009.11.003.

[10] Anderson, D. H., Mullins, R. F., Hageman, G. S., and Johnson, L. V. A role for local inflammation in the formation of drusen in the aging eye. *American Journal of Ophthalmology*, 134(3):411–431, September 2002. ISSN 0002-9394. doi: 10.1016/S0002-9394(02)01624-0.

[11] Morikis, D. and Lambris, J. D. Physical methods for structure, dynamics and binding in immunological research. *Trends in Immunology*, 25(12):700–707, December 2004. ISSN 1471-4906. doi: 10.1016/j.it.2004.09.009.

[12] Morikis, D. and Lambris, J. D. The Electrostatic Nature of C3d-Complement Receptor 2 Association. *J Immunol*, 172(12):7537–7547, June 2004. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.172.12.7537.

[13] Pyaram, K., Kieslich, C. A., Yadav, V. N., Morikis, D., and Sahu, A. Influence of electrostatics on the complement regulatory functions of Kaposica, the complement inhibitor of Kaposi's sarcoma-associated herpesvirus. *J. Immunol.*, 184(4):1956–1967, February 2010. ISSN 1550-6606. doi: 10.4049/jimmunol.0903261.

[14] Zhang, L. and Morikis, D. 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(9):3106–3119, May 2006. ISSN 0006-3495. doi: 10.1529/biophysj.105.068130.

[15] Kieslich, C. A., Vazquez, H., Goodman, G. N., de Victoria, A. L., and Morikis, D. The effect of electrostatics on factor H function and related pathologies. *Journal of Molecular Graphics and Modelling*, 29(8):1047–1055, August 2011. ISSN 1093-3263. doi: 10.1016/j.jmgm.2011.04.010.

[16] Gorham, R. D., Rodriguez, W., and Morikis, D. Molecular Analysis of the Interaction between Staphylococcal Virulence Factor Sbi-IV and Complement C3d. *Biophysical Journal*, 106(5):1164–1173, March 2014. ISSN 00063495. doi: 10.1016/j.bpj.2014.01.033.

[17] Kieslich, C. A. and Morikis, D. The Two Sides of Complement C3d: Evolution of Electrostatics in a Link between Innate and Adaptive Immunity. *PLoS Computational Biology*, 8(12):e1002840, December 2012. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002840.

[18] Mohan, R. R., Gorham Jr., R. D., and Morikis, D. A theoretical view of the C3d:CR2 binding controversy. *Molecular Immunology*, 64(1):112–122, March 2015. ISSN 0161-5890. doi: 10.1016/j.molimm.2014.11.006. 00002.

[19] E. S. Harrison, R., Gorham, R. D., and Morikis, D. Energetic evaluation of binding modes in the C3d and Factor H (CCP 19-20) complex. *Protein Science*, 24(5):789–802, May 2015. ISSN 1469-896X. doi: 10.1002/pro.2650. 00000.

[20] Gorham, R. D., Kieslich, C. A., and Morikis, D. Complement Inhibition by Staphylococcus aureus: Electrostatics of C3d–EfbC and C3d–Ehp Association. *Cel. Mol. Bioeng.*, 5(1):32–43, March 2012. ISSN 1865-5025, 1865-5033. doi: 10.1007/s12195-011-0195-6.

[21] Harrison, R. E. S., Mohan, R. R., Gorham, R. D., Kieslich, C. A., and Morikis, D. AESOP: A Python Library for Investigating Electrostatics in Protein Interactions. *Biophysical Journal*, 112(9):1761–1766, May 2017. ISSN 0006-3495. doi: 10.1016/j. bpj.2017.04.005.

[22] Lea, S. M. and Johnson, S. Putting the structure into complement. *Immunobiology*, 217(11):1117–1121, November 2012. ISSN 01712985. doi: 10.1016/j.imbio.2012.07.005.

[23] Janssen, B. J. C., Huizinga, E. G., Raaijmakers, H. C. A., Roos, A., Daha, M. R., Nilsson-Ekdahl, K., Nilsson, B., and Gros, P. Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature*, 437(7058): 505–511, September 2005. ISSN 0028-0836, 1476-4679. doi: 10.1038/nature04005.

[24] Janssen, B. J. C., Christodoulidou, A., McCarthy, A., Lambris, J. D., and Gros, P. Structure of C3b reveals conformational changes that underlie complement activity. *Nature*, 444(7116):213–216, November 2006. ISSN 0028-0836. doi: 10.1038/nature05172.

[25] Janssen, B. J. C., Halff, E. F., Lambris, J. D., and Gros, P. Structure of Compstatin in Complex with Complement Component C3c Reveals a New Mechanism of Complement Inhibition. *J. Biol. Chem.*, 282(40):29241–29247, May 2007. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.M704587200.

[26] Forneris, F., Ricklin, D., Wu, J., Tzekou, A., Wallace, R. S., Lambris, J. D., and Gros, P. Structures of C3b in Complex with Factors B and D Give Insight into Complement Convertase Formation. *Science*, 330(6012):1816–1820, December 2010. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1195821.

[27] van den Elsen, J. M. H. and Isenman, D. E. A Crystal Structure of the Complex Between Human Complement Receptor 2 and Its Ligand C3d. *Science*, 332(6029): 608–611, April 2011. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1201954.

[28] Morikis, D. F1000Prime Recommendation of [van den Elsen JM and Isenman DE, Science 2011, 332(6029):608-11]. F1000Prime.com/10371956#eval14018054, September 2011.

[29] Kohlhoff, K. J., Shukla, D., Lawrenz, M., Bowman, G. R., Konerding, D. E., Belov, D., Altman, R. B., and Pande, V. S. Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. *Nature Chemistry*, 6(1):15–21, January 2014. ISSN 1755-4330, 1755-4349. doi: 10.1038/nchem.1821.

[30] Schwantes, C. R. and Pande, V. S. Modeling Molecular Kinetics with tICA and the Kernel Trick. *Journal of Chemical Theory and Computation*, 11(2):600–608, February 2015. ISSN 1549-9618, 1549-9626. doi: 10.1021/ct5007357.

[31] Smith, C. A. and Kortemme, T. Structure-Based Prediction of the Peptide Sequence Space Recognized by Natural and Synthetic PDZ Domains. *Journal of Molecular*

Biology, 402(2):460–474, September 2010. ISSN 0022-2836. doi: 10.1016/j.jmb.2010.07.032.

[32] Smith, C. A. and Kortemme, T. Predicting the Tolerated Sequences for Proteins and Protein Interfaces Using RosettaBackrub Flexible Backbone Design. *PLOS ONE*, 6 (7):e20451, July 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0020451.

[33] Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *Journal of Pharmacological and Toxicological Methods*, 44(1):235–249, July 2000. ISSN 1056-8719. doi: 10.1016/S1056-8719(00)00107-6.

[34] Lavecchia, A. and Di Giovanni, C. Virtual screening strategies in drug discovery: A critical review. *Current medicinal chemistry*, 20(23):2839–2860, 2013. 00133.

[35] Bajic, G., Yatime, L., Sim, R. B., Vorup-Jensen, T., and Andersen, G. R. Structural insight on the recognition of surface-bound opsonins by the integrin I domain of complement receptor 3. *PNAS*, 110(41):16426–16431, October 2013. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1311261110.

Chapter 2

A mechanistic overview of the

C3d:CR2 binding controversy

2.1   Introduction

The complement system, consisting of several plasma proteins and cell surface receptors, is a crucial component of the innate immune system. Response to pathogens is carried out through several processes such as opsonization, inflammation, lysis via the membrane attack complex[1]. The interaction between complement fragment C3d and complement receptor 2 (CR2) is known to be a key component of the link between innate and adaptive immunity[2], through enhancement of B-cell mediated antibody production during initial complement response to infection[3]. For the past decade, the topology, geometric and physicochemical nature of the binding mode of C3d and CR2 has been an issue of major controversy[4].

The free structure of C3d was published in 1998, revealing a concave surface with an acidic patch, comprised of two acidic sequence segments, as a potential site of interaction to CR2[5]. A mutagenesis study in 2000 supported this hypothesis as it showed that acidic amino acid mutants in the acidic patch on C3d domain of iC3b decreased binding to CR2 in a Rosette assay, while a mutation of a basic amino acid near the acidic patch resulted in a two-fold enhancement in binding to CR2[6]. In 2001, a cocrystal structure of C3d and CR2 was published that contradicted this hypothesis (Protein Data Bank, PDB Code: 1GHQ, called hereafter the 1GHQ structure)[7]. Although the CR2 in the aforementioned structure had two N-terminal short consensus repeat (SCR) domains (known to interact with C3d[8–10]), the only interaction observed was between SCR2 and a side face (not the acidic patch) of C3d. This was different from the previously hypothesized binding site based on the free structure of C3d[5] and the 2000 mutagenesis study[6]. Some attributes of the 1GHQ cocrystal structure[7] raised concerns in the field, including discrepancies in stoichiometry, absence of salt bridges at the interface, lack of contact between C3d and SCR1, location of interface, and nonphysiological zinc ions at the binding interface, and became the subject of scrutiny in subsequent experimental and computational studies.

In 2005, a mutagenesis study on amino acids at the binding interface on CR2 lent further credence to the 1GHQ cocrystal structure of C3d:CR2[11]; however the same mutagenesis study, guided by previous epitope and peptide binding data, demonstrated that mutations of basic amino acids located on a single face of SCR1 were critical for C3d:CR2(SCR1-2) binding. Involvement of SCR1 and linker residues in binding to CR2(SCR1-2) was also supported by a subsequent NMR study using chemical shift perturbation analysis[12]. An-

other mutagenesis study published in 2010[13], confirmed the results of the 2000 mutagenesis study[6], using Surface Plasmon Resonance (SPR) instead of Rosette assay, and found that C3d mutations at the 1GHQ interface did not have a significant effect on binding. The authors of the study went on to question the physiological relevance of the cocrystal structure citing concerns brought up during the multi-year controversy. The use of 0.2 M zinc acetate in the crystallization medium was believed to be the cause of the nonphysiological zinc ions at the interface and was an area of concern regarding the cocrystal structure[14]. The authors of the 2010 study performed an ELISA study to demonstrate that zinc acetate actually abrogates the C3d:CR2(SCR1-2) interaction in solution, at conditions mimicking the crystallization pH and ionic strength. The authors of this study noted that the 1GHQ structure included two non-interaction molecules and suggested that the observed contacts were due to favorable crystal packing energetics[13].

In 2011, the authors of the 2010 mutagenesis study[13] published a new cocrystal structure of C3d:CR2(SCR1-2) (PDB Code: 3OED, called hereafter the 3OED structure) in which the binding interface is different from that of the 1GHQ structure, and is more consistent with existing biochemical data[15]. In this structure, both SCR1 and SCR2 modules of CR2 interact with the acidic patch of C3d, in agreement with the majority of the mutagenesis data.

During the course of the controversy, several computational studies were performed aiming to construct a unified model that accounts for all experimental data[16–19]. Although unambiguous construction of a unified model was not possible based on the 1GHQ structure, the computational studies proposed a two-step association process driven by long-

and short-range electrostatic interactions. A computational study in 2011 supported the claims alluding to the non-physiological nature of the crystal structure[17] and additionally, both experimental and theoretical studies have supported the possibility of another binding site[16,19–23].

The 3OED crystal structure[15] and newest mutagenesis studies[13,23] have experimentally proposed a resolution for the long-standing controversy regarding the binding mode of C3d:CR2(SCR1-2). Our study brings a theoretical basis and mechanistic insight to the C3d:CR2(SCR1-2) interaction. We use different computational tools to evaluate the electrostatic properties of the two crystal structures as well as their dynamic, energetic and dissociative properties. We demonstrate that the 3OED structure is energetically significantly more favorable and much more stable overall than the 1GHQ structure, and therefore 3OED is more likely to be the physiological structure. We also critically discuss the underlying factors of the 1GHQ structure formation, and we evaluate the potential effects of the presence of zinc ions at the binding mode of 1GHQ. Lastly, our study strongly supports the role of electrostatic interactions in the recognition and binding of C3d with CR2(SCR1-2).

2.2 Methods

2.2.1 Protein structures

The protein structures used in our analysis were acquired from the Protein Data Bank (PDB)[24] with the codes 1GHQ and 3OED for the 2001 cocrystal structure[7] and the 2011 cocrystal structure[15], respectively. We utilized chain A and chain B from the 1GHQ structure for C3d and CR2(SCR1-2) respectively and removed the two zinc ions present (one

each on chain A and chain B) and chain C which lacks contact with C3d. This complex will be referred to as 1GHQ hereafter. We also prepared a second 1GHQ complex in which the two zinc ions are not removed and this complex will be referred to as $1GHQ^{zn}$ hereafter. In the 3OED structure, there are two complexes for C3d:CR2(SCR1-2) but we used the complex with better electron density and lower B-factor values, corresponding to chain A and chain C for C3d and CR2(SCR1-2) respectively. This complex will be referred to as 3OED hereafter. Due to the discrepancy in residue numbering in the structures, we used the numbering in 3OED as our basis, which required subtracting 1 from all CR2 residue numbers in 1GHQ and $1GHQ^{zn}$.

2.2.2 Electrostatic analysis

Electrostatic effects play a crucial role in the C3d:CR2(SCR1-SCR2) interaction[16–19] so we utilized the AESOP (Analysis of Electrostatic Similarities of Proteins) framework[17,25–27] to study the effect of electrostatics on the two binding sites. The first step in our workflow was adding missing hydrogens, atomic radii and partial charges to our PDB structures using PDB2PQR[28] and the PARSE force field[29]. Next we performed computational mutagenesis on the protein complexes using prior mutants from the literature[6,11,13,15,30]. We also generated single alanine mutants of all ionizable amino acids of the complex. After the mutated protein complexes were generated, we calculated electrostatic potentials with the linearized Poisson–Boltzmann equation and electrostatic free energies of association, using APBS[31], and compared them to those of the unmodified parent protein complexes. The parameters applied to the APBS calculations, as justified in an earlier study[17,26], consist of dielectric coefficients of 20 and 78.54 for the protein and solvent dielec-

tric values, respectively, as well as ionic strengths of 0 or 150 mM. For 1GHQ and 1GHQ^{zn}, the number of grid points and mesh dimensions were set to $129 \times 97 \times 129$ and $126\,\text{Å} \times 96\,\text{Å} \times 105\,\text{Å}$, respectively, whereas for 3OED, they were set to $129 \times 97 \times 97$ and $98\,\text{Å} \times 95\,\text{Å} \times 90\,\text{Å}$, respectively.

Electrostatic free energies of association are calculated using a thermodynamic cycle, in order to remove self-energies and grid artifacts and the APBS Coulomb module is applied to incorporate nongrid-based Coulombic potentials for a more accurate representation of electrostatic free energies of association (ΔG_{assoc}) as previously described[17,25,32] and shown in Eq. 2.2

$$\Delta\Delta G_{solv} = \Delta G_{solv}^{C3d:CR2} - \Delta G_{solv}^{C3d} - \Delta G_{solv}^{CR2} \tag{2.1}$$

$$\Delta G_{assoc} = \Delta G_{coul} + \Delta\Delta G_{solv} \tag{2.2}$$

The electrostatic free energies of the mutants are represented relative to the parent protein as described in Eq. 2.3

$$\Delta G_{binding} = \Delta G_{assoc}^{mutant} - \Delta G_{assoc}^{parent} \tag{2.3}$$

where $\Delta G_{assoc}^{mutant}$ and $\Delta G_{assoc}^{parent}$ for C3d or CR2 are calculated using Eqs. 2.1, 2.2.

2.2.3 Explicit-solvent molecular dynamics simulations

We performed explicit-solvent molecular dynamics (MD) simulations for 1GHQ, 1GHQ^{zn} and 3OED in triplicate using NAMD[33] and the CHARMM forcefield[34]. The 3OED complex was solvated in TIP3P water boxes with dimensions of $83\,\text{Å} \times 83\,\text{Å} \times 83\,\text{Å}$ while

the 1GHQ and 1GHQ^{zn} complexes were solvated with box dimensions of 104 Å × 104 Å × 104 Å. The charges of the 3OED and 1GHQ systems were neutralized with sodium and chloride counterions at an ionic strength of 150 mM while 1GHQ^{zn} was neutralized with zinc and acetate counterions at an ionic strength of 0.2 M to reflect crystallization conditions. CHARMM parameters were not available for acetate ions so MATCH[35] was used to generate parameters from a PDB structure of acetate. The acetate ions were placed in the solvated structure by replacing randomly selected water molecules while satisfying certain distance criteria (>10 Å from protein). All three systems underwent 25,000 steps of conjugate gradient energy minimization followed by heating from 0 to 300 K in 62 ps with all protein atoms constrained to their positions after minimization. Following heating, the three systems were equilibrated through five stages for 50 ps/stage. In the case of 1GHQ^{zn}, the first equilibration stage ran for 1 ns to allow the acetate ions in the solvent to sufficiently equilibrate. During the first four stages of equilibration, force constants 41.84, 20.92, 8.368 and 4.184 kJ/mol/Å^2 respectively were used to harmonically constrain all protein atoms to their positions after minimization. During the final stage of equilibration all protein backbone atoms were harmonically constrained with a force constant of 4.184 kJ/mol/Å^2. Following the final equilibration stage, the production run was initiated with the following conditions: periodic boundary conditions, particle-mesh Ewald electrostatics (nonbonded interaction cutoff of 12 Å and switching distance of 10 Å), SHAKE algorithm (fixing hydrogen atom bond lengths), 2 fs integration timesteps, and Langevin pressure and temperature controls. We performed MD simulations for each of the systems in triplicate (total of nine

20 ns trajectories). An additional 100 ns simulation was performed for 1GHQ and 1GHQ^{zn} to evaluate if the systems stabilize over a longer time period.

2.2.4   Steered molecular dynamics simulations

We also performed steered molecular dynamics simulations (SMD) using NAMD. In order to have an independent starting point for the simulations, we took each of the trajectories from the explicit-solvent MD simulations and identified three representative conformations, each based on RMSD clustering, resulting in a total of 27 SMD trajectories (nine per system). First, we used UCSF Chimera[36] to calculate the normal vectors of the solvent accessible surface of C3d atoms within 8 Å of CR2. Subsequently we calculated the mean of the normal vectors and rotated each complex so that the mean normal vector was aligned with the $+z$ axis, allowing us to use the $+z$ axis as the direction of induced dissociation of CR2 from C3d. All three systems were solvated in TIP3P water boxes with boundaries from 12 Å from the protein in both the x and y directions. In order to allow for sufficient room for induced dissociation in the $+z$ direction, the boundary in the $+z$ direction is 70 Å from the protein for the 3OED complex and 1GHQ complex while that of 1GHQ^{zn} is 100 Å from the protein. All three systems were minimized and heated in the same manner as the explicit-solvent MD simulations as discussed above. Following heating, the three systems were equilibrated for 125 ps for the 3OED and 1GHQ systems and 250 ps for 1GHQ^{zn} while all proteins atoms were harmonically constrained with a force constant of 41.84 kJ/mol/Å^2. The SMD simulations were carried out using the equilibrated structures as input. During the SMD simulation, C3d residues >12 Å from CR2 were harmonically constrained to their post-equilibration positions with a force constant of 41.84 kJ/mol/Å^2,

while the center of mass of all CR2 atoms was constrained and pulled in the $+z$ direction at a constant velocity of $10\,\text{Å/ns}$.

2.2.5   Analysis of molecular dynamics simulations

Several tools were utilized in MD trajectory analysis. Hydrogen bond analysis and PDB and MD trajectory visual analysis were performed using UCSF Chimera, simulation videos as well as solvent accessible surface area (SASA) for both proteins and protein complexes were calculated using VMD[37] and several custom R scripts[38] in conjunction with the Bio3D package[39]. Interfacial SASA ($\Delta SASA$) of the trajectories (including hydrogen atoms) were calculated as described in Eq. 2.4

$$\Delta SASA = \Delta SASA_A + \Delta SASA_B - \Delta SASA_{AB} \tag{2.4}$$

The molecular mechanics-generalized Born surface area (MM-GBSA) method[40–42] was utilized to calculate association free energies for each MD trajectory, excluding the first 100 ps, as described previously[32,43]. The MM-GBSA calculations are outlined in the following equations

$$\Delta G = \Delta G_{C3d:CR2} - \Delta G_{CR2} - \Delta G_{C3d} \tag{2.5}$$

$$\Delta G_{elec} = \Delta E_{coul} + \Delta G_{elec:solv} \tag{2.6}$$

$$\Delta G_{np} = \Delta E_{vdw} + \Delta G_{np:solv} \tag{2.7}$$

$$\Delta G_{total} = \Delta G_{np} + \Delta G_{elec} \tag{2.8}$$

Free energy changes are calculated as described in Equation 2.5. The molecular mechanics Coulombic (ΔE_{coul}) and van der Waals (ΔE_{vdw}) terms are calculated using CHARMM[44] with the CHARMM27 force field and atomic radii[45]. An atomic radius of 1.47 was used for the nonphysiological zinc ion based on prior literature[46]. Generalized Born solvation ($\Delta G_{elec:solv}$) and nonpolar free energy ($\Delta G_{np:solv}$) terms were calculated using the CHARMM generalized Born approximation and by multiplying the interfacial surface area of the C3d:CR2 complex by a surface tension parameter of $0.005\,\mathrm{kcal/mol/\mathring{A}^2}$ respectively.

2.3 Results

2.3.1 Electrostatic calculations

In the following analysis, the $\Delta G_{binding}$ values are represented as described in Eq. 2.3. A positive $\Delta G_{binding}$ value corresponds to loss of binding due to mutation of the corresponding residue (to alanine unless specified otherwise) or in other words, a residue with a significant contribution to binding. The inverse also holds true, a negative $\Delta G_{binding}$ value corresponds to an increase in binding due to mutation of the corresponding residue.

The binding modes of 3OED and 1GHQ$^{\mathrm{zn}}$, the location of the SCR1 and SCR2 domains, and electrostatic complementarity of the structures can be visualized in Fig. 2.1. The interface at 3OED (Fig. 2.1C) demonstrates much better electrostatic complementarity than that of 1GHQ (2.1B). The results of our computational mutagenesis study using mutants from prior literature[6,11,13,15,30] are plotted in Fig. 2.2 along with previous experimental binding data labeled on the x-axis. It is apparent that 3OED is electrostatically

significantly more favorable than the 1GHQ and 1GHQ^{zn} structures based on the large number of residues displaying a significant increase in binding energy as well as the magnitudes of the increases. The disparity between the structures is especially evident when looking at the number of CR2 residues with increases in electrostatic free energy of binding. In 1GHQ, there are no individual residues exhibiting an increase in electrostatic free energy of binding $>5\,$kJ/mol while in 1GHQ^{zn}, there are a few residues on both C3d and CR2 that exhibit a significant increase in electrostatic free energy of binding and a few CR2 residues also exhibit a loss $>5\,$kJ/mol in electrostatic free energy of binding. In contrast, many CR2 residues destabilize binding in 3OED. It is important to note the net charges of the structures (1GHQ: $4e$, 1GHQ^{zn}: $8e$, 3OED: $7e$), as part of the difference between 1GHQ and 1GHQ^{zn} may be due to the charge discrepancy owed to the presence of nonphysiological zinc ions; therefore the zinc ions have an effect on binding. This effect is more obvious in the case of CR2. This discrepancy in electrostatic free energy of binding between the structures is highlighted In Fig. 2.3, where the free energies of the three systems are mapped to the surfaces of the complexes and colored correspondingly. In this representation, we observe that the 3OED binding interface shows many more residues with large contributions to binding in comparison to the 1GHQ and 1GHQ^{zn} structures.

Results of a general alanine scan of all ionizable amino acids of 3OED, 1GHQ, and 1GHQ^{zn} can be seen in Figs. A.1 and A.2. In 3OED, we observe several residues at the interface with increases in electrostatic free energy of binding $>5\,$kJ/mol with some CR2 residues exhibiting effects contradictory to those of the 1GHQ and 1GHQ^{zn} structures. As
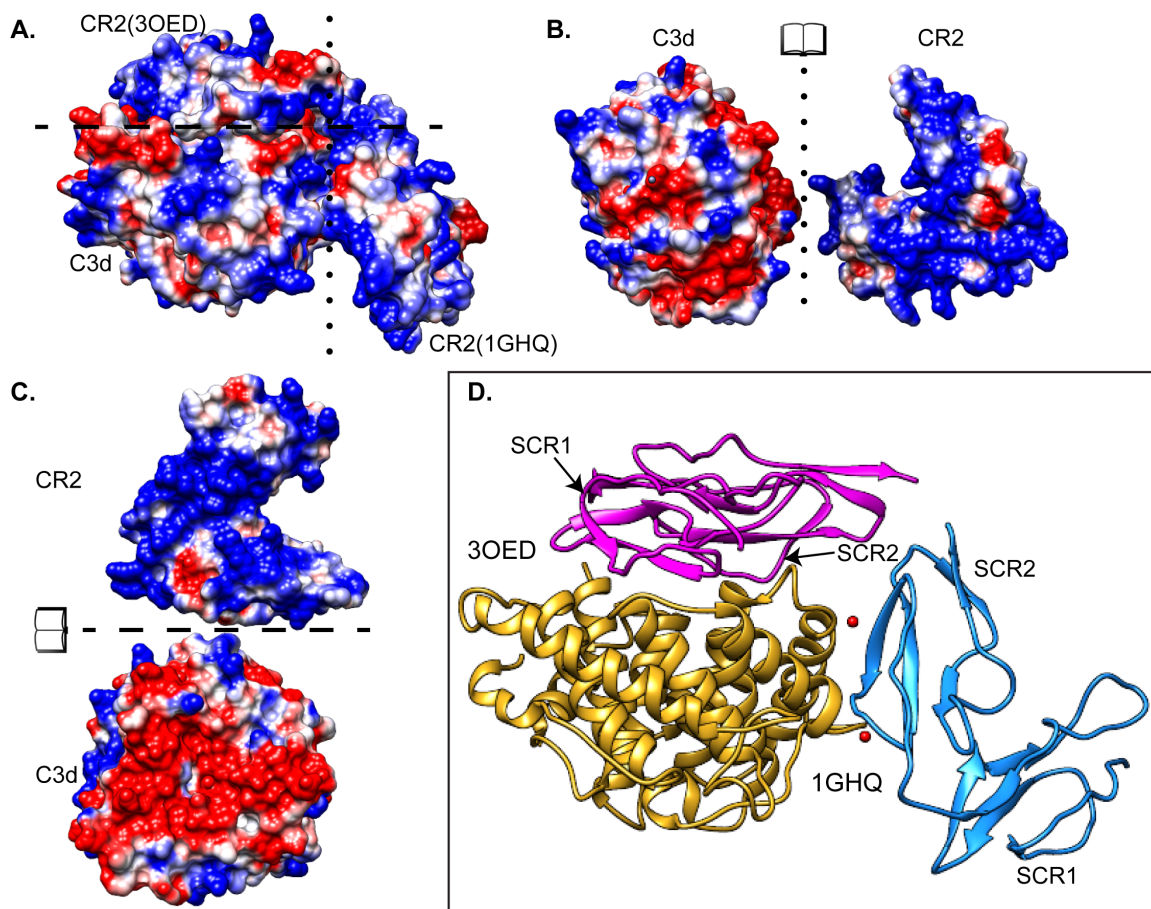
Figure 2.1: Electrostatic complementarity of the two structures and location of binding modes. (A) Both 1GHQ and 3OED C3d structures are superimposed to show the two different binding modes, and colored by electrostatic potentials. Red and blue indicate negative and positive electrostatic potentials, respectively, projected on the protein surfaces at $\pm 10 kBT/e$. (B) 1GHQ structure in open book representation splayed across the dotted y-axis. (C) 3OED structure in open book representation splayed across the dashed x-axis. (D) Both 1GHQ and 3OED C3d structures superimposed in ribbon form with the CR2 SCR1 and SCR2 domains labeled. The superimposed C3d is colored green while, 3OED CR2 is colored magenta and 1GHQ CR2 is colored cyan. Two small red spheres show the location of the nonphysiological zinc ions at the interface.
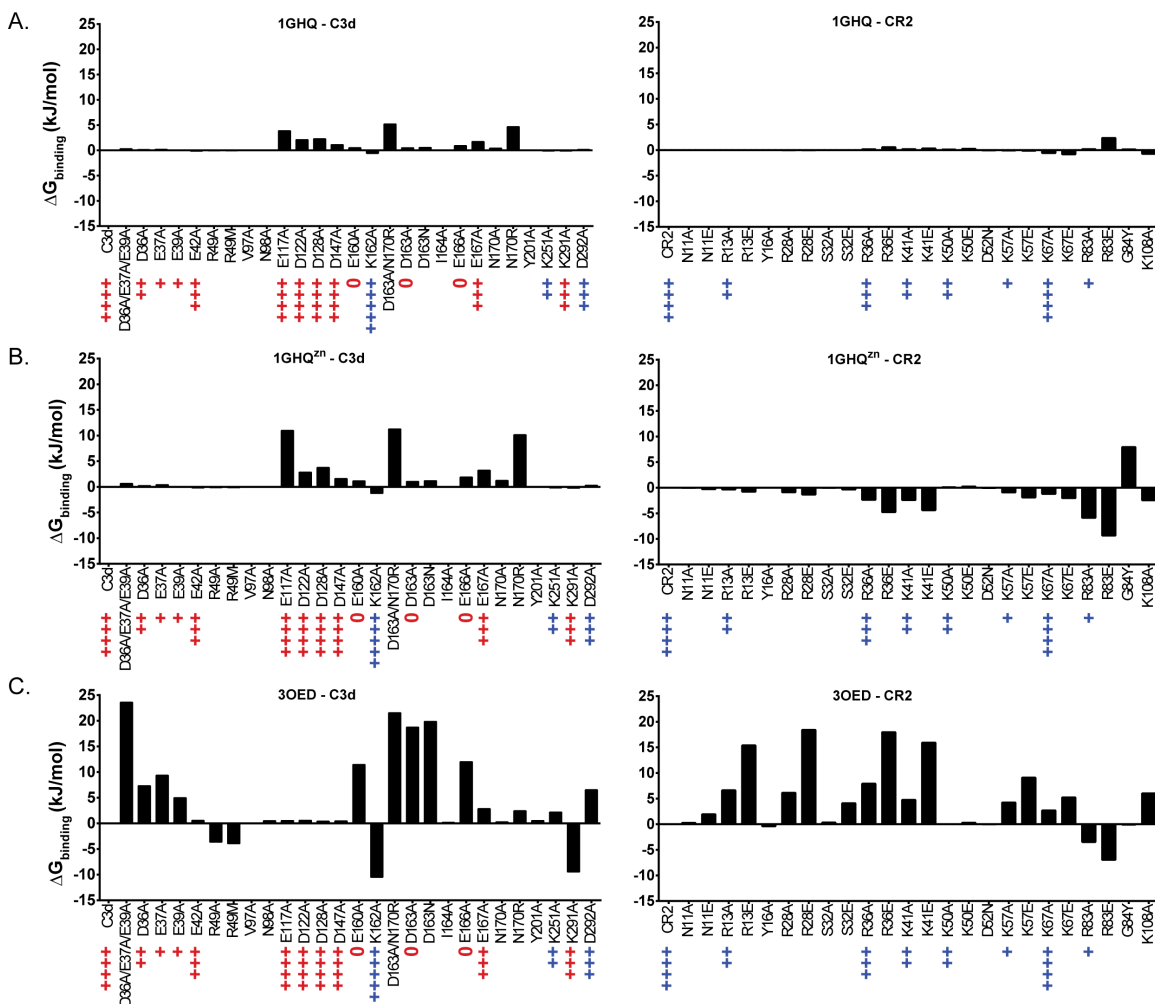
Figure 2.2: Results of electrostatic analysis using previous experimentally generated mutants. Changes in free energies of binding upon mutation are calculated using Eq. 2.3. A positive value denotes loss of binding, indicating that the mutated residue favors binding. A negative value denotes gain of binding, indicating that the mutated residue opposes binding. Binding data from previous literature can be found below the x-axis labels in the form of blue and red "+" marks corresponding to Rosette iC3b binding studies[6] and SPR C3d binding[13] respectively. The key to binding data is as follows: +++++: 200%, ++++: 90–120%, +++: 70–90%, ++: 40–70%, +: 20–40%, 0: 0–20%. The mutations without "+" marks were selected from (Hannan et al., 2005, Toapanta et al., 2010)[11,30].
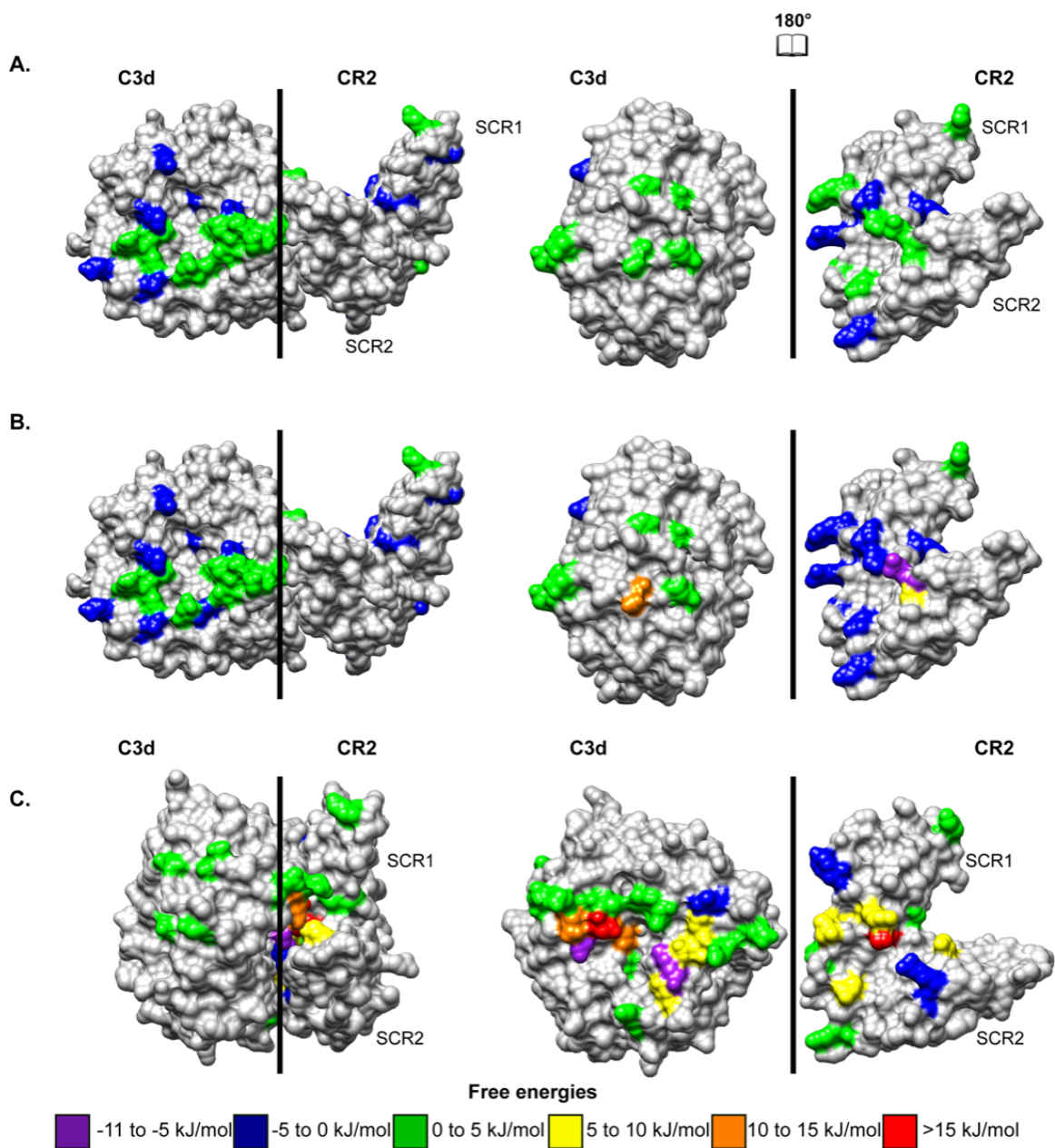
Figure 2.3: Mapping of the data of Fig. 2.2 on the surfaces of the 1GHQ (A), $1GHQ^{zn}$ (B), and 3OED (C) structures. The color code is described in the legend and represents electrostatic free energies of association.

with the computational mutagenesis study discussed above, in the case of 1GHQ, there are no residues with $>5\,\text{kJ/mol}$ increase in electrostatic free energy of binding. Similar trends are observed in 1GHQ$^{\text{zn}}$ but amplified, with the exception of one CR2 residue (D92) exhibiting a significant reversal in effect. The results are mostly in agreement with the computational mutagenesis study described above as there is a large overlap in mutants.

2.3.2   MD simulations

Crystallographic structures can be limiting due to their static nature so in order to evaluate the dynamic stability of the three systems, we performed MD simulations and analyzed the resulting trajectories. We plotted the root-mean-square deviation (RMSD) of atomic positions in Fig. 2.4 to visualize the fluctuations of the structure throughout the simulation. Fig. 2.4A–C shows the RMSD of the triplicate $20\,\text{ns}$ simulations for each system. In the case of 3OED, the RMSD stabilizes approximately around the $1\,\text{ns}$ mark, whereas both 1GHQ and 1GHQ$^{\text{zn}}$ never quite stabilize and continue to fluctuate through the $20\,\text{ns}$ simulation. As a result, we extended a simulation each from 1GHQ and 1GHQ$^{\text{zn}}$ to $100\,\text{ns}$ to evaluate if the stabilization occurs later. Based on the RMSD plots of the extended simulations, as seen in Fig. 2.4D, both 1GHQ and 1GHQ$^{\text{zn}}$ continue to fluctuate without stabilizing.

To compare the stability of the interfaces in the crystallographic structures, we measured the changes in interfacial SASA over the duration of the MD simulation. In Fig. 2.5, the interfacial SASA of each system's triplicate $20\,\text{ns}$ simulations were averaged and plotted, with their upper and lower-bound SEM shaded. Although all three systems
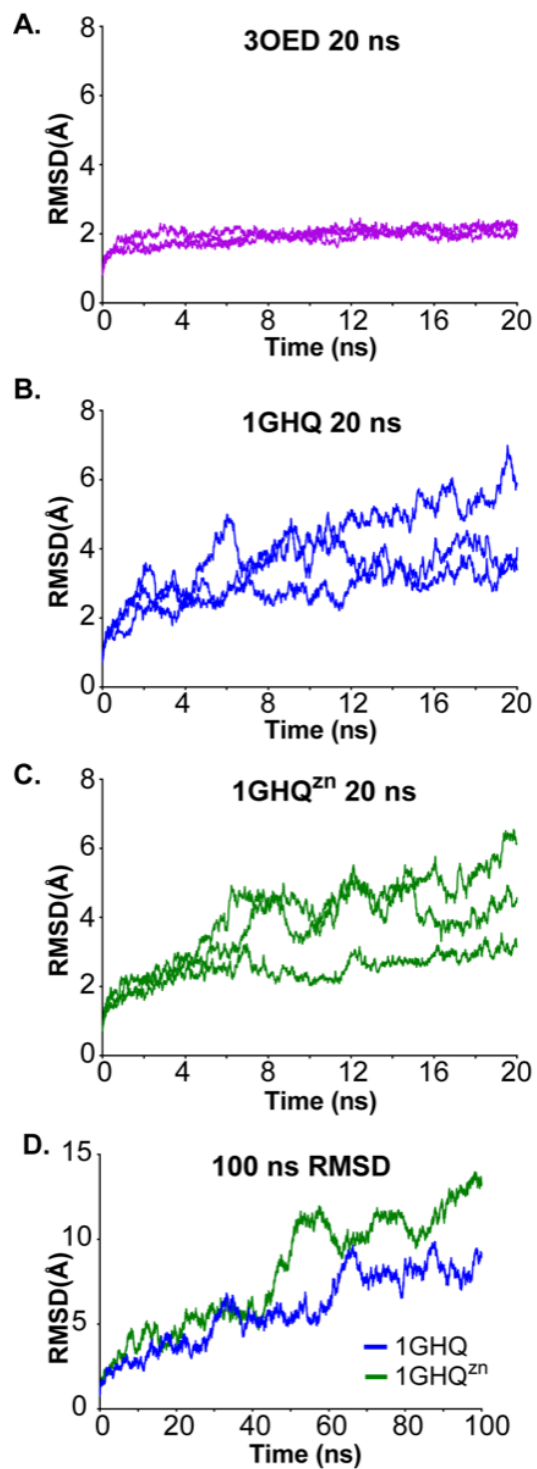
Figure 2.4: Plots of RMSD from the MD trajectories. (A–C) show a comparison of the 20 ns triplicate trajectories while (D) shows a comparison of the two 100 ns trajectories.
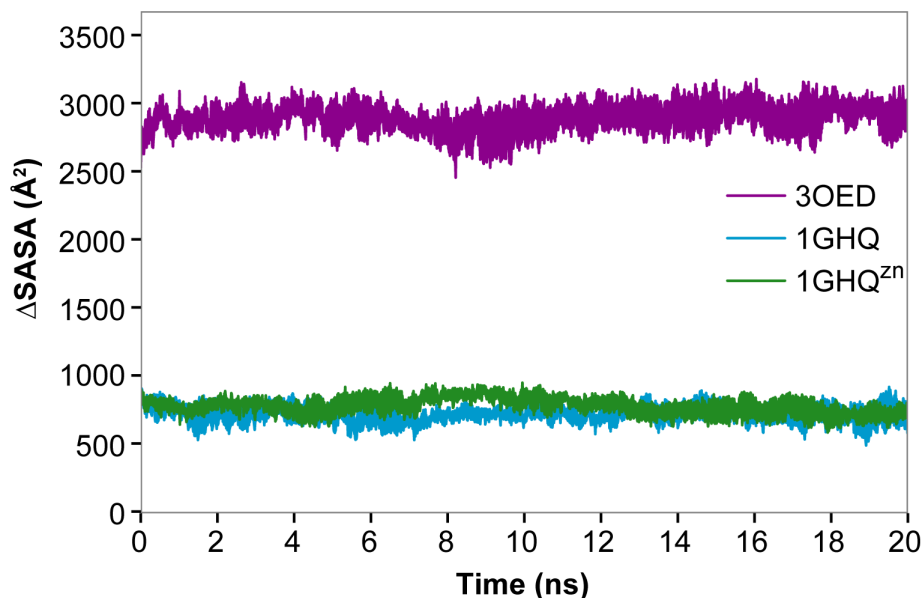
Figure 2.5: Interfacial SASA averaged over the triplicate 20 ns simulations for each system and plotted with shaded SEM. The color code is as follows: magenta for 3OED, blue for 1GHQ, and green for $1GHQ^{zn}$.

display fairly constant interfacial SASA over time, 3OED has a much higher interfacial SASA (2500—3000 Å) than the other systems (500—800 Å).

Fig. 2.6 shows the percent occupancy plots for intermolecular interactions, calculated for the triplicate simulations in each system. 3OED has a larger number of non-polar interactions (Fig. 2.6C), salt bridges (Fig. 2.6F), and hydrogen bonds (Fig. 2.6I) and also has higher percent occupancies in the aforementioned intermolecular interactions compared to 1GHQ and $1GHQ^{zn}$. $1GHQ^{zn}$ has if not more, at least higher percent occupancies in hydrogen bonds and non-polar interactions.

MM-GBSA calculations were performed in order to provide further insight into the energetics of the three systems. The frequency distribution of the electrostatic, nonpolar and total free energies of the triplicate trajectories of each system can be seen in Fig.
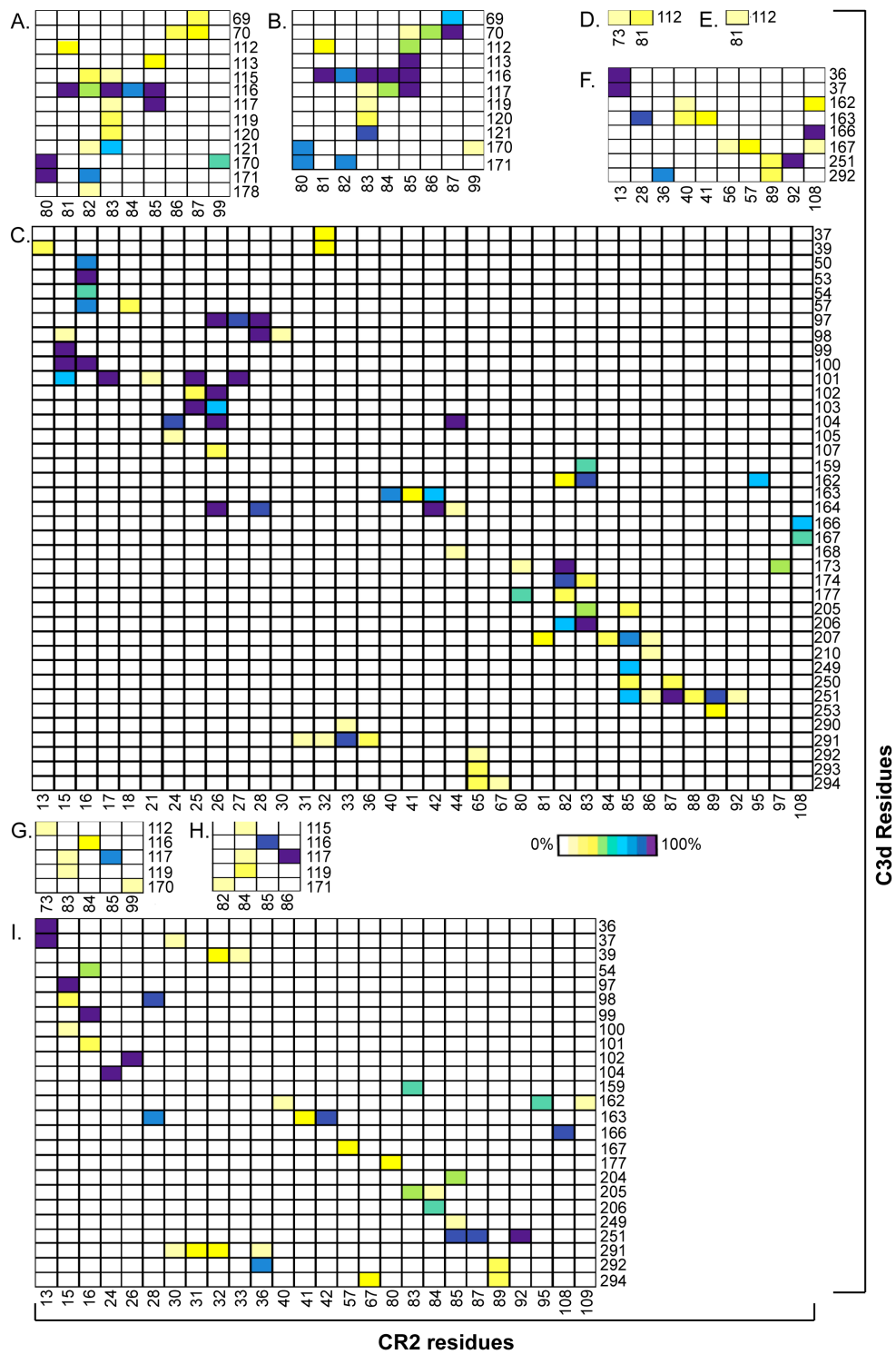
Figure 2.6: Occupancy plots from the 20 ns MD trajectories for non-polar interactions (A–C), salt bridges within 5 Å (D–F), and hydrogen bonds (G–I) for 1GHQ, 1GHQ^{zn}, and 3OED respectively.
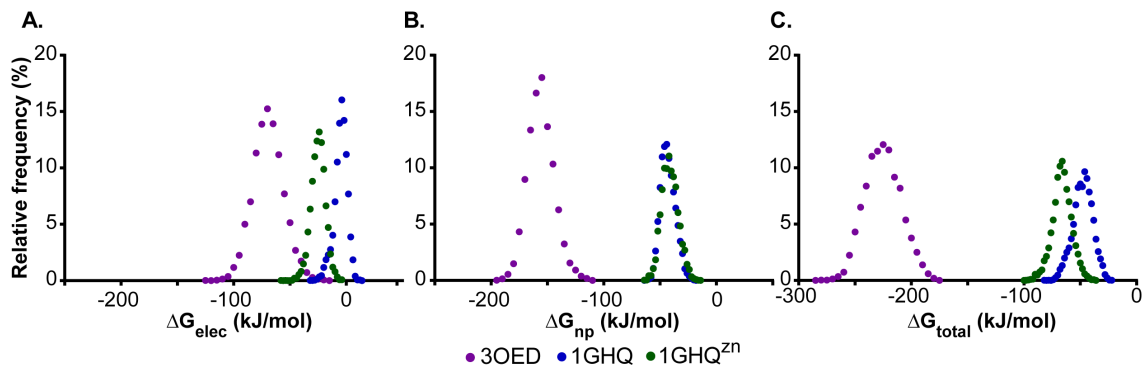
Figure 2.7: Relative frequency histograms of MM-GBSA free energies: (A) Electrostatic free energies. (B) Nonpolar free energies. (C) Total free energies.

2.7. The importance of electrostatic interactions (Fig. 2.7A) and nonpolar interactions (Fig. 2.7B) in the 3OED interface is confirmed and the overall energetic favorability (more negative MM-GBSA energies) of 3OED is exemplified by these results (Fig. 2.7C). We observe that 1GHQ and 1GHQ^{zn} demonstrate similar nonpolar free energies to each other (Fig. 2.7B) but 1GHQ^{zn} is more favorable in terms of electrostatic free energies (Fig. 2.7A), confirming the role of the nonphysiological zinc ions in electrostatic effects. Based on the criteria of RMSD, interfacial SASA, intermolecular interaction occupancies, and MM-GBSA calculations, 3OED is much more physiologically favorable than 1GHQ and 1GHQ^{zn}, and the nonphysiological zinc ions appear to be crucial to the 1GHQ^{zn} interface.

2.3.3   SMD simulations

In the SMD simulations[47,48], we induced dissociation to examine the C3d:CR2 interactions as they unbind. During the SMD simulations, the center of mass of CR2 is pulled away from C3d at a constant velocity and the resultant forces acting on CR2 are plotted over time. In Fig. 2.8, the force-time curves for each system were averaged and

plotted with their upper and lower-bound SEM shaded. For the majority of the 3OED SMD trajectories, the SCR2 domain dissociated from C3d around 0.3—0.6 ns but the time at which complete dissociation of CR2 occurred varied from 1.1 to 1.7 ns (Movie S1 of ref[49]). There was one trajectory where CR2 completely dissociated by 0.8 ns and two trajectories where CR2 did not completely dissociate by the end of the trajectory but that only further supports that the 3OED structure has an energetically highly favorable binding interface. In the case of 1GHQ, CR2 experienced complete dissociation from C3d within 0.6—0.7 ns in most of the simulations (Movie S2 of ref[49]). There was one trajectory (out of 9) where CR2 did not dissociate until 1 ns into the simulation. None of the 1GHQ^{zn} simulations exhibited complete dissociation by 2.4 ns and CR2 continued to stay connected to C3d via a nonphysiological zinc ion. As a result, we extended the 1GHQ^{zn} simulations to 3.5 ns and even then we only observed complete dissociation in one simulation at 3.2 ns. In the majority of the 1GHQ^{zn} simulations, a C3d (E117) and a CR2 (D92) residue remained connected via a nonphysiological zinc ion (Movie S3 of ref[49]). As the simulation progressed, this connection remained persistent as CR2 was stretched. Based on these results, it is apparent that the interface in 3OED is significantly more favorable energetically than 1GHQ. The amount of force required to separate just the SCR2 domain from C3d is higher in 3OED than in the other structures. We can also conclude that the presence of zinc does indeed play a role in the interface of 1GHQ based on the lack of complete dissociation and the disparity in forces between 1GHQ and 1GHQ^{zn}.
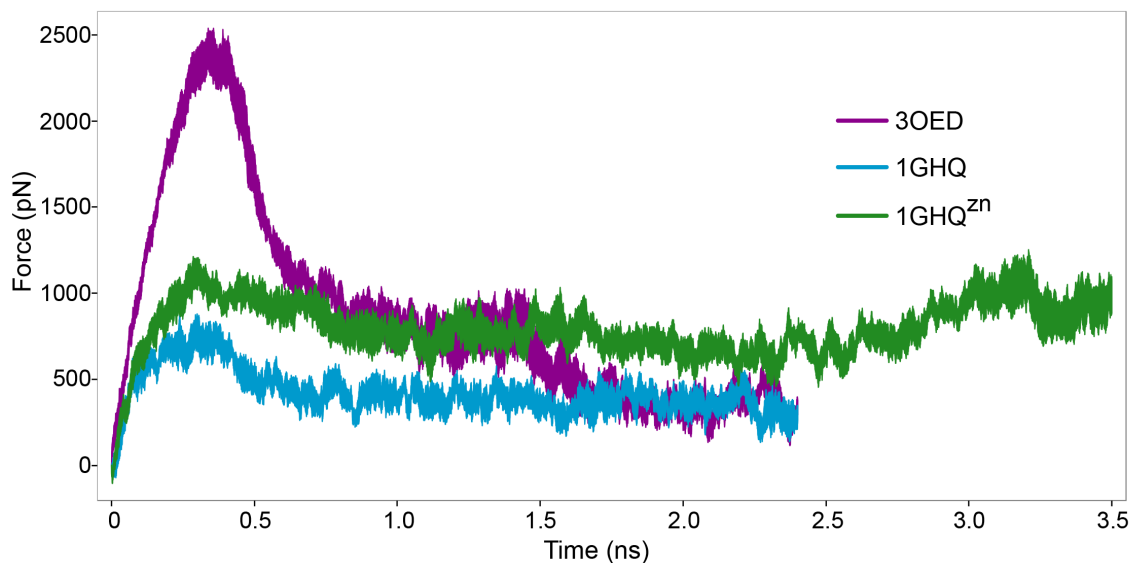
Figure 2.8: Force-time plots from the SMD simulations. Force-time curves for the nine trajectories from each system were averaged and plotted with the SEM shaded.

2.4   Discussion

We have thoroughly explored the two structures of 1GHQ and 3OED using several different computational tools to illuminate the nature of C3d:CR2(SCR1-SCR2) complex formation. A large amount of existing biochemical data and biophysical data, both experimental and theoretical, has been at odds with the 1GHQ structure and supported the possibility of a binding site at the acidic patch within the concave surface of C3d[6,11–13,15–23,30,50]. Past efforts to account for all sets of data by incorporating the concept of long range electrostatic interactions, which are indeed operative, were not sufficient in explaining all experimental data. Here, our results are in agreement with existing biochemical data and strongly suggests that 3OED is the energetically favorable binding mode of C3d:CR2(SCR1-SCR2),

and at the same time they provide insight into the influence of nonphysiological zinc ions on the formation of the 1GHQ structure.

2.4.1 Favorability of 3OED and importance of the SCR1 and SCR2 domains

Numerous C3d and CR2 residues, in the SCR1 domain in particular, are involved in the 3OED binding mode. The CR2 SCR1 domain plays a key role in electrostatic effects on binding (Fig. 2.2C), stabilizes the interface through highly conserved intermolecular interactions during MD simulations (Fig. 2.6C, F and I) and contributes to the dissociative stability during SMD simulations (Movie S1 of ref[49]). In our electrostatic analysis (Fig. 2.2C), SCR1 residues R13, R28, R36, K41, and K57 all exhibit significant contributions to binding (>5 kJ/mol increase in electrostatic free energy of binding). During the course of the MD trajectories, the majority of highly conserved nonpolar interactions (Fig. 2.6C), salt bridges (Fig. 2.6F) and hydrogen bonds (Fig. 2.6I) involve SCR1 residues R13, S15, Y16, Y17, G24, T25, V26, I27, R28, G33, R36, E40, S42, and L44. In the case of the SMD simulations, SCR1 continues to remain connected to C3d long after the detachment of SCR2 until eventually complete dissociation occurs, further reinforcing the importance of SCR1. The contributions of SCR2 to the 3OED binding mode are important as well, but as demonstrated by our results, SCR2 contributions are outmatched by those of SCR1. As for C3d, our electrostatic analysis (Fig. 2.2C) showed several C3d residues to have significant contributions to binding (>5 kJ/mol increase in electrostatic free energy of binding). We also observe that C3d residues D36, E37, Q50, L53, K57, V97, N98, L99, I100, A101, I102, D103, S104, K162, D163, I164, E166, P173, G174, M206, G207, K251, and K291 all help stabilize

the interface during MD simulations through highly conserved intermolecular interactions (Fig. 2.6C, F, and I). Overall, we observe that 3OED intermolecular interactions span a large sequence region. 3OED also demonstrates favorable electrostatic and nonpolar energetics (Fig. 2.7) as well as dissociative stability (Fig. 2.8 and Movie S1(of ref[49])).

2.4.2 Effects of nonphysiological zinc ions in 1GHQ

As expected from the structure of 1GHQ, the only CR2 residues involved in the binding mode are of the SCR2 domain. The electrostatic effects of CR2 residues in the absence of the nonphysiological zinc ions are mostly negligible but in the presence of the nonphysiological zinc ions, CR2 SCR2 domain residues E73 and D92 demonstrate major increases ($>5\,$kJ/mol) in electrostatic free energies of binding (Fig. S2). SCR2 domain residues Y80, K81, I82, R83, G84, and S85 stabilize the binding mode of 1GHQ through intermolecular interactions (Fig. 2.6A and G) but the presence of the nonphysiological zinc ions results in drastic increased conservation of interactions involving T86 and P87 (Fig. 2.5B and H). During the SMD simulations, CR2 dissociates from C3d fairly quickly in the absence of nonphysiological zinc ions but complete dissociation does not occur in the majority of the trajectories when the nonphysiological zinc ions are present. Specifically, as CR2 is pulled away from C3d, there is a single persistent interaction between E117 on C3d and D92 on CR2 (Figs. A.1 and A.2) mediated by a nonphysiological zinc ion as seen in Movie S3(of ref[49]). C3d E117 and CR2 D92 also show the highest increases in electrostatic free energy of binding in the presence of the nonphysiological zinc ions during our electrostatic analysis (Figs. A.1 and A.2).

2.4.3 Comparison of the two binding modes

When comparing the binding modes of 1GHQ and 3OED, it is evident that 3OED is more favorable. In 3OED, the large number of residues with increases in electrostatic free energy of binding, and the magnitude of the increases, overshadows those of 1GHQ (Fig. 2.2, Figs. A.1 and A.2). This is most clearly visualized in the open-book molecular graphics of the structures (Fig. 2.3) where it is clear that there are many more residues contributing to binding at the interface in 3OED. We also see that the intermolecular interactions contributing to the stability of 3OED span a much larger sequence region (Fig. 2.6). 3OED demonstrates much higher dynamic stability than 1GHQ through RMSD analysis (Fig. 2.4), interfacial SASA analysis (Fig. 2.5), intermolecular interaction occupancies (Fig. 2.6), and MM-GBSA calculations (Fig. 2.7). The force-time plots of SMD trajectories (Fig. 2.8) and the movies (Movies S1–S3 of ref[49]) of the trajectories themselves emphasize the dissociative stability of 3OED in comparison to 1GHQ. Although 1GHQ^{zn} doesn't exhibit complete dissociation in comparison to 3OED, it requires much less force to dissociate from the complex (Fig. 2.8), indicating the role of zinc in stabilizing the complex. The discrepancy is due to the unraveling of CR2 as it is pulled away and the interaction between C3d E117 and CR2 D92 mediated by a nonphysiological zinc ion persists (Movie S3 of ref[49]). The contributions of the SCR1 domain underlie the favorability of the 3OED binding mode as emphasized by major contributing residues in electrostatic effects (Fig. 2.2C, and Figs. A.1 and A.2), conserved intermolecular interactions (Fig. 2.6C, F, and I) and the delay in dissociation from C3d between the SCR2 and SCR1 domains (Movie S1 of ref[49]).

The potential role of the two nonphysiological zinc ions in crystal packing energetics in the binding mode of 1GHQ is evident through the observation of increases in electrostatic free energies of binding (Fig. 2.2B, Figs. A.1 and A.2), increases in the number and strength of intermolecular interactions (Fig. 2.6B, E, and H), increases in energetic favorability (Fig. 2.7), and increase in dissociative stability (Fig. 2.8 and Movie S3 of ref[49]) when the zinc ions are present in 1GHQ.

2.4.4   Results in light of previous experimental data

The authors of the 1GHQ structure published a mutagenesis study in 2005[11] in which they propose that the cluster of C3d residues E117, D122, D128 and D147 may be a good interface for SCR1 residues. In our electrostatic analysis (Fig. 2.2), we find that only E117 displays significant increases ($>5\,\text{kJ/mol}$) in electrostatic free energies of binding and that too only occurs in the 1GHQ$^{\text{zn}}$ structure. In the case of 3OED, we observe that this cluster of residues has almost no effect on binding, in agreement with data from the 2010 SPR study[13] by the same authors of the 3OED structure. The 2010 SPR data also suggested that C3d D36, E37, and E39 are important to binding and our 3OED electrostatic analysis (Fig. 2.2C) supports this as well. C3d K162 was proposed as a potential inhibitor for binding in the 2000 Rosette study[6] by the same authors of the 2010 study and our 3OED electrostatic analysis is in agreement again. In the case of CR2, there is a lot of overlap between the experimental data in the 2005 study[11] and the 2000 Rosette study[6]. Our data is generally in agreement with both studies with minor discrepancies in CR2 residues R36A, K50A, K50E, K57A, R83A and R83E (mutants from the 2005 mutagenesis study), which slightly deviate from the results of both the 2000 and 2005 study in the context of

3OED. Our electrostatic analysis regarding R83A also contradicts the experimental data published along with the 3OED structure[15]. Importantly, our 1GHQ CR2 data, both with and without the nonphysiological zinc ions, differ from both the 2005 and 2010 data with a few exceptions, indicating that the 1GHQ binding mode conflicts with the majority of biochemical and biophysical data.

Concerns regarding the 1GHQ structure included the lack of contact between SCR1 and C3d, absence of classical salt bridges, the sole side chain-side chain interaction observed was between C3d N170 and CR2 Y80, and the presence of nonphysiological zinc ions at the binding interface[13,15]. We have demonstrated the importance of SCR1 contacts to C3d in the context of 3OED through electrostatic analysis (Fig. 2.2), conserved intermolecular interactions (Fig. 2.6C, F, and I) and dissociative stability (Movie S1 of ref[49]). We find that the 3OED binding mode experiences several highly conserved salt bridges (Fig. 2.6F) during MD simulations while 1GHQ experiences minimal salt bridges with low percent occupancies (Fig. 2.6D and E). In the 2010 study[13], the authors found that mutating N170A did not affect the binding activity of C3d and our electrostatic data (Fig. 2.2C) supports this. Additionally, we found that during the MD trajectories, the only side chain-side chain interaction between N170 and Y80 observed is a nonpolar interaction in 1GHQ (Fig. 2.6A and B). The authors of the 2010 study also established that 0.2 M zinc acetate abrogates the interaction between C3d and CR2, thus questioning the crystallization conditions of 1GHQ and proposing that crystallization of 1GHQ occurred with the help of the nonphysiological zinc ions to optimize crystal packing energetics. Our results further support this hypothesis as we have shown the two nonphysiological zinc ions to have significant effects on the binding

interface of 1GHQ while at the same time establishing that 3OED is a much more favorable binding mode.

2.4.5   Electrostatic interactions in the complement system

A recent computational study has shown that the development of an evolutionary conserved acidic patch that is resistant to perturbations, is present in species that have both innate and adaptive immunity, compared to more primitive species that have innate immunity only[51]. The study demonstrated that the appearance of the functional acidic patch within the concave surface of C3d coincides with the onset of adaptive immunity in species of higher complexity, as the immune system evolved from noncellular defense to multicellular defense with memory. In this context, C3d evolved to become a link between innate and adaptive immunity, in addition to being an opsonin, therefore obtaining dual functionality. It is not accidental that CR2 interacts with C3d at the acidic patch, but also Factor H (SCR4 and SCR20) and S. aureus proteins Sbi, Efb, and Ecb, all interact with C3d at the same acidic patch[32,52,53]. All of these interactions are highly electrostatic in nature It appears that the C3d acidic patch not only is the site for interaction with CR2, and therefore contributing to linking innate and adaptive immunity (through augmentation of antibody production through the B cell receptor-coreceptor complex), but also a site of regulation by Factor H, and the site of immune evasion by bacterial proteins Sbi, Efb, and Ecb.

Regulators of complement activation that contain SCR modules are known to be highly charged, with net charges shown variability in sign and magnitude, which is believed to be functional. This is more evident in the functional diversity of the 20-SCR module Fac-

tor H[54], but also CR2, discussed here[16–19,55], and complement inhibiting proteins secreted from pox viruses (vaccinia and variola)[56–58] and Kaposi's sarcoma-associated herpesvirus[59]. A recent review discusses collectively the role of electrostatics in complement inhibition by SCR-containing viral proteins, targeting C3b and the C3d domain of C3b[60].

Based on the observed charge diversity, it appears that charge and electrostatics is a driving force for the regulation and inhibition of the complement system. The two-step model for protein-protein association consists of recognition and binding. Recognition refers to the electrostatic acceleration of the formation of a weak and non-specific complex (called the encounter complex), driven by long-range electrostatic interactions between protein macrodipoles; binding to structural and entropic re-arrangements to form the final and specific bound complex, which is stabilized by local pairwise energetic effects such as hydrophobic, van der Waals, and electrostatic (hydrogen bonds and salt bridges), as well as entropic effects such as loss of solvent molecules from the binding interface and gain by the bulk solvent, loss of overall translational and rotational degrees of freedom, and loss of local conformational freedom of side chains at the interface. Overall, there is plenty of theoretical and experimental evidence that our previously proposed two-step model for interactions of SCR-containing regulators of complement activation is operative in the regulation and inhibition of the complement system.

2.5   Conclusion

Through a combination of electrostatic analysis and analysis of MD and SMD simulations, we have established that the 3OED structure is energetically more favorable

than 1GHQ while also providing insight into the role of nonphysiological zinc ions at the 1GHQ interface. This observation is in favor of 3OED being the physiological binding state. We have addressed the majority of the discrepancies regarding the 1GHQ structure such as the lack of SCR1 contact with C3d, the smaller interfacial SASA, the absence of classical salt bridges, the sole side chain-side chain interaction between C3d N170 and CR2 Y80 and nonphysiological zinc ions at the binding interface. We have established 3OED to be more favorable and physiologically relevant in all these facets through an array of computational methods. We have also established the stabilizing effects of nonphysiological zinc ions on the 1GHQ structure, lending further credence to the potential role of crystal packing energetics in the crystallization of 1GHQ complex. With this study, we provide a theoretical basis for the closure to the binding controversy. Finally, we provided further evidence that electrostatics are a main driving force of C3d:CR2 association, in agreement with our previous studies.

2.6 References

[1] Roozendaal, R. and Carroll, M. C. Complement receptors CD21 and CD35 in humoral immunity. *Immunological reviews*, 219(1):157–166, 2007.

[2] Fearon, D. T. The complement system and adaptive immunity. *Seminars in Immunology*, 10(5):355–361, October 1998. ISSN 1044-5323. doi: 10.1006/smim.1998.0137.

[3] Carter, R. H. and Fearon, D. T. CD19: Lowering the Threshold for Antigen Receptor Stimulation of B Lymphocytes. *Science*, 256(5053):105–107, April 1992.

[4] Morikis, D. F1000Prime Recommendation of [van den Elsen JM and Isenman DE, Science 2011, 332(6029):608-11]. F1000Prime.com/10371956#eval14018054, September 2011.

[5] Nagar, B., Jones, R. G., Diefenbach, R. J., Isenman, D. E., and Rini, J. M. X-ray Crystal Structure of C3d: A C3 Fragment and Ligand for Complement Receptor 2. *Science*, 280(5367):1277–1281, May 1998. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.280.5367.1277.

[6] Clemenza, L. and Isenman, D. E. Structure-guided identification of C3d residues essential for its binding to complement receptor 2 (CD21). *The Journal of Immunology*, 165(7):3839–3848, 2000.

[7] Szakonyi, G., Guthridge, J. M., Li, D., Young, K., Holers, V. M., and Chen, X. S. Structure of Complement Receptor 2 in Complex with Its C3d Ligand. *Science*, 292 (5522):1725–1728, January 2001. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1059118.

[8] Carel, J. C., Myones, B. L., Frazier, B., and Holers, V. M. Structural requirements for C3d, g/Epstein-Barr virus receptor (CR2/CD21) ligand binding, internalization, and viral infection. *Journal of Biological Chemistry*, 265(21):12293–12299, 1990.

[9] Guthridge, J. M., Rakstang, J. K., Young, K. A., Hinshelwood, J., Aslam, M., Robertson, A., Gipson, M. G., Sarrias, M.-R., Moore, W. T., Meagher, M., Karp, D., Lambris, J. D., Perkins, S. J., and Holers, V. M. Structural Studies in Solution of the Recombinant N-Terminal Pair of Short Consensus/Complement Repeat Domains of Complement Receptor Type 2 (CR2/CD21) and Interactions with Its Ligand C3dg. *Biochemistry*, 40(20):5931–5941, May 2001. ISSN 0006-2960, 1520-4995. doi: 10.1021/bi0101749.

[10] Guthridge, J. M., Young, K., Gipson, M. G., Sarrias, M.-R., Szakonyi, G., Chen, X. S., Malaspina, A., Donoghue, E., James, J. A., Lambris, J. D., and others. Epitope mapping using the X-ray crystallographic structure of complement receptor type 2 (CR2)/CD21: Identification of a highly inhibitory monoclonal antibody that directly recognizes the CR2-C3d interface. *The Journal of Immunology*, 167(10):5758–5766, 2001.

[11] Hannan, J. P., Young, K. A., Guthridge, J. M., Asokan, R., Szakonyi, G., Chen, X. S., and Holers, V. M. Mutational Analysis of the Complement Receptor Type 2 (CR2/CD21)–C3d Interaction Reveals a Putative Charged SCR1 Binding Site for C3d. *Journal of Molecular Biology*, 346(3):845–858, February 2005. ISSN 0022-2836. doi: 10.1016/j.jmb.2004.12.007.

[12] Kovacs, J. M., Hannan, J. P., Eisenmesser, E. Z., and Holers, V. M. 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(14):9513–9520, March 2009. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.M808404200.

[13] Isenman, D. E., Leung, E., Mackay, J. D., Bagby, S., and van den Elsen, J. M. H. 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. *The Journal of Immunology*, 184(4):1946–1955, January 2010. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.0902919.

[14] Prota, A. E., Sage, D. R., Stehle, T., and Fingeroth, J. D. The crystal structure of human CD21: Implications for Epstein–Barr virus and C3d binding. *Proceedings of the National Academy of Sciences*, 99(16):10641–10646, 2002.

[15] van den Elsen, J. M. H. and Isenman, D. E. A Crystal Structure of the Complex Between Human Complement Receptor 2 and Its Ligand C3d. *Science*, 332(6029): 608–611, April 2011. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1201954.

[16] Cheung, A. S., Kieslich, C. A., Yang, J., and Morikis, D. Solvation effects in calculated electrostatic association free energies for the C3d-CR2 complex, and comparison to experimental data. *Biopolymers*, 2010. ISSN 00063525, 10970282. doi: 10.1002/bip. 21388.

[17] Kieslich, C. A., Morikis, D., Yang, J., and Gunopulos, D. Automated computational framework for the analysis of electrostatic similarities of proteins. *Biotechnology Progress*, 27(2):316–325, March 2011. ISSN 87567938. doi: 10.1002/btpr.541.

[18] Morikis, D. and Lambris, J. D. The Electrostatic Nature of C3d-Complement Receptor 2 Association. *J Immunol*, 172(12):7537–7547, June 2004. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.172.12.7537.

[19] Zhang, L., Mallik, B., and Morikis, D. Immunophysical Exploration of C3d–CR2(CCP1-2) Interaction Using Molecular Dynamics and Electrostatics. *Journal of Molecular Biology*, 369(2):567–583, June 2007. ISSN 0022-2836. doi: 10.1016/j.jmb. 2007.02.101.

[20] Gilbert, H. E., Asokan, R., Holers, V. M., and Perkins, S. J. The 15 SCR Flexible Extracellular Domains of Human Complement Receptor Type 2 can Mediate Multiple

Ligand and Antigen Interactions. *Journal of Molecular Biology*, 362(5):1132–1147, October 2006. ISSN 0022-2836. doi: 10.1016/j.jmb.2006.08.012.

[21] Gilbert, H. E., Eaton, J. T., Hannan, J. P., Holers, V. M., and Perkins, S. J. Solution Structure of the Complex between CR2 SCR 1-2 and C3d of Human Complement: An X-ray Scattering and Sedimentation Modelling Study. *Journal of Molecular Biology*, 346(3):859–873, February 2005. ISSN 0022-2836. doi: 10.1016/j.jmb.2004.12.006.

[22] Li, K., Okemefuna, A. I., Gor, J., Hannan, J. P., Asokan, R., Holers, V. M., and Perkins, S. J. Solution Structure of the Complex Formed between Human Complement C3d and Full-length Complement Receptor Type 2. *Journal of Molecular Biology*, 384 (1):137–150, December 2008. ISSN 0022-2836. doi: 10.1016/j.jmb.2008.08.084.

[23] Shaw, C. D., Storek, M. J., Young, K. A., Kovacs, J. M., Thurman, J. M., Holers, V. M., and Hannan, J. P. Delineation of the Complement Receptor Type 2–C3d Complex by Site-Directed Mutagenesis and Molecular Docking. *Journal of Molecular Biology*, 404 (4):697–710, December 2010. ISSN 0022-2836. doi: 10.1016/j.jmb.2010.10.005.

[24] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. The Protein Data Bank. *Nucl. Acids Res.*, 28(1): 235–242, January 2000. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/28.1.235.

[25] Gorham, R. D., Kieslich, C. A., and Morikis, D. Electrostatic Clustering and Free Energy Calculations Provide a Foundation for Protein Design and Optimization. *Ann Biomed Eng*, 39(4):1252–1263, April 2011. ISSN 0090-6964. doi: 10.1007/s10439-010-0226-9.

[26] Gorham, R. D., Kieslich, C. A., Nichols, A., Sausman, N. U., Foronda, M., and Morikis, D. An evaluation of poisson-boltzmann electrostatic free energy calculations through comparison with experimental mutagenesis data. *Biopolymers*, 95(11):746–754, 2011. ISSN 00063525. doi: 10.1002/bip.21644.

[27] Kieslich, C. A., Gorham Jr., R. D., and 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*, 357(2):707–716, January 2011. ISSN 0022-3093. doi: 10.1016/j.jnoncrysol.2010.05.087.

[28] Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res*, 32(Web Server issue):W665–W667, July 2004. ISSN 0305-1048. doi: 10.1093/nar/gkh381.

[29] Sitkoff, D., Sharp, K. A., and Honig, B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.*, 98(7):1978–1988, February 1994. ISSN 0022-3654. doi: 10.1021/j100058a043.

[30] Toapanta, F. R., DeAlmeida, D. R., Dunn, M. D., and Ross, T. M. C3d adjuvant activity is reduced by altering residues involved in the electronegative binding of C3d to CR2. *Immunology Letters*, 129(1):32–38, March 2010. ISSN 01652478. doi: 10.1016/j.imlet.2009.12.022.

[31] Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *PNAS*, 98(18):10037–10041, August 2001. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.181342398.

[32] Gorham, R. D., Rodriguez, W., and Morikis, D. Molecular Analysis of the Interaction between Staphylococcal Virulence Factor Sbi-IV and Complement C3d. *Biophysical Journal*, 106(5):1164–1173, March 2014. ISSN 00063495. doi: 10.1016/j.bpj.2014.01.033.

[33] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16):1781–1802, December 2005. ISSN 0192-8651, 1096-987X. doi: 10.1002/jcc.20289.

[34] MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B*, 102(18):3586–3616, April 1998. ISSN 1520-6106. doi: 10.1021/jp973084f.

[35] Yesselman, J. D., Price, D. J., Knight, J. L., and Brooks, C. L. MATCH: An atom-typing toolset for molecular mechanics force fields. *J Comput Chem*, 33(2):189–202, January 2012. ISSN 1096-987X. doi: 10.1002/jcc.21963.

[36] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. UCSF Chimera?A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, October 2004. ISSN 0192-8651, 1096-987X. doi: 10.1002/jcc.20084.

[37] Humphrey, W., Dalke, A., and Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, February 1996. ISSN 0263-7855. doi: 10.1016/0263-7855(96)00018-5.

[38] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.

[39] Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A., and Caves, L. S. D. Bio3d: An R package for the comparative analysis of protein structures. *Bioinformatics*, 22(21):2695–2696, January 2006. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btl461.

[40] Chen, J., Im, W., and Brooks, C. L. Balancing Solvation and Intramolecular Interactions: Toward a Consistent Generalized Born Force Field. *Journal of the American Chemical Society*, 128(11):3728–3736, March 2006. ISSN 0002-7863, 1520-5126. doi: 10.1021/ja057216r.

[41] Gohlke, H., Kiel, C., and Case, D. A. Insights into Protein–Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras–Raf and Ras–RalGDS Complexes. *Journal of Molecular Biology*, 330(4):891–913, July 2003. ISSN 0022-2836. doi: 10.1016/S0022-2836(03)00610-7.

[42] Im, W., Lee, M. S., and Brooks, C. L. Generalized born model with a simple smoothing function. *Journal of computational chemistry*, 24(14):1691–1702, 2003.

[43] Kieslich, C. A., Tamamis, P., D. Gorham Jr., R., Lopez de Victoria, A., U. Sausman, N., Archontis, G., and Morikis, D. Exploring Protein-Protein and Protein-Ligand Interactions in the Immune System using Molecular Dynamics and Continuum Electrostatics. *Current Physical Chemistry*, 2(4):324–343, 2012-12-01T00:00:00///.

[44] Brooks, B., Brooks, C., MacKerell, A., Nilsson, L., Petrella, R., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R., Post, C., Pu, J., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D., and Karplus, M. CHARMM: The Biomolecular Simulation Program. *J Comput Chem*, 30(10):1545–1614, July 2009. ISSN 0192-8651. doi: 10.1002/jcc.21287.

[45] Nina, M., Beglov, D., and Roux, B. Atomic Radii for Continuum Electrostatics Calculations Based on Molecular Dynamics Free Energy Simulations. *J. Phys. Chem. B*, 101(26):5239–5248, June 1997. ISSN 1520-6106. doi: 10.1021/jp970736r.

[46] Wu, J. C., Piquemal, J.-P., Chaudret, R., Reinhardt, P., and Ren, P. Polarizable molecular dynamics simulation of Zn(II) in water using the AMOEBA force field. *J Chem Theory Comput*, 6(7):2059–2070, July 2010. ISSN 1549-9618. doi: 10.1021/ct100091j.

[47] Cuendet, M. A. and Michielin, O. Protein-protein interaction investigated by steered molecular dynamics: The TCR-pMHC complex. *Biophys. J.*, 95(8):3575–3590, October 2008. ISSN 1542-0086. doi: 10.1529/biophysj.108.131383.

[48] Randjelović, J., Erić, S., and Savić, V. Computational study and peptide inhibitors design for the CDK9 – cyclin T1 complex. *Journal of Molecular Modeling*, 19(4): 1711–1725, April 2013. ISSN 1610-2940, 0948-5023. doi: 10.1007/s00894-012-1735-2.

[49] Mohan, R. R., Gorham Jr., R. D., and Morikis, D. A theoretical view of the C3d:CR2 binding controversy. *Molecular Immunology*, 64(1):112–122, March 2015. ISSN 0161-5890. doi: 10.1016/j.molimm.2014.11.006. 00002.

[50] Wan, H., Hu, J.-p., Tian, X.-h., and Chang, S. Molecular dynamics simulations of wild type and mutants of human complement receptor 2 complexed with C3d. *Physical Chemistry Chemical Physics*, 15(4):1241, 2013. ISSN 1463-9076, 1463-9084. doi: 10.1039/c2cp41388d.

[51] Kieslich, C. A. and Morikis, D. The Two Sides of Complement C3d: Evolution of Electrostatics in a Link between Innate and Adaptive Immunity. *PLoS Computational Biology*, 8(12):e1002840, December 2012. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002840.

[52] El-Assaad, A. M., Kieslich, C. A., Gorham Jr., R. D., and Morikis, D. Electrostatic exploration of the C3d–FH4 interaction using a computational alanine scan. *Molecular Immunology*, 48(15–16):1844–1850, September 2011. ISSN 0161-5890. doi: 10.1016/j.molimm.2011.05.007.

[53] Gorham, R. D., Kieslich, C. A., and Morikis, D. Complement Inhibition by Staphylococcus aureus: Electrostatics of C3d–EfbC and C3d–Ehp Association. *Cel. Mol. Bioeng.*, 5(1):32–43, March 2012. ISSN 1865-5025, 1865-5033. doi: 10.1007/s12195-011-0195-6.

[54] Kieslich, C. A., Vazquez, H., Goodman, G. N., de Victoria, A. L., and Morikis, D. The effect of electrostatics on factor H function and related pathologies. *Journal of Molecular Graphics and Modelling*, 29(8):1047–1055, August 2011. ISSN 1093-3263. doi: 10.1016/j.jmgm.2011.04.010.

[55] Morikis, D. and Zhang, L. An immunophysical study of the complement system: Examples for the pH dependence of protein binding and stability. *Journal of Non-Crystalline Solids*, 352(42–49):4445–4450, November 2006. ISSN 0022-3093. doi: 10.1016/j.jnoncrysol.2006.01.104.

[56] Sfyroera, G., Katragadda, M., Morikis, D., Isaacs, S. N., and Lambris, J. D. Electrostatic Modeling Predicts the Activities of Orthopoxvirus Complement Control Proteins. *J Immunol*, 174(4):2143–2151, February 2005. ISSN 0022-1767.

[57] Yadav, V. N., Pyaram, K., Ahmad, M., and Sahu, A. Species selectivity in poxviral complement regulators is dictated by the charge reversal in the central complement control protein modules. *J. Immunol.*, 189(3):1431–1439, August 2012. ISSN 1550-6606. doi: 10.4049/jimmunol.1200946.

[58] Zhang, L. and Morikis, D. 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(9):3106–3119, May 2006. ISSN 0006-3495. doi: 10.1529/biophysj.105.068130.

[59] Pyaram, K., Kieslich, C. A., Yadav, V. N., Morikis, D., and Sahu, A. Influence of electrostatics on the complement regulatory functions of Kaposica, the complement

inhibitor of Kaposi's sarcoma-associated herpesvirus. *J. Immunol.*, 184(4):1956–1967, February 2010. ISSN 1550-6606. doi: 10.4049/jimmunol.0903261.

[60] Ojha, H., Panwar, H. S., Gorham, R. D., Morikis, D., and Sahu, A. Viral regulators of complement activation: Structure, function and evolution. *Molecular Immunology*, 61(2):89–99, October 2014. ISSN 01615890. doi: 10.1016/j.molimm.2014.06.004.

Chapter 3

The role of electrostatic steering in the acceleration of C3d:CR2 association

3.1 Introduction

Electrostatics plays an important role in accelerating biomolecular reactions, such as diffusional encounters and catalytic processes [1–5]. Both long-range and short-range electrostatic interactions have been shown to affect the protein–protein association rate [6–8] and have demonstrated important contributions to improving the efficiency of enzymatic reactions, often by several orders of magnitude [9,10]. An area of research that has received significant attention in recent computational studies is the role of electrostatics in the function and regulation of the complement system, as well as the electrostatic mechanisms

that bacterial and viral proteins have evolved to infiltrate host cells or evade the immune system[11–20].

The complement system is a vital component of innate immunity, acting as a rapid-response surveillance system that identifies and eliminates, or contributes to the elimination of foreign pathogens through the processes of inflammation, opsonization, phagocytosis, and direct cell lysis[21]. In addition, the complement system contributes to clearance of apoptotic cells, damaged cells and cellular debris, and immune complexes[22–26]. The complement system is tightly regulated to discriminate self-from nonself[27], and when such regulation fails, the complement system contributes to autoimmune and inflammatory diseases[28–30]. Overall, the complement system senses and responds to danger signals, contributing to host homeostasis[23,25].

A direct result of complement activation is the role it plays as a link between innate and adaptive immunity, through the interaction between complement fragment C3d and complement receptor 2 (CR2). This is a property of mammals and higher species, because invertebrates have complement immune response but lack adaptive immunity. The formation of the C3d:CR2 complex contributes to the formation of the B cell receptor–coreceptor complex and to the enhancement of B cell-mediated antibody production by up to 3–4 orders of magnitude[31–33]. Due to the importance of the C3d:CR2 complex to the development of autoantibodies, the C3d:CR2 interaction is also implicated in the pathology of autoimmune and inflammatory diseases. Thus, a comprehensive understanding of the nature of the C3d:CR2 interaction not only will contribute to mechanistic knowledge of a fundamental

immune response process, but also can serve as the basis for improvements in therapeutic development.

Complement activation occurs through three different pathways: the classical, alternative, and lectin pathways. All three pathways converge at complement component C3[34]. Complement C3 undergoes a series of cleavage steps that activate, inactivate, and redirect its activation. The first cleavage step produces the opsonin fragment C3b that covalently (through a thioester bond) attaches to pathogens and other surfaces, tagging them for recognition and elimination by phagocytic cells, and the fragment C3a that contributes to inflammatory response and phagocytosis. The second cleavage step produces the so-called inactivated C3b, iC3b, which also contributes to phagocytosis. The final cleavage steps produce C3c and C3dg, which is immediately transformed to C3d. Although C3d is the final cleavage product that remains on cell surfaces for the life of the cell, it is not just a degradation product. Nature has evolved mechanisms that utilize antigen-bound C3d and B cell expressed CR2 as a site of interaction between innate and adaptive immunity. B cell expressed antibodies also opsonize pathogens by binding to antigens on pathogen surfaces. Thus, the combined function of B cell bound antibodies and the CR2-C3d complex cross-link B cells to pathogens, forming the so-called B cell receptor (antibody)–coreceptor (CR2) complex. This cross-linking initiates a cascade of intracellular signaling reactions, involving protein kinases.

Several structural and computational studies have proposed that the interaction between C3d and CR2 is predominantly electrostatic in nature[35–39], occurring through a negatively charged patch on a concave surface of C3d and the first two modules of CR2

that are positively charged (Figure 3.1). Experimental studies involving mutagenesis, pH, and ionic strength effects are in agreement with the dominant role of electrostatics in the association between C3d and CR2[35,36,40–42]. Although there was ambiguity and controversy for several years because of an older nonphysiological crystallographic structure of the C3d:CR2 complex, this controversy is now resolved with new crystallographic, mutagenesis and binding, and computational data[19,36,41,43]. A recent study has evaluated the physicochemical origins and strength of the C3d:CR2 interaction, using the physiological and the controversial crystallographic structures, and has demonstrated the electrostatic mechanism of binding[19]. This and earlier studies[37–39] have proposed a two-step model for C3d:CR2 association, consisting of recognition and binding for highly and oppositely charged proteins. This model was based on earlier work on electrostatic steering in enzymatic reactions and protein interactions by McCammon and co-workers[1–10,44–46]. During the recognition step, long-range electrostatic interactions between protein macrodipoles accelerate the formation of a transient encounter complex, followed by the binding step, which is marked by the stabilization of the bound complex through short-range, pairwise polar and nonpolar interactions and entropic effects.

Another recent study used the concept of "electrostatic hotspots" to evaluate the origin of the C3d–CR2 interaction throughout evolution[47]. An "electrostatic hotspot" was defined as a surface patch of like-charged residues that is resistant to perturbation. Such hotspots contribute to the formation of the encounter complex and rapid association. Because of the high concentration of like charges, a hotspot forms an unfavorable electrostatic environment, which is amenable to the formation of favorable interactions with proteins
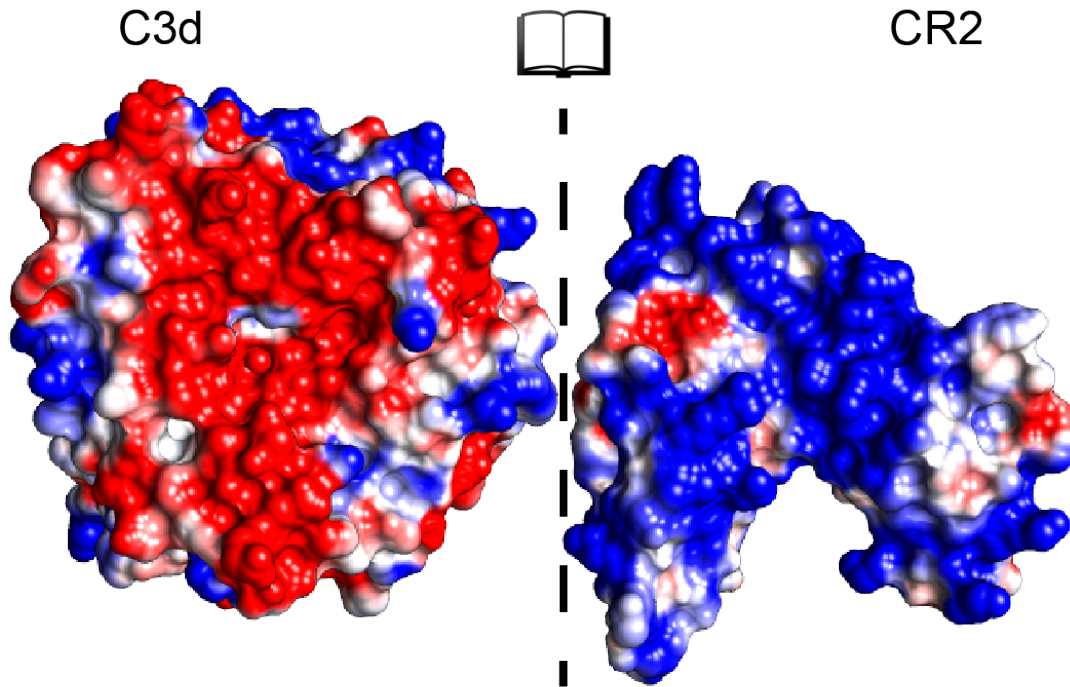
Figure 3.1: Electrostatic potentials mapped onto the protein surfaces of C3d and CR2 in open book representation. Electrostatic potentials were calculated at ionic strength corresponding to 150 mM monovalent counterion concentration. The color transitions from red to white to blue represent electrostatic potential values of $-5kT/e$ to $0kT/e$ to $5kT/e$.

that have areas with complementary charges when such an encounter occurs. The study was performed using C3d sequences from 24 species, homology modeling based on the most recent (physiological) crystallographic structure, perturbations based on alanine scan of ionizable residues and molecular dynamics simulations, Poisson–Boltzmann electrostatic calculations, and electrostatic potential similarity clustering. The study proposed that C3d has two "electrostatic hotspots" located at opposite faces; one hotspot is predominantly positively charged and contains the thioester bond that opsonizes pathogen surfaces, and the other hotspot is predominantly negatively charged and forms the concave surface that is the site of interaction with CR2. The study concluded that the appearance of the negatively

charged hotspot coincides with the onset of adaptive immunity at the level of jawless fish and beyond, and it is stronger in mammals, but it is not present in invertebrates. Therefore, C3d evolved its electrostatic properties to acquire the negatively charged hotspot to interact with the positively charged CR2, resulting to enhancement of adaptive immunity.

Given the large number of background studies on the electrostatic character of the C3d–CR2 interaction, and that the formation of the encounter complex is a diffusion-limited process[48], we initiated a Brownian dynamics (BD) simulation study to evaluate k_{on} reaction rate constants for C3d:CR2 complexes. We perform our study for native C3d and CR2, as well as for a number of mutants with available experimental binding data. We evaluate the ionic strength dependence of the C3d:CR2 interaction to investigate the impact of salt concentration on electrostatic screening. We utilize computational mutagenesis to elucidate the contributions of specific mutants, with known experimental binding data, to the association of the C3d:CR2 complex. We demonstrate the inverse relationship between the association rate constant and ionic strength.We also find that mutations of residues shown to enhance or hinder binding in experimental binding data[35,40,41] result in slower or higher association rate constants, respectively. The examined mutations involved ionizable residues at the binding interface and were introduced to disrupt association. Our results are in agreement with the experimental data, as well as with a previous computational study[19], and they indicate that the electrostatic steering accelerates the interaction between C3d and CR2.

3.2 Methods

3.2.1 Protein Structure Preparation

We utilized the more recent crystallographic structure of the C3d:CR2 complex (Protein Data Bank, PDB, code: 3OED)[36] in our study. From the three-dimensional coordinates of this structure, we used chains A and C, corresponding to C3d and CR2, respectively, as they had better electron density and lower B-factor compared to chains B and D of another complex present in the structure. It should be noted that the structure of CR2 contains only the two modules that contact C3d, SCR1, and SCR2, out of a total of 15 or 16 SCR modules. C3d consists of 292 amino acids with a net charge of $-1e$ while CR2 consists of 130 amino acids with a net charge of $+8e$. To alleviate crystal packing effects in the crystallographic structure, 25,000 steps of conjugate-gradient energy minimization were performed using NAMD[49].

Subsequent to energy minimization, missing hydrogens, atomic radii, and partial charges were added to the coordinates of the structure, using PDB2PQR version 2.0[50] and the PARSE force field[51], thus converting the PDB file to a PQR file. No atypical protonation states were observed using PROPKA[52,53]. Histidine residues were neutral with a hydrogen attached to $N^{\delta 1}$ atom. Computational mutagenesis was performed on the protein complex using the analysis of electrostatic similarities of proteins (AESOP) computational framework[15,39,54,55]. Mutants were chosen from prior literature that had reported experimental binding data[35,40,41]. Electrostatic potentials were calculated for the parent (wild-type) and mutated protein complexes using the Adaptive Poisson–Boltzmann Solver (APBS) version 1.4[56]. The number of grid points was set to $129 \times 161 \times 161$.

Coarse and fine mesh dimensions were set to $1000\,\text{Å} \times 1000\,\text{Å} \times 1000\,\text{Å}$ and $150\,\text{Å} \times 150\,\text{Å} \times 150\,\text{Å}$, respectively, as discussed previously[54]. Protein and solvent dielectric values were set to 20 and 78.54, respectively[54]. Ionic strengths corresponding to monovalent counterion concentrations of 50, 75, 100, 125, 150, 200, and 300 mM were used for the evaluation of ionic strength dependence of parent C3d:CR2 and the alanine scan mutants. An ionic concentration of 150 mM was used in the electrostatic analysis of the alanine scan mutants.

3.2.2 Brownian Dynamics Simulations

Brownian dynamics simulations and the corresponding rate calculations were performed using the BrownDye package[57], according to the Northrup–Allison–McCammon algorithm[44], which is based on the original Brownian dynamics algorithm by Ermak and McCammon[58]. The two molecules start separated at a center-to-center radius (represented by the inner circle in Figure 3.2A), and the simulation progresses until termination due to formation of the encounter complex or due to the molecule reaching an escape radius (represented by the outer circle in Figure 3.2A). The center-to-center radius is calculated by BrownDye for each system, and through an improvement to the Northrup–Allison–McCammon algorithm, the escape radius is no longer a necessary input[59]. Pairwise residue interactions with cutoff distances of $5.0\,\text{Å}$ were calculated from the PQR files generated as described above. Criteria for a successful reaction required that at least two of the atom pairs from the calculated list of pairwise interactions approach within $3.495\,\text{Å}$ of each other (Figure 3.2B).
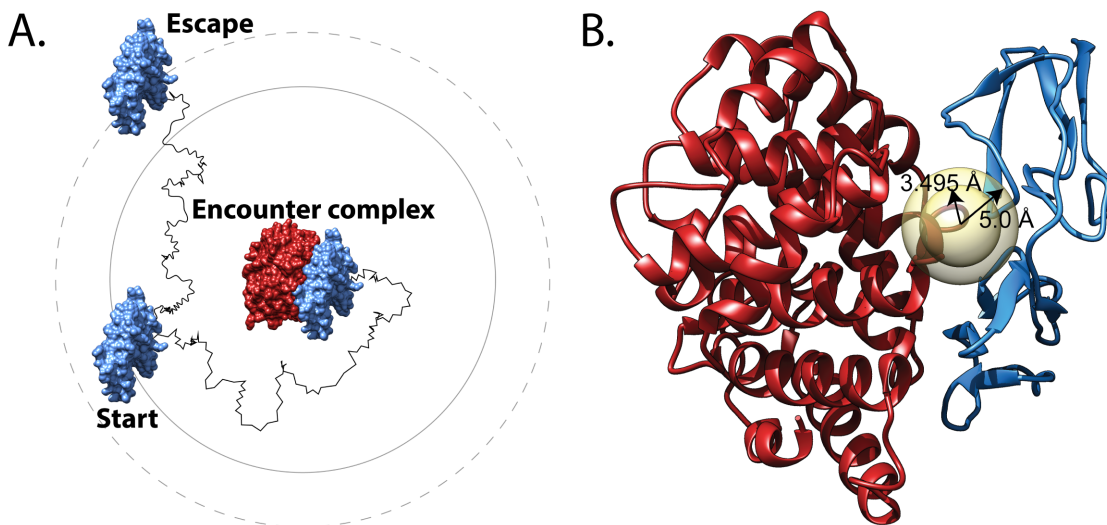
Figure 3.2: Schematic and molecular graphics illustrating the Brownian dynamics (BD) simulation of C3d:CR2 and the corresponding reaction criteria with C3d and CR2 in blue and red, respectively. (A) At the beginning of the BD simulation, CR2 starts at a center-to-center radius away from C3d, as represented by the inner circle. The simulation terminates when either the formation of the encounter complex occurs or if CR2 reaches an escape radius as represented by the outer circle. (B) The concentric circles represent the reaction criteria for the C3d:CR2 BD simulations where $5.0\,\text{Å}$ is the cutoff for determining potential pairwise residue interactions between C3d and CR2. The circle with radius $3.495\,\text{Å}$ represents the distance within which at least two atom pairs of the previously determined pairwise residue interactions must occur for a successful reaction.

Reaction criteria were selected to match known rate constants of C3d:CR2 association. Initially, the BrownDye program rates_of_distances was utilized to generate a list of reaction constants corresponding to minimum reaction distances. Then, the criteria were fine-tuned to the known association rate constant at 125 mM NaCl from experimental data[40]. There are two experimental association rate constants, at 50 and 125 mM NaCl ionic strength, and we chose for calibration the value at 125 mM because of its proximity to the physiological ionic strength of 150 mM. Additional input files were generated using the program bd_top. BD simulations were carried out using the weighted-ensemble method to account for low probabilities of reactions[60]. The association rate constant, k_{on}, and

corresponding reaction probabilities were calculated using weighted-ensemble simulations carried out for 2,000,000 steps with 200 copies of each system to guarantee convergence of the results. The acceleration of protein–protein association, as well as enzymatic reactions, by electrostatic steering has been explored previously, which suggested the feasibility of this study[61,62].

3.3 Results and Discussion

3.3.1 Ionic Strength Dependence of Association

The weighted-ensemble BD simulations of C3d:CR2 demonstrate that under constraints of reaction criteria of 3.495 Å and at least two successful pairwise interactions, the association rate constant decreases with increasing ionic strength (Figure 3.3), which is expected when ionic screening of Coulombic interactions is present. The experimental data utilized for calibration of the reaction criteria is plotted in Figure 3.3 as well. Additionally, Figure 3.3 shows the ionic strength dependence of the Debye length, suggesting the importance of ionic screening for association. Both the ionic strength dependence and Debye length curves follow similar trends.

Because C3d and CR2 have surfaces that are both highly and oppositely charged (Figure 3.1), we expect that association follows the two-step model of recognition (formation of the intermediate encounter complex) and binding (formation of the final bound complex). This is demonstrated by the ionic strength dependence of the association, which is possible if electrostatics drives association. From our results, the acceleration of the interaction
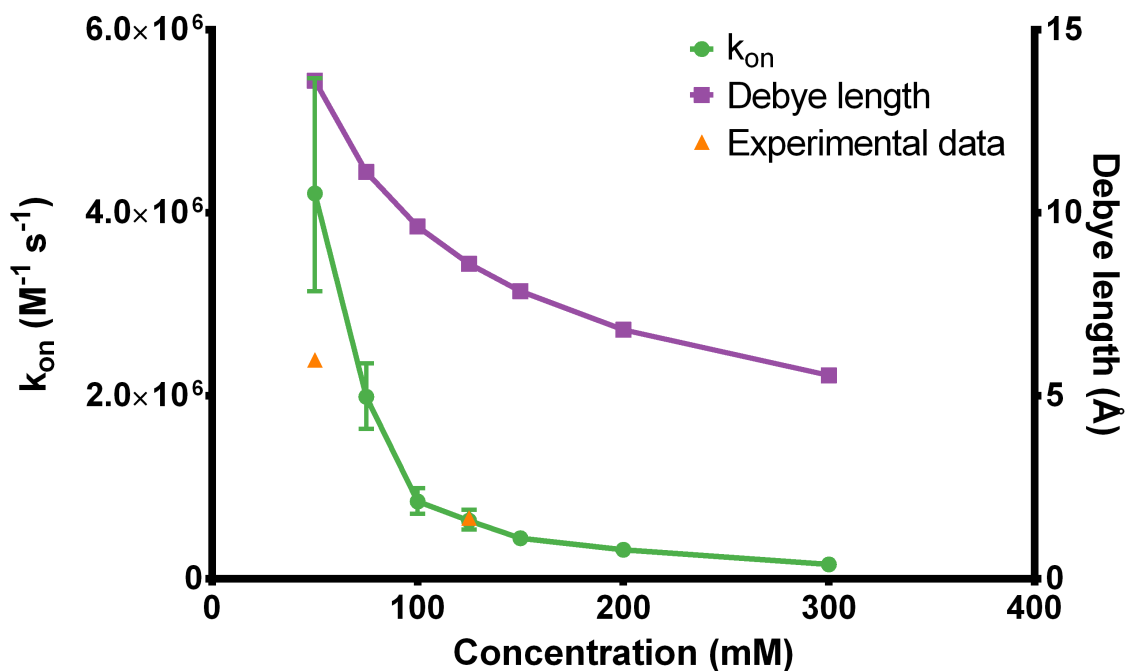
Figure 3.3: Association rate constant of CR2 binding to C3d and the calculated Debye length at varying ionic strengths. The mean association rate constant is represented by a green circle with 95% confidence intervals represented as error bars. The Debye length at each ionic strength is represented as a purple square. Experimentally known association rate constants at two ionic strengths, 50 mM and 125 mM NaCl, are plotted and represented as orange triangles[40]. Note: the experimental data point at 125 mM was utilized for calibration of the BD simulation reaction criteria.

through electrostatic steering underscores the plausibility of the model. The kinetic rate constants acquired through these BD simulations reflect the recognition step, given the calculation limitation, such as the rigid body assumption of the protein structures and lack of modeling of short-range interactions such as salt bridges, van der Waals forces, and hydrogen bonding[63]. These limitations may affect only the binding step of the aforementioned two-step model of protein–protein association and may not be necessary for the evaluation of electrostatic steering during the diffusion-limited recognition step[64].

3.3.2   Impact of Specific Residues on Electrostatic Steering

We investigated the effects of electrostatic steering when various C3d:CR2 residues are mutated. Computational mutations were chosen from available experimental data and BD simulations were performed for each mutant system. We observe that mutations of certain residues result in a decrease in the association rate constant, or in other words, the residue is significant to the formation of the encounter complex of C3d:CR2 (Figure 3.4A,B). The reverse holds true as well where mutations of residues resulting in an increase in the association rate constant signify that the residue hinders the formation of the encounter complex. These results are in line with a previous computational alanine scan and electrostatic analysis study using AESOP[19] with a few small discrepancies such as K251A on C3d and R83A on CR2, which are close to the threshold of the error. Thus, we establish that the acceleration of the formation of the encounter complex due to electrostatic steering is affected by individual charged residues contributions as well.

Mutants displaying significant variance from the wild-type (outside the shaded area of Figure 3.4A,B) were selected for ionic strength dependence analysis (Figure 3.4C,D). We find that the mutants exhibit ionic strength dependence as well, and the trends are overall in line with what we would expect from the mutagenesis analysis of Figure 3.4A,B. As expected from the mutagenesis analysis, certain mutants demonstrate more drastic electrostatic steering, suggesting that they play a more significant role in the formation of the encounter complex.
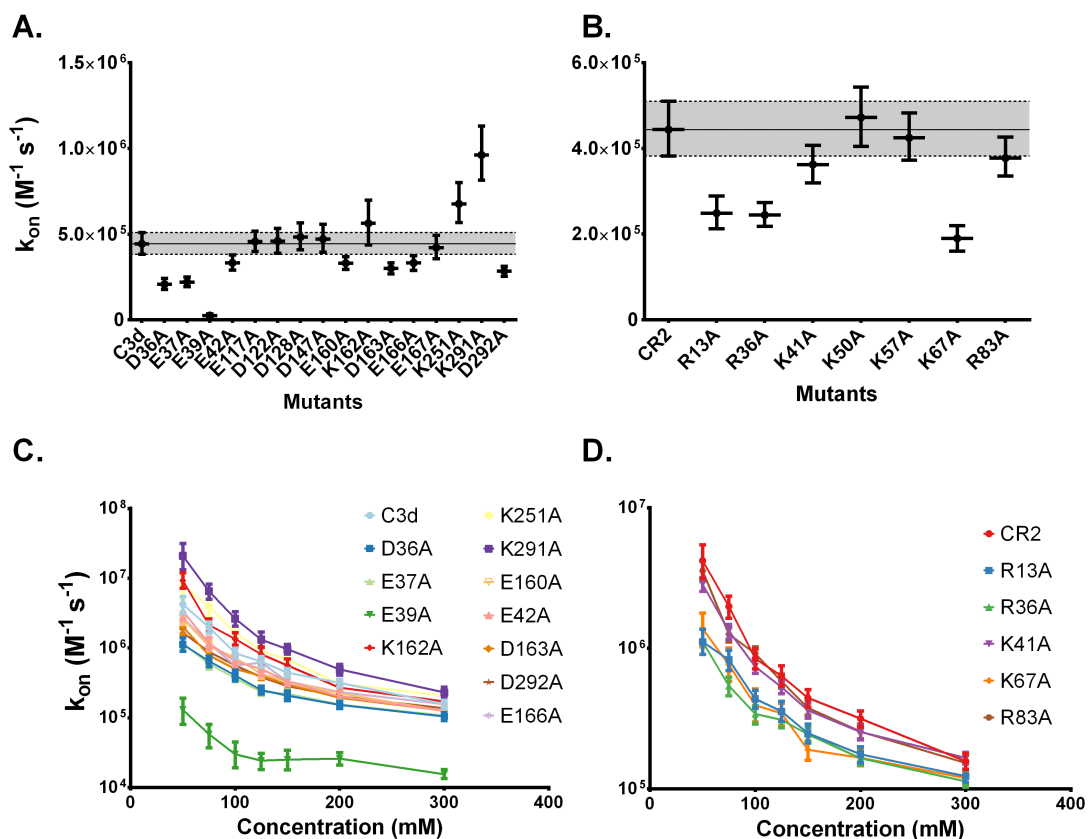
Figure 3.4: Effects of mutagenesis on association rate constant and ionic strength dependence. (A, B) The mean association rate at 150 mM ionic strength is plotted for each mutant for C3d and CR2 with 95% confidence intervals represented as error bars. The shaded region represents the upper and lower bound threshold of the association rate constant values of the parent (C3d or CR2). Computational mutations were performed based on previous experimentally generated mutants[35,40,41]. (C, D) Ionic strength dependence of selected mutants demonstrating significant deviations from the wild type association rate constant. The mutant notation denotes the residue number surrounded by the replaced residue on the left and the replacing residue on the right.

3.3.3 Comparison with Prior Experimental and Computational Studies

A recent computational study investigating the binding mode of C3d:CR2 using molecular dynamics (MD) simulations (both explicit-solvent and steered), including MM-GBSA analysis, and electrostatic calculations, including AESOP alanine scan analysis, quantifies why the acidic patch on C3d plays a key role in driving the C3d:CR2 interaction[19]. In particular, two clusters of C3d residues (D36, E37, and E39; E160, K162, D163, E166, and E167) demonstrate significant contributions to electrostatic interactions, intermolecular interaction occupancies (hydrogen bonds, salt bridges, and nonpolar interactions) and steered MD (SMD) simulations. The computational study also emphasizes the importance of both the SCR1 and SCR2 domain to the stability and energetics of the complex. This is supported by the electrostatic and MD simulation analysis and also was strongly suggested by the slow unbinding of the SCR1 domain in SMD simulations, in contrast to the SCR2 domain.

Our results are generally in agreement with prior experimental and computational data. C3d residues at the acidic patch such as D36, E37, and E39 (Figure 3.5) were suggested to be important to binding by a 2000 rosette immunological assay study[35] and a 2010 surface plasmon resonance study[41], in addition to the AESOP analysis[19]. Our BD results support the importance of D36, E37, and E39 in binding as well. C3d residues E117, D122, D128, and D147, a cluster of residues not located at the acidic patch, demonstrated minimal contributions to electrostatic steering, in contrast to a mutagenesis study[40] performed by authors of the older crystallographic C3d:CR2 structure[65]. This is in line with the previous computational study[19] and the 2000 and 2010 mutagenesis studies[35,41].
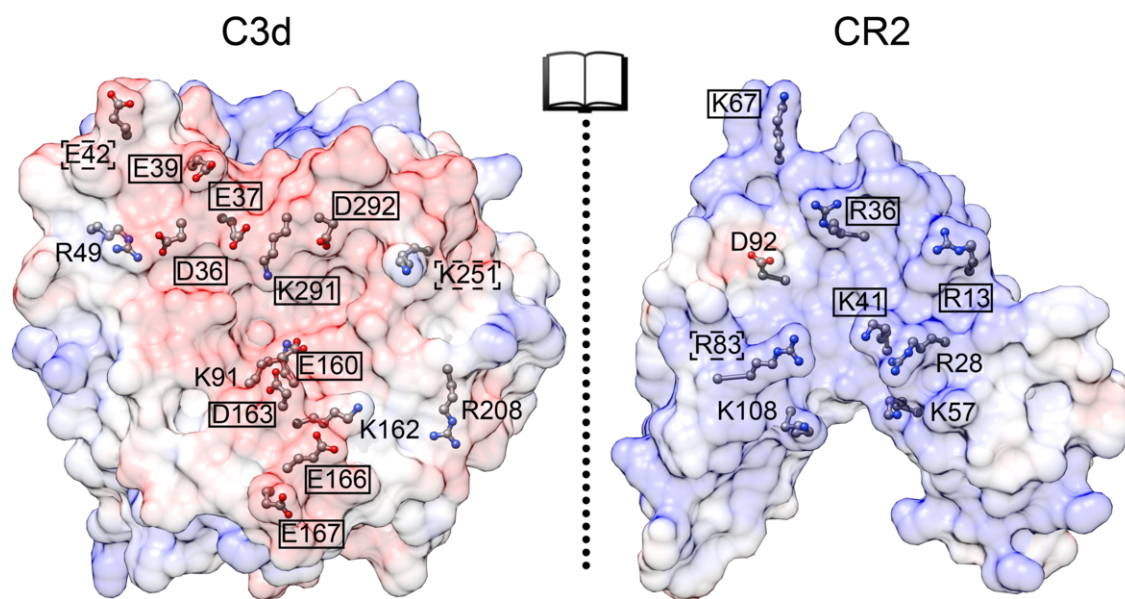
Figure 3.5: Molecular graphic of C3d:CR2 in open-book form and locations of residues with significant contributions to electrostatic steering. The surfaces of C3d and CR2 are rendered translucent and colored from red to white to blue according to electrostatic potential values (calculated at 150 mM ionic strength) from $-5kT/e$ to $0kT/e$ to $5kT/e$. Residues are displayed in ball-and-stick form with atoms colored according to atom type (carbon in gray, oxygen in red, nitrogen in blue). Residues found to demonstrate significant contributions to electrostatic steering (falling outside the shaded area in Figure 4A,B) in this ionic strength dependence analysis and significant contributions to electrostatic interactions (greater or less than ± 2.5 kJ/mol) in a previous AESOP computational alanine scan study [19](19) are labeled with a solid black box. Residues found to demonstrate significant contributions to electrostatic steering in this ionic strength dependence analysis but less significant interactions (within ± 2.5 kJ/mol) in the AESOP computational alanine scan study are labeled with a dashed black box. Additional residues found to have significant contributions to electrostatic interactions in the AESOP computational alanine scan study are labeled without a box. This figure is an updated version of the one presented in ref[66].

C3d residues at the other acidic patch cluster (E160, D163, and E166) demonstrated significant contributions to electrostatic steering, in line with the previous computational study[19], but are at odds with the 2000 mutagenesis study. However, the influence of electrostatic steering of C3d residues K162 and E167 are in contrast to the previous computational study[19] and 2000 mutagenesis study[35]. Here, we found that removal of K162, which is located within the acidic patch of C3d, did not affect the electrostatic steering of the formation of the encounter complex as dramatically as it was suggested by the results of the 2000 rosette study[35] and the previous AESOP study[19]. It is likely that the role of K162 is to stabilize the negative surface patch, and its removal produces local structural rearrangements to optimize the remaining electrostatic interactions. Such structural effects are not taken into account by the rigid model BD and AESOP studies. We should keep in mind that the BD and AESOP analyses introduce theoretical perturbation to assess the importance of each mutated residue in binding of the parent proteins, and they do not aim to model the structures of the mutated proteins.

The importance of the SCR1 domain of CR2, which was overlooked by the original C3d:CR2 crystallographic structure, is further emphasized in our BD results, as evidenced by the effects of mutagenesis and electrostatic steering by CR2 residues R13, R36, and K41. This is in line with the previous computational study where it was established that the contributions of the SCR1 domain to the C3d:CR2 interface are significant[19].

3.3.4 Pharmacological Significance

Understanding the physicochemical origins of the C3d:CR2 interaction is important for the design of biomarkers and therapeutic interventions. Recent studies have utilized

information from the C3d–CR2 interaction to design C3d-binding biomarkers for imaging of complement activation, such as fluorescently labeled antibodies[67],(66) small fluorescent molecules[68],(67) and CR2-bound iron nanoparticles[69].(68)

The complement system has been a target of inhibition in several studies, but only two anticomplement drugs are currently in the clinic[70]. Although, a lack of proper regulation of complement response has been implicated in several autoimmune and inflammatory diseases, inhibiting complement activation may reduce the efficacy of response to infection and injury. Recent studies have demonstrated in animal models effective targeted delivery of CR2-attached protein complement inhibitors, through the CR2:C3d interaction, to sites of local inflammation, which are abundant of C3d-opsonized tissues[71–76]. Such inhibitors included fragments of the alternative pathway regulator Factor H, or the complement inhibitory protein Crry, but can also be low-affinity C3d-bound or CR2-bound peptidic or nonpeptidic molecules. Therefore, the C3d:CR2 interaction is an ideal target for therapeutic development because it allows for inhibition of complement response through the alternative pathway while retaining the ability to fight infection through the lectin and classical pathways. The C3d:CR2 interaction can also be a target for therapeutic intervention in cases of autoimmunity because it allows for inhibition of complement-enhanced adaptive immune response, through the inhibition of the formation of the B cell receptor–coreceptor complex.

3.4 Conclusion

We have investigated the role of electrostatic steering on the C3d:CR2 interaction, using BD simulations. We demonstrate that the predicted k_{on} reaction rate constant depends on ionic strength, which is possible only if electrostatics contributes significantly to the C3d:CR2 interaction. We have also evaluated the contributions of specific ionizable residues to the electrostatic acceleration of the interaction through computational mutagenesis and ionic strength dependence analysis. We demonstrate that computational mutations of ionizable residues previously known from experimental studies to be significant to the C3d:CR2 interaction result in a reduced k_{on} rate constant. Therefore, the replaced (original) residues contribute to the acceleration of the interaction in the native complex. These results are in agreement with a previous computational analysis, based on the calculation of electrostatic free energies of association for a family of experimentally known mutants and an alanine scan family of C3d:CR2 complexes[19]. Interestingly, acidic residues within an evolutionarily significant "electrostatic hotspot" in C3d are the major contributors to the complex formation. As was suggested in a previous work, this acidic patch in the complement degradation product C3d may have evolved to establish a link between innate immunity (complement system) and adaptive immunity (B cell bound antibodies)[47]. Although this C3d "electrostatic hotspot" may be destabilizing local structure, it accelerates interaction with CR2 and function. As was suggested by Professor J. Andrew McCammon in his 2009 "Darwinian Biophysics" article, evolution favors speed and acceleration of function[77]; and evolution favors function over stability[78].

3.5 References

[1] Sines, J. J., Allison, S. A., and McCammon, J. A. Point charge distributions and electrostatic steering in enzyme/substrate encounter: Brownian dynamics of modified copper/zinc superoxide dismutases. *Biochemistry*, 29(40):9403–9412, October 1990. ISSN 0006-2960. doi: 10.1021/bi00492a014. 00131.

[2] Tan, R. C., Truong, T. N., McCammon, J. A., and Sussman, J. L. Acetylcholinesterase: Electrostatic steering increases the rate of ligand binding. *Biochemistry*, 32(2):401–403, January 1993. ISSN 0006-2960. doi: 10.1021/bi00053a003. 00138.

[3] Gabdoulline, R. R. and Wade, R. C. Simulation of the diffusional association of barnase and barstar. *Biophysical Journal*, 72(5):1917–1929, May 1997. ISSN 0006-3495. doi: 10.1016/S0006-3495(97)78838-6. 00245.

[4] Elcock, A. H. and McCammon, J. A. Evidence for Electrostatic Channeling in a Fusion Protein of Malate Dehydrogenase and Citrate Synthase. *Biochemistry*, 35(39):12652–12658, January 1996. ISSN 0006-2960. doi: 10.1021/bi9614747. 00052.

[5] Gabdoulline, R. R. and Wade, R. C. Biomolecular diffusional association. *Current Opinion in Structural Biology*, 12(2):204–213, April 2002. ISSN 0959-440X. doi: 10.1016/S0959-440X(02)00311-1. 00171.

[6] Elcock, A. H., Sept, D., and McCammon, J. A. Computer Simulation of Protein-Protein Interactions. *The Journal of Physical Chemistry B*, 105(8):1504–1518, March 2001. ISSN 1520-6106, 1520-5207. doi: 10.1021/jp003602d.

[7] Chang, C.-E., Shen, T., Trylska, J., Tozzini, V., and McCammon, J. A. Gated Binding of Ligands to HIV-1 Protease: Brownian Dynamics Simulations in a Coarse-Grained Model. *Biophysical Journal*, 90(11):3880–3885, June 2006. ISSN 0006-3495. doi: 10.1529/biophysj.105.074575. 00066.

[8] Spaar, A., Dammer, C., Gabdoulline, R. R., Wade, R. C., and Helms, V. Diffusional Encounter of Barnase and Barstar. *Biophysical Journal*, 90(6):1913–1924, March 2006. ISSN 00063495. doi: 10.1529/biophysj.105.075507. 00100.

[9] Wade, R. C., Luty, B. A., Demchuk, E., Madura, J. D., Davis, M. E., Briggs, J. M., and McCammon, J. A. Simulation of enzyme–substrate encounter with gated active sites. *Nat Struct Mol Biol*, 1(1):65–69, January 1994. doi: 10.1038/nsb0194-65. 00071.

[10] Metzger, V. T., Eun, C., Kekenes-Huskey, P. M., Huber, G., and McCammon, J. A. Electrostatic Channeling in P. falciparum DHFR-TS: Brownian Dynamics and Smoluchowski Modeling. *Biophysical Journal*, 107(10):2394–2402, November 2014. ISSN 0006-3495. doi: 10.1016/j.bpj.2014.09.039. 00005.

[11] Zhang, L. and Morikis, D. 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(9):3106–3119, May 2006. ISSN 0006-3495. doi: 10.1529/biophysj.105.068130.

[12] Pyaram, K., Kieslich, C. A., Yadav, V. N., Morikis, D., and Sahu, A. Influence of electrostatics on the complement regulatory functions of Kaposica, the complement inhibitor of Kaposi's sarcoma-associated herpesvirus. *J. Immunol.*, 184(4):1956–1967, February 2010. ISSN 1550-6606. doi: 10.4049/jimmunol.0903261.

[13] El-Assaad, A. M., Kieslich, C. A., Gorham Jr., R. D., and Morikis, D. Electrostatic exploration of the C3d–FH4 interaction using a computational alanine scan. *Molecular Immunology*, 48(15–16):1844–1850, September 2011. ISSN 0161-5890. doi: 10.1016/j. molimm.2011.05.007.

[14] Kieslich, C. A., Vazquez, H., Goodman, G. N., de Victoria, A. L., and Morikis, D. The effect of electrostatics on factor H function and related pathologies. *Journal of Molecular Graphics and Modelling*, 29(8):1047–1055, August 2011. ISSN 1093-3263. doi: 10.1016/j.jmgm.2011.04.010.

[15] Gorham, R. D., Kieslich, C. A., and Morikis, D. Electrostatic Clustering and Free Energy Calculations Provide a Foundation for Protein Design and Optimization. *Ann Biomed Eng*, 39(4):1252–1263, April 2011. ISSN 0090-6964. doi: 10.1007/s10439-010-0226-9.

[16] Kieslich, C. A., Tamamis, P., D. Gorham Jr., R., Lopez de Victoria, A., U. Sausman, N., Archontis, G., and Morikis, D. Exploring Protein-Protein and Protein-Ligand Interactions in the Immune System using Molecular Dynamics and Continuum Electrostatics. *Current Physical Chemistry*, 2(4):324–343, 2012-12-01T00:00:00///.

[17] Gorham, R. D., Rodriguez, W., and Morikis, D. Molecular Analysis of the Interaction between Staphylococcal Virulence Factor Sbi-IV and Complement C3d. *Biophysical Journal*, 106(5):1164–1173, March 2014. ISSN 00063495. doi: 10.1016/j.bpj.2014.01. 033.

[18] Ojha, H., Panwar, H. S., Gorham, R. D., Morikis, D., and Sahu, A. Viral regulators of complement activation: Structure, function and evolution. *Molecular Immunology*, 61(2):89–99, October 2014. ISSN 01615890. doi: 10.1016/j.molimm.2014.06.004.

[19] Mohan, R. R., Gorham Jr., R. D., and Morikis, D. A theoretical view of the C3d:CR2 binding controversy. *Molecular Immunology*, 64(1):112–122, March 2015. ISSN 0161-5890. doi: 10.1016/j.molimm.2014.11.006. 00002.

[20] E. S. Harrison, R., Gorham, R. D., and Morikis, D. Energetic evaluation of binding modes in the C3d and Factor H (CCP 19-20) complex. *Protein Science*, 24(5):789–802, May 2015. ISSN 1469-896X. doi: 10.1002/pro.2650. 00000.

[21] Roozendaal, R. and Carroll, M. C. Complement receptors CD21 and CD35 in humoral immunity. *Immunological reviews*, 219(1):157–166, 2007.

[22] Walport, M. J. Advances in Immunology: Complement (First of Two Parts). *New England Journal of Medicine*, 344(14):1058–1066, April 2001. ISSN 0028-4793. doi: 10.1056/NEJM200104053441406. 02220.

[23] Merle, N. S., Church, S. E., Fremeaux-Bacchi, V., and Roumenina, L. T. Complement system part I – molecular mechanisms of activation and regulation. *Front. Immunol.*, page 262, 2015. doi: 10.3389/fimmu.2015.00262. 00057.

[24] Merle, N. S., Noe, R., Halbwachs-Mecarelli, L., Fremeaux-Bacchi, V., and Roumenina, L. T. Complement system part II: Role in immunity. *Front. Immunol.*, 6:257, 2015. doi: 10.3389/fimmu.2015.00257. 00030.

[25] Zipfel, P. F. and Skerka, C. Complement regulators and inhibitory proteins. *Nat Rev Immunol*, 9(10):729–740, October 2009. ISSN 1474-1733. doi: 10.1038/nri2620. 00538.

[26] Kolev, M., Friec, G. L., and Kemper, C. Complement — tapping into new sites and effector systems. *Nat Rev Immunol*, 14(12):811–820, December 2014. ISSN 1474-1733. doi: 10.1038/nri3761. 00018.

[27] Fearon, D. T. The complement system and adaptive immunity. *Seminars in Immunology*, 10(5):355–361, October 1998. ISSN 1044-5323. doi: 10.1006/smim.1998.0137.

[28] Ricklin, D. and Lambris, J. D. Complement in Immune and Inflammatory Disorders: Pathophysiological Mechanisms. *The Journal of Immunology*, 190(8):3831–3838, April 2013. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.1203487. 00107.

[29] Liszewski, M. K. and Atkinson, J. P. Complement regulators in human disease: Lessons from modern genetics. *Journal of Internal Medicine*, 277(3):294–305, March 2015. ISSN 1365-2796. doi: 10.1111/joim.12338.

[30] Morgan, B. P. and Harris, C. L. Complement, a target for therapy in inflammatory and degenerative diseases. *Nature Reviews Drug Discovery*, 14(12):857–877, October 2015. ISSN 1474-1776, 1474-1784. doi: 10.1038/nrd4657. 00002.

[31] Carroll, M. C. The complement system in regulation of adaptive immunity. *Nat Immunol*, 5(10):981–986, October 2004. ISSN 1529-2908. doi: 10.1038/ni1113. 00720.

[32] Carroll, M. C. and Isenman, D. E. Regulation of Humoral Immunity by Complement. *Immunity*, 37(2):199–207, August 2012. ISSN 1074-7613. doi: 10.1016/j.immuni.2012.08.002. 00081.

[33] Holers, V. M. Complement and Its Receptors: New Insights into Human Disease. *Annual Review of Immunology*, 32(1):433–459, 2014. doi: 10.1146/annurev-immunol-032713-120154. 00043.

[34] Zipfel, P. F. and Skerka, C. Complement: The Alternative Pathway. *eLS*, 2001.

[35] Clemenza, L. and Isenman, D. E. Structure-guided identification of C3d residues essential for its binding to complement receptor 2 (CD21). *The Journal of Immunology*, 165(7):3839–3848, 2000.

[36] van den Elsen, J. M. H. and Isenman, D. E. A Crystal Structure of the Complex Between Human Complement Receptor 2 and Its Ligand C3d. *Science*, 332(6029): 608–611, April 2011. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1201954.

[37] Morikis, D. and Lambris, J. D. The Electrostatic Nature of C3d-Complement Receptor 2 Association. *J Immunol*, 172(12):7537–7547, June 2004. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.172.12.7537.

[38] Zhang, L., Mallik, B., and Morikis, D. Immunophysical Exploration of C3d–CR2(CCP1-2) Interaction Using Molecular Dynamics and Electrostatics. *Journal of Molecular Biology*, 369(2):567–583, June 2007. ISSN 0022-2836. doi: 10.1016/j.jmb. 2007.02.101.

[39] Kieslich, C. A., Morikis, D., Yang, J., and Gunopulos, D. Automated computational framework for the analysis of electrostatic similarities of proteins. *Biotechnology Progress*, 27(2):316–325, March 2011. ISSN 87567938. doi: 10.1002/btpr.541.

[40] Hannan, J. P., Young, K. A., Guthridge, J. M., Asokan, R., Szakonyi, G., Chen, X. S., and Holers, V. M. Mutational Analysis of the Complement Receptor Type 2 (CR2/CD21)–C3d Interaction Reveals a Putative Charged SCR1 Binding Site for C3d. *Journal of Molecular Biology*, 346(3):845–858, February 2005. ISSN 0022-2836. doi: 10.1016/j.jmb.2004.12.007.

[41] Isenman, D. E., Leung, E., Mackay, J. D., Bagby, S., and van den Elsen, J. M. H. 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. *The Journal of Immunology*, 184(4):1946–1955, January 2010. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.0902919.

[42] Toapanta, F. R., DeAlmeida, D. R., Dunn, M. D., and Ross, T. M. C3d adjuvant activity is reduced by altering residues involved in the electronegative binding of C3d to CR2. *Immunology Letters*, 129(1):32–38, March 2010. ISSN 01652478. doi: 10.1016/ j.imlet.2009.12.022.

[43] Morikis, D. F1000Prime Recommendation of [van den Elsen JM and Isenman DE, Science 2011, 332(6029):608-11]. F1000Prime.com/10371956#eval14018054, September 2011.

[44] Northrup, S. H., Allison, S. A., and McCammon, J. A. Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *The Journal of Chemical Physics*, 80(4): 1517–1524, February 1984. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.446900. 00278.

[45] Gabdoulline, R. R. and Wade, R. C. On the protein-protein diffusional encounter complex. *Journal of Molecular Recognition*, 12(4):226–234, 1999. 00081.

[46] Spaar, A. and Helms, V. Ionic strength effects on the association funnel of barnase and barstar investigated by Brownian dynamics simulations. *Journal of Non-Crystalline Solids*, 352(42-49):4437–4444, November 2006. ISSN 00223093. doi: 10.1016/j.jnoncrysol.2006.03.117. 00006.

[47] Kieslich, C. A. and Morikis, D. The Two Sides of Complement C3d: Evolution of Electrostatics in a Link between Innate and Adaptive Immunity. *PLoS Computational Biology*, 8(12):e1002840, December 2012. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002840.

[48] Elcock, A. H. Molecular Simulations of Diffusion and Association in Multimacromolecular Systems. *Methods in Enzymology*, 383:166–198, 2004. 00023.

[49] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16):1781–1802, December 2005. ISSN 0192-8651, 1096-987X. doi: 10.1002/jcc.20289.

[50] Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res*, 32(Web Server issue):W665–W667, July 2004. ISSN 0305-1048. doi: 10.1093/nar/gkh381.

[51] Sitkoff, D., Sharp, K. A., and Honig, B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.*, 98(7):1978–1988, February 1994. ISSN 0022-3654. doi: 10.1021/j100058a043.

[52] Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M., and Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.*, 7(2):525–537, February 2011. ISSN 1549-9618. doi: 10.1021/ct100578z. 00000.

[53] Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M., and Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.*, 7(7):2284–2295, July 2011. ISSN 1549-9618. doi: 10.1021/ct200133y. 00000.

[54] Gorham, R. D., Kieslich, C. A., Nichols, A., Sausman, N. U., Foronda, M., and Morikis, D. An evaluation of poisson-boltzmann electrostatic free energy calculations through comparison with experimental mutagenesis data. *Biopolymers*, 95(11):746–754, 2011. ISSN 00063525. doi: 10.1002/bip.21644.

[55] Kieslich, C. A., Gorham Jr., R. D., and Morikis, D. Is the rigid-body assumption reasonable?: Insights into the effects of dynamics on the electrostatic analysis of bar-

nase–barstar. *Journal of Non-Crystalline Solids*, 357(2):707–716, January 2011. ISSN 0022-3093. doi: 10.1016/j.jnoncrysol.2010.05.087.

[56] Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *PNAS*, 98(18):10037–10041, August 2001. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.181342398.

[57] Huber, G. A. and McCammon, J. A. Browndye: A software package for Brownian dynamics. *Computer Physics Communications*, 181(11):1896–1905, November 2010. ISSN 00104655. doi: 10.1016/j.cpc.2010.07.022.

[58] Ermak, D. L. and McCammon, J. A. Brownian dynamics with hydrodynamic interactions. *The Journal of Chemical Physics*, 69(4):1352–1360, August 1978. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.436761. 01920.

[59] Luty, B. A., McCammon, J. A., and Zhou, H.-X. Diffusive reaction rates from Brownian dynamics simulations: Replacing the outer cutoff surface by an analytical treatment. *The Journal of Chemical Physics*, 97(8):5682–5686, October 1992. ISSN 0021-9606, 1089-7690. doi: 10.1063/1.463777. 00053.

[60] Huber, G. A. and Kim, S. Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophys J*, 70(1):97–110, January 1996. ISSN 0006-3495. 00138.

[61] Gabdoulline, R. R. and Wade, R. C. Brownian Dynamics Simulation of Protein–Protein Diffusional Encounter. *Methods*, 14(3):329–341, March 1998. ISSN 1046-2023. doi: 10.1006/meth.1998.0588.

[62] Huang, Y.-m. M., Huber, G., and Andrew McCammon, J. Electrostatic steering enhances the rate of cAMP binding to phosphodiesterase: Brownian dynamics modeling: Brownian Dynamics Modeling. *Protein Science*, 24(11):1884–1889, November 2015. ISSN 09618368. doi: 10.1002/pro.2794. 00000.

[63] Votapka, L. W. and Amaro, R. E. Multiscale Estimation of Binding Kinetics Using Brownian Dynamics, Molecular Dynamics and Milestoning. *PLOS Computational Biology*, 11(10):e1004381, October 2015. ISSN 1553-7358. doi: 10.1371/journal.pcbi. 1004381. 00000.

[64] Spaar, A. and Helms, V. Free Energy Landscape of Protein-Protein Encounter Resulting from Brownian Dynamics Simulations of Barnase:Barstar. *Journal of Chemical Theory and Computation*, 1(4):723–736, July 2005. ISSN 1549-9618, 1549-9626. doi: 10.1021/ct050036n. 00034.

[65] Szakonyi, G., Guthridge, J. M., Li, D., Young, K., Holers, V. M., and Chen, X. S. Structure of Complement Receptor 2 in Complex with Its C3d Ligand. *Science*, 292 (5522):1725–1728, January 2001. ISSN 0036-8075, 1095-9203. doi: 10.1126/science. 1059118.

[66] Mohan, R. R., Huber, G. A., and Morikis, D. Electrostatic Steering Accelerates C3d:CR2 Association. *J. Phys. Chem. B*, April 2016. ISSN 1520-6106. doi: 10.1021/acs.jpcb.6b02095. 00000.

[67] Thurman, J. M., Kulik, L., Orth, H., Wong, M., Renner, B., Sargsyan, S. A., Mitchell, L. M., Hourcade, D. E., Hannan, J. P., Kovacs, J. M., Coughlin, B., Woodell, A. S., Pickering, M. C., Rohrer, B., and Holers, V. M. Detection of complement activation using monoclonal antibodies against C3d. *Journal of Clinical Investigation*, 123(5): 2218–2230, May 2013. ISSN 0021-9738. doi: 10.1172/JCI65861. 00022.

[68] Gorham, R. D., Nuñez, V., Lin, J., Rooijakkers, S. H. M., Vullev, V. I., and Morikis, D. Discovery of Small Molecules for Fluorescent Detection of Complement Activation Product C3d. *J. Med. Chem.*, November 2015. ISSN 0022-2623. doi: 10.1021/acs. jmedchem.5b01062. 00000.

[69] Serkova, N. J., Renner, B., Larsen, B. A., Stoldt, C. R., Hasebroock, K. M., Bradshaw-Pierce, E. L., Holers, V. M., and Thurman, J. M. Renal Inflammation: Targeted Iron Oxide Nanoparticles for Molecular MR Imaging in Mice. *Radiology*, 255(2):517–526, April 2010. ISSN 0033-8419. doi: 10.1148/radiol.09091134. 00041.

[70] Ricklin, D. and Lambris, J. D. Complement in Immune and Inflammatory Disorders: Therapeutic Interventions. *J Immunol*, 190(8):3839–3847, April 2013. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.1203200. 00094.

[71] Song, H., He, C., Knaak, C., Guthridge, J. M., Holers, V. M., and Tomlinson, S. Complement receptor 2–mediated targeting of complement inhibitors to sites of complement activation. *J Clin Invest*, 111(12):1875–1885, June 2003. ISSN 0021-9738. doi: 10.1172/JCI200317348. 00097.

[72] Atkinson, C., Song, H., Lu, B., Qiao, F., Burns, T. A., Holers, V. M., Tsokos, G. C., and Tomlinson, S. Targeted complement inhibition by C3d recognition ameliorates tissue injury without apparent increase in susceptibility to infection. *J Clin Invest*, 115 (9):2444–2453, September 2005. ISSN 0021-9738. doi: 10.1172/JCI25208. 00122.

[73] Rohrer, B., Long, Q., Coughlin, B., Wilson, R. B., Huang, Y., Qiao, F., Tang, P. H., Kunchithapautham, K., Gilkeson, G. S., and Tomlinson, S. A Targeted Inhibitor of the Alternative Complement Pathway Reduces Angiogenesis in a Mouse Model of Age-Related Macular Degeneration. *Invest Ophthalmol Vis Sci*, 50(7):3056–3064, July 2009. ISSN 0146-0404. doi: 10.1167/iovs.08-2222. 00107.

[74] Schmidt, C. Q., Bai, H., Lin, Z., Risitano, A. M., Barlow, P. N., Ricklin, D., and Lambris, J. D. Rational Engineering of a Minimized Immune Inhibitor with Unique Triple-Targeting Properties. *J Immunol*, 190(11):5712–5721, January 2013. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.1203548. 00032.

[75] Fridkis-Hareli, M., Storek, M., Mazsaroff, I., Risitano, A. M., Lundberg, A. S., Horvath, C. J., and Holers, V. M. Design and development of TT30, a novel C3d-targeted C3/C5

convertase inhibitor for treatment of human complement alternative pathway–mediated diseases. *Blood*, 118(17):4705–4713, October 2011. ISSN 0006-4971, 1528-0020. doi: 10.1182/blood-2011-06-359646. 00052.

[76] Holers, V. M., Rohrer, B., and Tomlinson, S. CR2-Mediated Targeting of Complement Inhibitors: Bench-to-Bedside Using a Novel Strategy for Site-Specific Complement Modulation. In Lambris, J. D., Holers, V. M., and Ricklin, D., editors, *Complement Therapeutics*, volume 735, pages 137–154. Springer US, Boston, MA, 2013. ISBN 978-1-4614-4117-5 978-1-4614-4118-2. 00000.

[77] McCammon, J. A. Darwinian biophysics: Electrostatics and evolution in the kinetics of molecular binding. *PNAS*, 106(19):7683–7684, December 2009. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0902767106. 00015.

[78] Schreiber, G., Buckle, A. M., and Fersht, A. R. Stability and function: Two constraints in the evolution of barstar and other proteins. *Structure*, 2(10):945–951, October 1994. ISSN 0969-2126. doi: 10.1016/S0969-2126(94)00096-4. 00156.

Chapter 4

High-throughput screening for discovery of C3d-binding small molecules

4.1 Introduction

The complement system is integral to innate immunity and plays a crucial role as the link between innate and adaptive immunity[1]. Through cascading responses, the complement system is the first line of defense against pathogens or injury via processes such as opsonization, phagocytosis, and cell lysis[2]. Through regulation complement response is capable of discerning self vs. non-self but lack of regulation or misregulation can result in the pathology of autoimmune and inflammatory diseases (such as age-related macular degeneration)[3–6].

Complement activation can occur through the classical, alternative, and lectin pathways with all three pathways converging at complement C3[7]. C3b and iC3b, cleavage products of C3, are opsonins through the covalent attachment of the TED domain, also known as C3d. Standalone C3d acts as a link between innate and adaptive immunity. Because C3d is the final degradation product of C3, it is a natural biomarker of complement activation. Although there have been previous efforts to target C3d as a biomarker for complement response, these efforts are primarily focused on biologics[8,9] which have limitations in bioavailability and scalability in production. Small molecules address these limitations through their higher bioavailability, modes of delivery and cost of production. Virtual screening studies have shown promise in targeting complement proteins[10]. Specifically, a previous study to identify C3d-binding small molecules with fluorescence properties has demonstrated success[11] and in this study we continue those efforts with enhanced methods and a more chemically diverse database. Here we outline a virtual screening workflow utilizing conformer generation, tICA clustering-based receptor ensemble, and robust docking and scoring to identify new small molecules with C3d-binding capability (as seen in Figure 4.1).
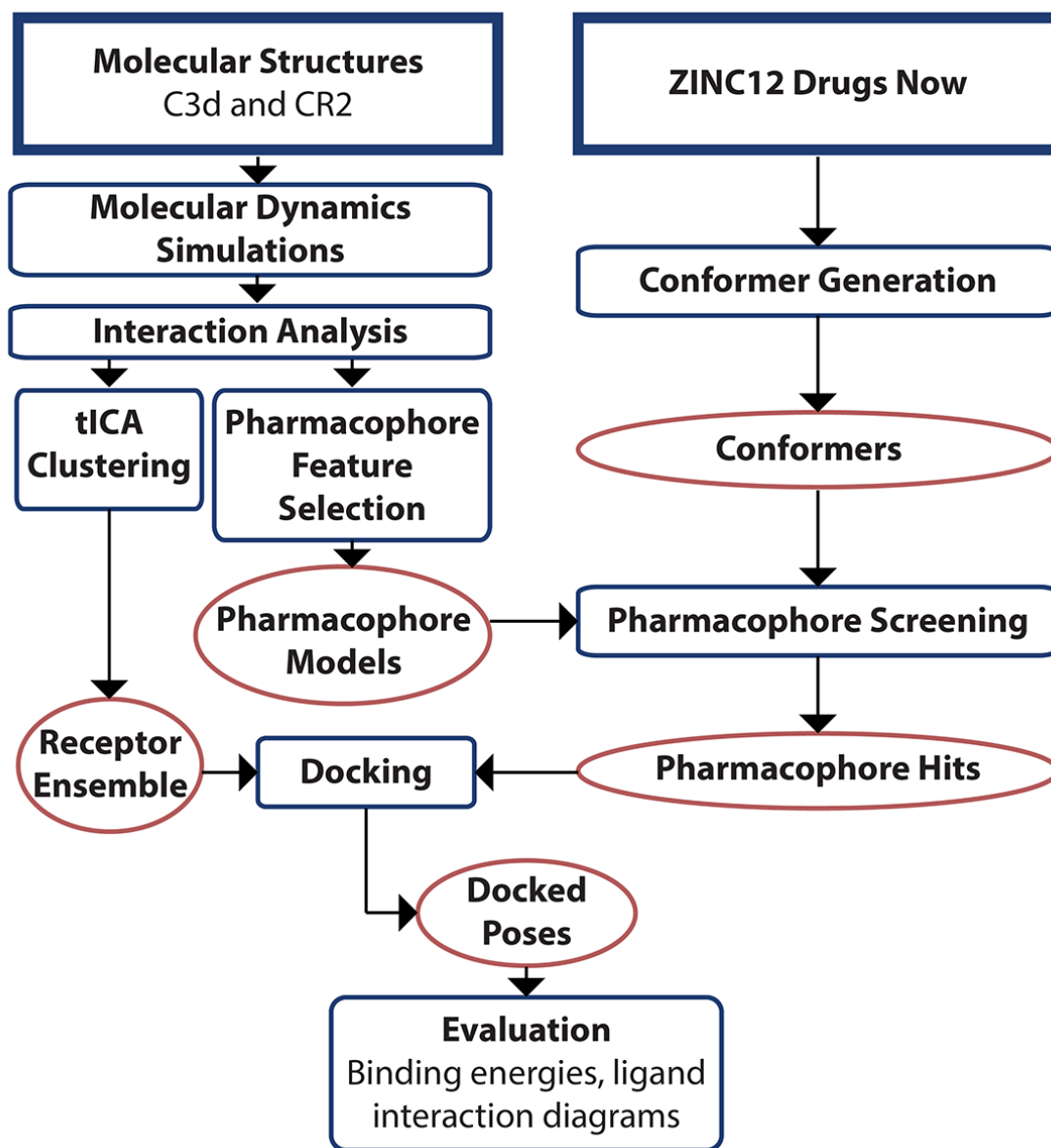
Figure 4.1: A schematic of the virtual screening workflow.

4.2 Methods

4.2.1 Molecular Dynamics Simulations

The crystallographic structure of C3d:CR2 (PDB: 3OED[12]) was obtained from the Protein Data Bank[13]. In the structure, there are two complexes for C3d:CR2 but we used the complex with better electron density and lower B-factor values, corresponding to chain A and chain C for C3d and CR2 respectively. Crystallographic waters were retained in the structure.

Explicit-solvent molecular dynamics (MD) simulations were carried out using the input structure. Initial minimization of the structure without solvent was carried out in NAMD[14] using the CHARMM36 forcefield[15]. The structures were then solvated using TIP3P water boxes with dimensions of $83\,\text{Å} \times 83\,\text{Å} \times 80\,\text{Å}$ and neutralized with sodium and chloride counterions at an ionic strength of 150 mM. Subsequent to addition of solvent, the structures were underwent 25,000 steps of conjugate gradient energy minimization followed by heating from 0 to $300\,\text{K}$ in $62\,\text{ps}$ with all protein atoms constrained to their post-minimization positions. Following heating, the system was equilibrated through five stages for $10\,\text{ns}$ in the first stage and $5\,\text{ns/stage}$ for the last four stages. Force constants 41.84, 20.92, 8.368 and $4.184\,\text{kJ/mol/Å}^2$ were applied during the first four stages respectively to harmonically constrain all protein atoms to their post-minimization positions. During the final stage of equilibration a force constant of $4.184\,\text{kJ/mol/Å}^2$ was applied to constrain only protein backbone atoms to their post-minimization positions. Following equilibration, production runs were carried out for $100\,\text{ns}$ in triplicate using AMBER16[16] with: periodic boundary conditions, Langevin dynamics, nonbonded interaction cutoff of $12\,\text{Å}$, SHAKE

algorithm, and an integration timestep of 2 fs. Analysis of the simulation trajectories was carried out using MDTraj[17].

4.2.2   Receptor Ensemble

In order to extract representative structures for the creation of a receptor ensemble, tICA decomposition was performed on the phi and psi angles observed throughout the trajectory using MSMBuilder[18]. The MiniBatchKMeans method in MSMBuilder was utilized to cluster the components to five distinct clusters and cluster centers were extracted as representative structures for the C3d:CR2 complex, from which representative C3d structures were exracted. Visualization of clustering and decomposition was performed using MSMExplorer[19].

4.2.3   Pharmacophore Model Generation

A set of mock pharmacophore features (from a prior study[11]) representing conserved hydrogen bonding interactions and favorable hydrophobic contacts in C3d:CR2 informed the choice of pharmacophore features. Pharmacophore feature positions and tolerance radii were assigned by calculating center-of-mass positions and conformational flexibility of successful hits from the prior virtual screening study[11]. After identifying a total of 7 pharmacophore features (Figure 4.2), interfeature distances and inclusion of hydrophobic features informed the selection of subsets of 4-5 features as individual pharmacophore models. A total of 10 pharmacophore models were defined with each model having at least one hydrophobic feature, a requirement chosen for its proposed importance in anchoring the small molecule in the C3d cavity[11]. The models were screened against the Drugs Now

subset in ZINC12[20], consisting of 10.6 million small molecules. Conformer generation was performed using Phase[21,22], with 203 conformers generated per molecule on average. Exclusion volumes were utilized to prevent overlap with C3d residues.

4.2.4 Docking and Scoring

Molecular docking was performed using Glide[23–25] using the XP scoring method for added accuracy. All molecules identified in pharmacophore models with total number of hits $<$ 10,000 were docked to the CR2 binding site of C3d using the each structure from the receptor ensemble generated through analysis of the MD trajectory. The inner grid and outer grid dimensions were set to $12\,\text{Å} \times 12\,\text{Å} \times 12\,\text{Å}$ and $32\,\text{Å} \times 32\,\text{Å} \times 32\,\text{Å}$ respectively. A total of 5 docking runs, for each representative C3d structure from the receptor ensemble, were carried out. Epik[26,27] penalties were incorporated into scoring so that higher energy states are accounted for. The top 20 docked poses from each molecule were retained.

For pharmacophore models with $>$10,000 hits, the hits were docked first using the HTVS method in Glide and the top 5000 docked molecules were then inputs for the XP docking method as described above. Pharmacophore models with $>$100,000 hits were excluded from analysis as the models are too lenient.
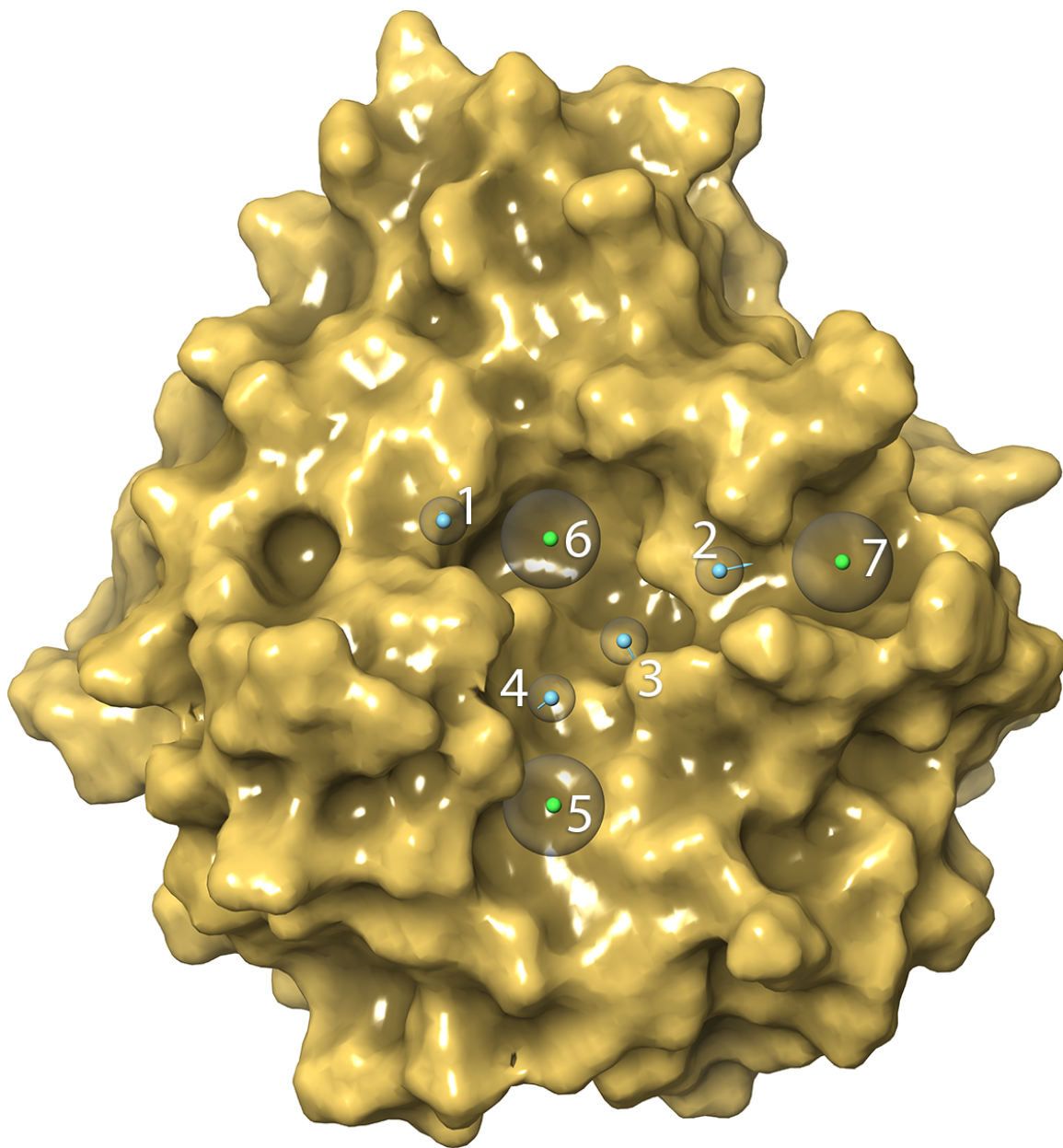
Figure 4.2: Molecular graphics demonstrating the distribution of pharmacophore features identified for pharmacophore model generation. Features 1, 2, 3, and 4 are hydrogen bond donor/acceptor features while features 5, 6, and 7 correspond to hydrophobic and/or aromatic features.

4.3 Results and Discussion

The results of the pharmacophore screen are outlined in Table 4.1. Each pharmacophore model included at least one hydrophobic feature in an attempt to identify compounds that are anchored into the hydrophobic cavity of C3d. The three identified hydrophobic features (H5, H6, and H7) were chosen based on persistent chemical moieties from previously identified C3d-binding ligands [11] and simulation trajectories of the C3d:CR2 complex. Additionally, the features were chosen based on their proximity to hydrophobic C3d amino acids as follows: H5 - M206, I176, K162; H6 - L95, H33, I30; H7 - L248, F253, and Y289.

Model 5 is the only model with 0 hits and it may be due to the constrained angles the model lends itself to. Models 7 and 10 resulted in 500,000 hits (the max allowed during screening) and are too lenient in criteria, likely due to the ubiquity of hydrophobic/aromatic features and the close proximity of hydrogen bond donor/acceptor features.

During docking of the results from the pharmacophore screens, models 4 and 6 went through two rounds of docking to filter the large number of hits: 1) an initial high-throughput docking with HTVS scoring and 2) a second docking round with the top 5000 hits from the first round using XP scoring as with the other models. The top 10 compounds identified during docking are outlined in Table 4.2 and are visualized in Figures 4.3 and 4.4.

All of the top 10 compounds were identified through screening model 1. This may be due to the compounds matching six key pharmacophore features including the hy-

Table 4.1: Pharmacophore Screening Results

Model #	Pharmacophore Features	Hits
1	1, 2, 3, 4, 6	1368
2	2, 3, 4, 5, 6	64
3	1, 2, 3, 6	1254
4	1, 2, 4, 6	57,524
5	1, 3, 4, 6	0
6	2, 3, 4, 5	50,175
7	2, 3, 4, 6	500,000
8	2, 3, 5, 6	45
9	2, 4, 5, 6	4389
10	3, 4, 5, 6	500,000

Table of pharmacophore models and corresponding features and results from screening. Features 1, 2, 3, and 4 are capable of matching either hydrogen bond donor or acceptor features while features 5, 6, and 7 are capable of matching hydrophobic and/or aromatic features.

Table 4.2: Top 10 Docked Compounds

Rank	ZINCID	Calculated Binding Affinity (kcal/mol)
1	ZINC14709426	-9.5
2	ZINC20590128	-9.0
3	ZINC01236065	-8.5
4	ZINC35456133	-8.5
5	ZINC31158776	-8.5
6	ZINC59585888	-8.4
7	ZINC01692153	-8.1
8	ZINC12111684	-8.1
9	ZINC27541822	-8.0
10	ZINC35465754	-8.0
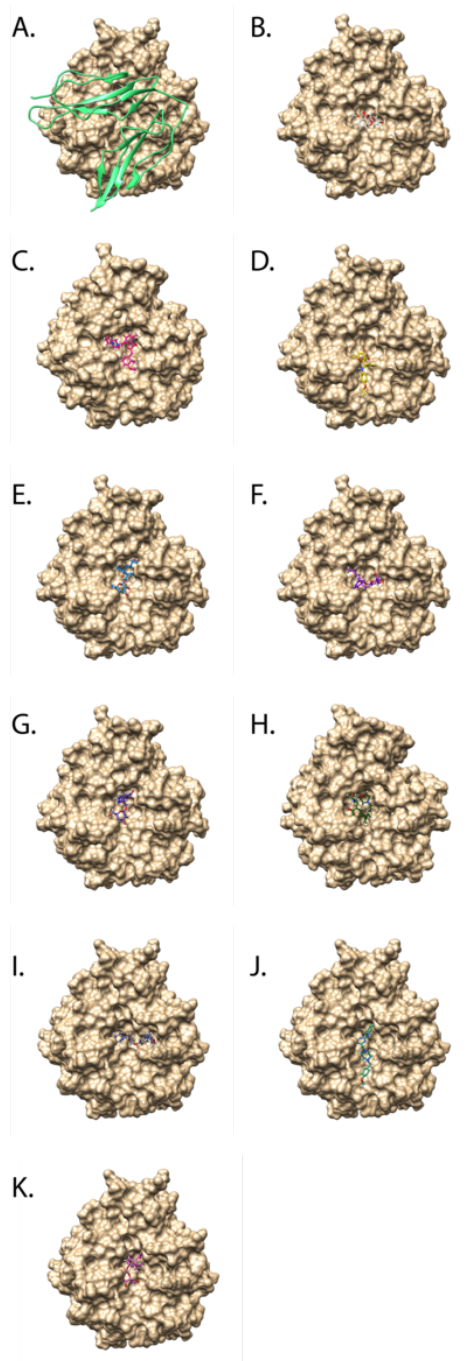
Figure 4.3: Molecular graphics of (A) CR2 bound to C3d and (B-K) the top ten compounds docked to C3d.
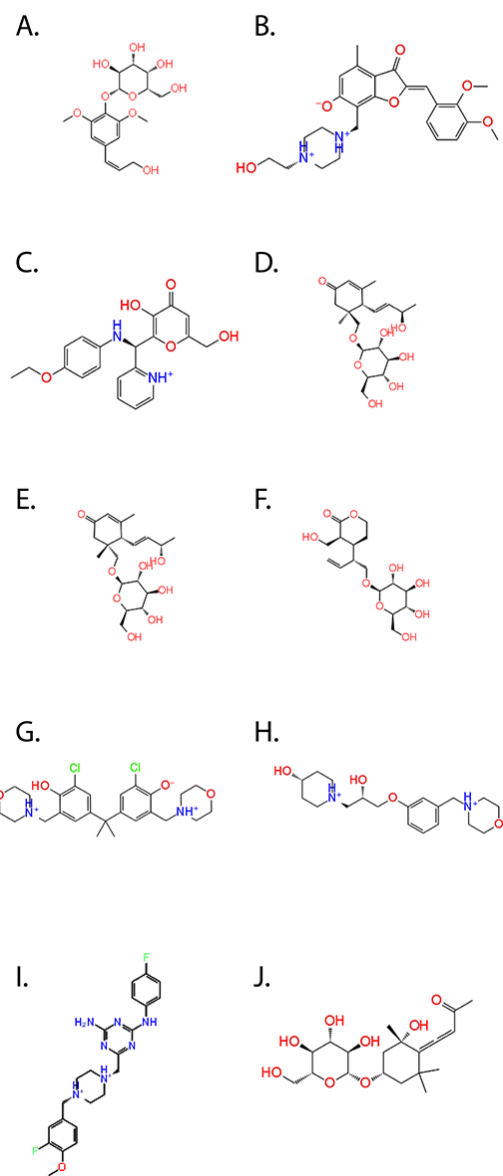
Figure 4.4: 2D structures of (A-J) the top ten compounds.

drophobic feature in the cavity through the course of screening. Of the compounds that were docked, only one of the compounds tested experimentally in the prior C3d virtual screening study[11] was identified during screening (ZINC14749391) but it was not found within the top 10 docked compounds. Ligand interaction diagrams of the top 4 compounds demonstrate key conserved interactions between C3d and the docked compounds (Figure 4.5). C3d residues participating in conserved interactions include Asp163, Lys91, and Gln205 which are all suggested to participate in key salt bridges and hydrogen bonds with CR2[28]. One of the residues participating in conserved interactions that was not identified in prior analyses of the C3d:CR2 complex is Glu160, which is located further into the cavity of C3d. This suggests that the inclusion of the hydrophobic feature in the cavity of C3d is allowing for identification of additional favorable interactions that may aid in anchoring the small molecule to the cavity of C3d.

Using an in-silico high-throughput screening approach we have identified a subset of small molecules that exhibit favorable physicochemical features and predicted binding affinities. Binding capabilities and fluorescence need to be characterized through further experimental methodologies. These compounds provide a basis for the potential development of theranostic small molecules for characterizing complement-induced disease progression, such as age-related macular degeneration, through noninvasive imaging.
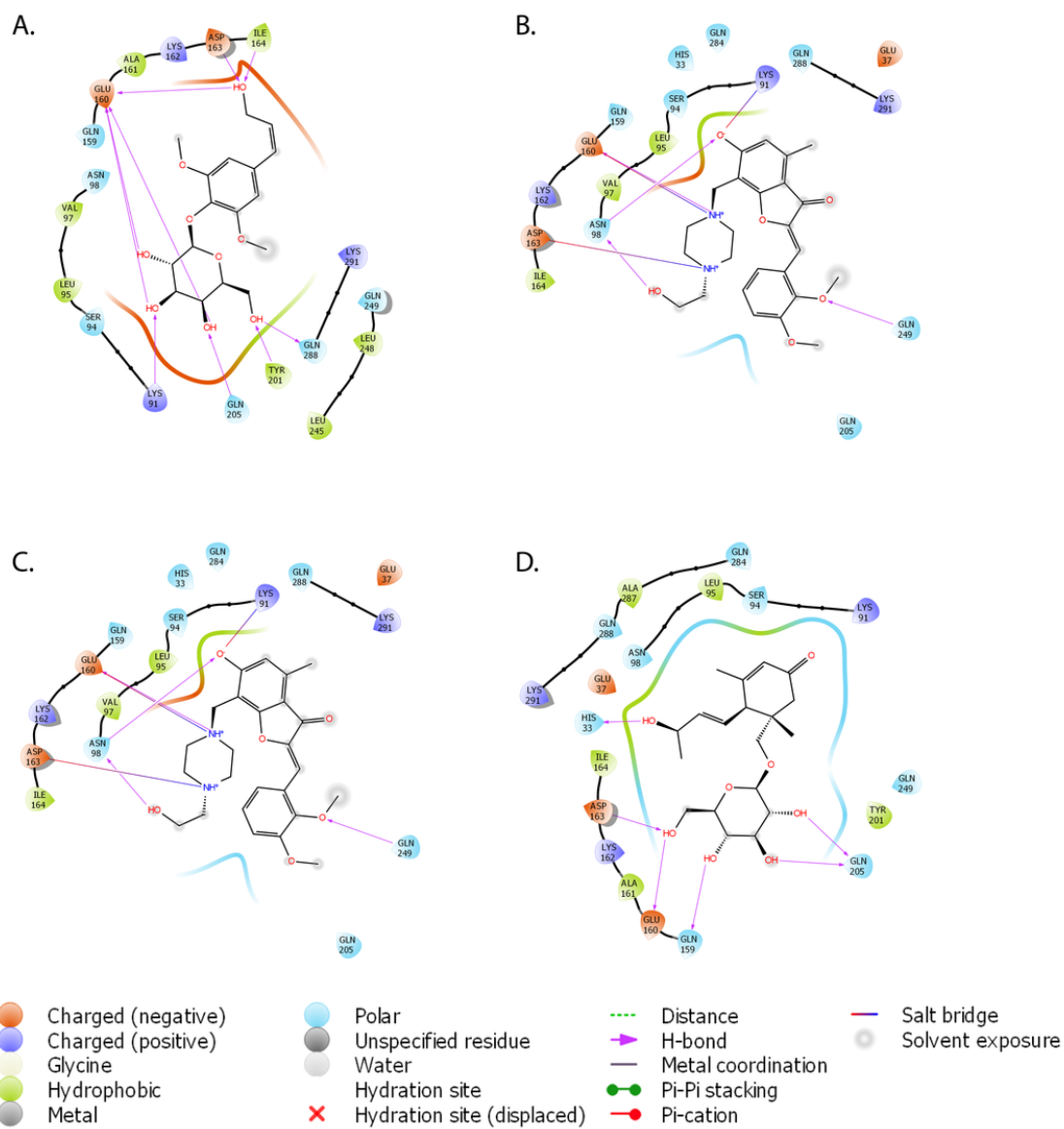
Figure 4.5: (A-D) Ligand interaction diagrams generated in Schrodinger Maestro outlining the various interactions between key C3d residues and the top 4 docked compounds.

4.4 References

[1] Fearon, D. T. The complement system and adaptive immunity. *Seminars in Immunology*, 10(5):355–361, October 1998. ISSN 1044-5323. doi: 10.1006/smim.1998.0137.

[2] Roozendaal, R. and Carroll, M. C. Complement receptors CD21 and CD35 in humoral immunity. *Immunological reviews*, 219(1):157–166, 2007.

[3] Ricklin, D. and Lambris, J. D. Complement in Immune and Inflammatory Disorders: Pathophysiological Mechanisms. *The Journal of Immunology*, 190(8):3831–3838, April 2013. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.1203487. 00107.

[4] Morgan, B. P. and Harris, C. L. Complement, a target for therapy in inflammatory and degenerative diseases. *Nature Reviews Drug Discovery*, 14(12):857–877, October 2015. ISSN 1474-1776, 1474-1784. doi: 10.1038/nrd4657. 00002.

[5] Liszewski, M. K. and Atkinson, J. P. Complement regulators in human disease: Lessons from modern genetics. *Journal of Internal Medicine*, 277(3):294–305, March 2015. ISSN 1365-2796. doi: 10.1111/joim.12338.

[6] Schramm, E. C., Clark, S. J., Triebwasser, M. P., Raychaudhuri, S., Seddon, J. M., and Atkinson, J. P. Genetic variants in the complement system predisposing to age-related macular degeneration: A review. *Molecular Immunology*, 61(2):118–125, October 2014. ISSN 01615890. doi: 10.1016/j.molimm.2014.06.032. 00029.

[7] Zipfel, P. F. and Skerka, C. Complement: The Alternative Pathway. *eLS*, 2001.

[8] Serkova, N. J., Renner, B., Larsen, B. A., Stoldt, C. R., Hasebroock, K. M., Bradshaw-Pierce, E. L., Holers, V. M., and Thurman, J. M. Renal Inflammation: Targeted Iron Oxide Nanoparticles for Molecular MR Imaging in Mice. *Radiology*, 255(2):517–526, April 2010. ISSN 0033-8419. doi: 10.1148/radiol.09091134. 00041.

[9] Thurman, J. M., Kulik, L., Orth, H., Wong, M., Renner, B., Sargsyan, S. A., Mitchell, L. M., Hourcade, D. E., Hannan, J. P., Kovacs, J. M., Coughlin, B., Woodell, A. S., Pickering, M. C., Rohrer, B., and Holers, V. M. Detection of complement activation using monoclonal antibodies against C3d. *Journal of Clinical Investigation*, 123(5): 2218–2230, May 2013. ISSN 0021-9738. doi: 10.1172/JCI65861. 00022.

[10] Mohan, R. R., Wilson, M., Gorham, R. D., Harrison, R. E. S., Morikis, V. A., Kieslich, C. A., Orr, A. A., Coley, A. V., Tamamis, P., and Morikis, D. Virtual Screening of Chemical Compounds for Discovery of Complement C3 Ligands. *ACS Omega*, 3(6): 6427–6438, June 2018. ISSN 2470-1343. doi: 10.1021/acsomega.8b00606.

[11] Gorham, R. D., Nuñez, V., Lin, J., Rooijakkers, S. H. M., Vullev, V. I., and Morikis, D. Discovery of Small Molecules for Fluorescent Detection of Complement Activation Product C3d. *J. Med. Chem.*, November 2015. ISSN 0022-2623. doi: 10.1021/acs.jmedchem.5b01062. 00000.

[12] van den Elsen, J. M. H. and Isenman, D. E. A Crystal Structure of the Complex Between Human Complement Receptor 2 and Its Ligand C3d. *Science*, 332(6029): 608–611, April 2011. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1201954.

[13] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. The Protein Data Bank. *Nucl. Acids Res.*, 28(1): 235–242, January 2000. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/28.1.235.

[14] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16):1781–1802, December 2005. ISSN 0192-8651, 1096-987X. doi: 10.1002/jcc.20289.

[15] MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B*, 102(18): 3586–3616, April 1998. ISSN 1520-6106. doi: 10.1021/jp973084f.

[16] Case, D. A., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., Iii, V. W. D. C., Darden, T. A., Duke, R. E., Ghoreishi, D., Gilson, M. K., Gohlke, H., Goetz, A. W., Greene, D., Harris, R., Homeyer, N., Izadi, S., Kovalenko, A., Kurtzman, T., Lee, T. S., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Mermelstein, D. J., Merz, K. M., Miao, Y., Monard, G., Nguyen, C., Nguyen, H., Omelyan, I., Onufriev, A., Pan, F., Qi, R., Roe, D. R., Roitberg, A., Sagui, C., Schott-Verdugo, S., Shen, J., Simmerling, C. L., Smith, J., Salomon-Ferrer, R., Swails, J., Walker, R. C., Wang, J., Wei, H., Wolf, R. M., Wu, X., Xiao, L., York, D. M., and Kollman, P. A. *AMBER 2018*. University of California, San Francisco, 2018.

[17] McGibbon, R. T., Beauchamp, K. A., Harrigan, M. P., Klein, C., Swails, J. M., Hernández, C. X., Schwantes, C. R., Wang, L.-P., Lane, T. J., and Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal*, 109(8):1528–1532, October 2015. ISSN 0006-3495. doi: 10.1016/j.bpj.2015.08.015.

[18] Beauchamp, K. A., Bowman, G. R., Lane, T. J., Maibaum, L., Haque, I. S., and Pande, V. S. MSMBuilder2: Modeling Conformational Dynamics on the Picosecond to Millisecond Scale. *J. Chem. Theory Comput.*, 7(10):3412–3419, October 2011. ISSN 1549-9618. doi: 10.1021/ct200463m.

[19] Hernández, C. X., Harrigan, M. P., Sultan, M. M., and Pande, V. S. MSMExplorer: Data Visualizations for Biomolecular Dynamics. http://joss.theoj.org, April 2017.

[20] Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., and Coleman, R. G. ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.*, 52(7):1757–1768, July 2012. ISSN 1549-9596. doi: 10.1021/ci3001277.

[21] Dixon, S. L., Smondyrev, A. M., and Rao, S. N. PHASE: A Novel Approach to Pharmacophore Modeling and 3D Database Searching. *Chemical Biology & Drug Design*, 67(5):370–372, May 2006. ISSN 1747-0285. doi: 10.1111/j.1747-0285.2006.00384.x.

[22] Dixon, S. L., Smondyrev, A. M., Knoll, E. H., Rao, S. N., Shaw, D. E., and Friesner, R. A. PHASE: A new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J Comput Aided Mol Des*, 20(10-11):647–671, October 2006. ISSN 0920-654X, 1573-4951. doi: 10.1007/s10822-006-9087-6.

[23] Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.*, 47(7):1739–1749, March 2004. ISSN 0022-2623. doi: 10.1021/jm0306430.

[24] Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., and Mainz, D. T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.*, 49(21):6177–6196, October 2006. ISSN 0022-2623. doi: 10.1021/jm051256o.

[25] Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., and Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.*, 47(7):1750–1759, March 2004. ISSN 0022-2623. doi: 10.1021/jm030644s.

[26] Shelley, J. C., Cholleti, A., Frye, L. L., Greenwood, J. R., Timlin, M. R., and Uchimaya, M. Epik: A software program for pKaprediction and protonation state generation for drug-like molecules. *J Comput Aided Mol Des*, 21(12):681–691, December 2007. ISSN 1573-4951. doi: 10.1007/s10822-007-9133-z.

[27] Greenwood, J. R., Calkins, D., Sullivan, A. P., and Shelley, J. C. Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *J Comput Aided Mol Des*, 24(6):591–604, June 2010. ISSN 1573-4951. doi: 10.1007/s10822-010-9349-1.

[28] Mohan, R. R., Gorham Jr., R. D., and Morikis, D. A theoretical view of the C3d:CR2 binding controversy. *Molecular Immunology*, 64(1):112–122, March 2015. ISSN 0161-5890. doi: 10.1016/j.molimm.2014.11.006. 00002.

Chapter 5

CR2-based peptide design

5.1 Introduction

Within the complement system, the interaction between complement fragment C3d and complement receptor CR2 plays an important role as the link between innate and adaptive immunity. C3d is a marker for complement activation, but also functions to stimulate B-cell activation and production of antibodies through its interaction with CR2. As unregulated complement response is implicated in several inflammatory disorders and other diseases, C3d can be utilized both as a biomarker for tracking disease progression as well as a target for inhibition due to its role as an opsonin and in augmenting B-cell mediated antibody production. One of the major factors in the development and progression of age-related macular degeneration (AMD) is complement response [1,2]. As AMD progresses and drusen accumulation continues, C3d is observed locally at the RPE-Bruch's membrane interface [3,4]. This behavior of C3d as an opsonin and marker of complement activation makes it amenable as a biomarker target to characterize the progression of and diagnose AMD.

By designing CR2-based peptides we can leverage the specific intermolecular interactions that drive the stability and mechanics of the C3d:CR2 complex.

Previous in-silico studies have characterized the dynamics and persistent intermolecular interactions of the C3d:CR2 interaction through molecular dynamics simulations[5], computational mutagenesis[6], and Poisson Boltzmann electrostatics calculations[7]. These findings are corroborated by experimental mutagenesis[8–10] and epitope mapping studies[11,12]. These studies emphasize the importance of specific amino acids from the SCR1 and SCR2 domains of CR2 to the binding interface of C3d. By leveraging the mechanistic information of the C3d:CR2 interaction gleaned through prior studies, we implement rational and combinatorial peptide design derived from key CR2 contacts with C3d. We explore different structural approaches with different cyclization schemes and explore variations in physicochemical properties of the amino acids and their effects on solubility and more importantly, binding. After an iterative approach and We identify two peptides from the SCR1 and SCR2 domain respectively with C3d-binding capability. Subsequent to identification of peptides with C3d-binding capability, we explore conjugation with fluorophores that emit in near infrared, which is suitable for *in vivo* research, thus allowing for eventual clinical applications as a theranostic for elucidating progression of and aiding in the diagnosis of AMD.

5.2   Methods

5.2.1   Structural Analysis and Rational Design

The initial three proposed peptides were chosen using β sheets on the SCR1 and SCR2 domains of CR2 containing significant intermolecular interactions identified through prior molecular dynamics, computational mutagenesis, experimental mutagenesis, and epitope mapping studies[5,6,8–10] as seen in Figure 5.1. Proposed peptides were limited to no longer than 22 peptides in length as longer peptides tend to have structural and biological stability issues. Additionally, all the initial peptides are cyclic as it is preferred for therapeutic applications. Cyclization reduces the conformational space available to the peptide, structurally restraining the peptide to a smaller number of conformations and thus making the peptide more amenable to conformational selection for a binding target. Structures of the C3d:peptide complexes were generated using Chimera[13] by truncating CR2 from the crystallographic structure of the C3d:CR2 complex (PDB: 3OED)[14]. Subsequent iterations and mutations chosen through rational design were carried out using Chimera as well. More complicated mutation and truncation schemes as with peptide 2.2_1 were performed using modeller[15]. Rational design, in addition to being informed by the prior studies outlined above, was informed by evaluation of conservative mutations as outlined in the Blosum62 matrix[16] as well as the β-turns propensity potentials of amino acids to be introduced (in cases where the mutation was performed at the β-turn region)[17].
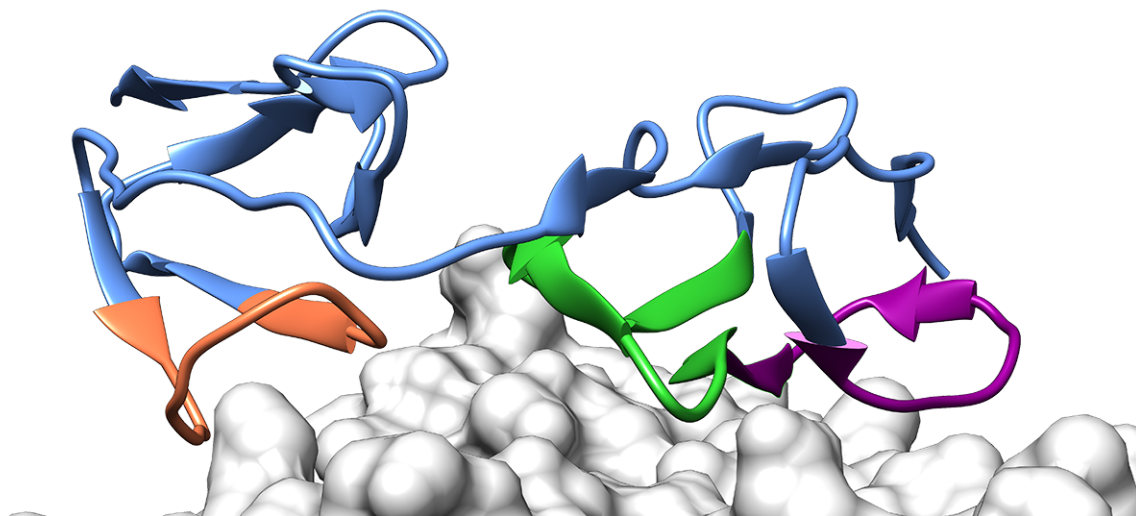
Figure 5.1: Molecular graphics of the three proposed peptides and the surface of the binding site on C3d. (B-D) show Peptides 1-3 respectively.

5.2.2 Combinatorial Optimization

Combinatorial optimization was carried out using the sequence_tolerance application from Rosetta[18,19] with the structures generated above as input. First, a resfile is generated specifying which input peptide residues should be mutated as well as the possible mutations e.g all but cysteine. C3d residues within 5 Å of the input peptide were included in the resfile as well. Ex1, ex2, extrachi_cutoff parameters were included to increase the resolution of the rotamer library. The fitness_master_weights parameter was set to 0.4, 0.4, 0.4 and 1.0 respectively to prioritize intermolecular contact optimization. The number of generations in the genetic algorithm was set to 10 with the number of sequences per generation set to 4000. The number of sequences to include in the initial population that are

created by mutation of the sequence determined by full redesign was set to 1. The reference energy of histidine was increased by 1.2 score units as outlined in the protocol for the sequence_tolerance application for added accuracy [18].

5.2.3 Microscale Thermophoresis

Selected peptides were synthesized for evaluation of C3d-binding capabilities. Peptides 1.1, 2.1, 3.2_1, and 3.2_2 were synthesized by Abgent, a Wuxi-Apptec company (San Diego, CA, USA) while peptides 1.2, 1.3, 1.4, 2.2, 2.2_1, 2.5, 2.6, 3.3, 3.6, and 3.9 were synthesized by ELIM Biopharm (Hayward, CA, USA).

The binding affinity of the peptides synthesized were evaluated in a competitive microscale thermophoresis (MST) assay, using a Monolith NT.115 instrument (NanoTemper Technologies GmbH, Munich, Germany). A 1:1 serial dilution series of each of the peptides was performed in MST buffer (50 mM Tris-HCl, 150 mM NaCl, 10 mM MgCl2, 0.05% Tween 20). Each dilution series was incubated with purified C3d (Complement Technology, Inc., Tyler, TX, USA), labeled with NT547 dye using the kit from NanoTemper Technologies, at a final C3d concentration of 96 nM for 5 minutes at room temperature in the dark. Samples were loaded into Monolith NT.115 Premium Capillaries from NanoTemper Technologies and thermophoretic response was measured. Each dilution series was performed in triplicate and estimation of the K_D was performed through nonlinear regression.

5.3 Results and Discussion

5.3.1 Rational and Combinatorial Optimization

Design of all three proposed peptides was informed by the mechanistic analysis of the C3d:CR2 interaction. Evaluation of significant intermolecular and intramolecular interactions through electrostatic and molecular dynamics simulation analysis, in addition to physicochemical properties such as hydrophobicity, aromaticity, and propensity for formation of β-turns played a significant role in the choice of potential peptide sequences. Prior epitope mapping studies identifying regions of CR2 significant to the C3d:CR2 interaction[11,12] informed the design process as well. In early iterations of the peptides optimization was introduced through minor conservative mutations while some of the later iterations expanded to larger number of mutations through combinatorial and rational design.

CR2-based Peptide 1

The proposed peptide 1 spanned the residue range of 12-31 in the SCR1 domain of CR2. The results of several rounds of iteration and rational optimization are listed in Table 5.1. Peptide 1.1 entails mutating G12 to C in order to cyclize the peptide through a disulfidge bridge with C31. Cyclization was included in the design to maintain the structure of the β-sheet. Peptide 1.1 demonstrated poor solubility during experimental validation so two conservative mutations to Threonine were introduced in Peptide 1.2 but this did not completely resolve the solubility issues. In an attempt to address the solubility issue while retaining amino acids important to intermolecular interactions, smaller, linearized Peptide 1.3 and Peptide 1.4 were designed and tested. Although the reduction in overall

hydrophobicity of the peptides resolved the solubility issues these peptides did not exhibit any binding to C3d.

Table 5.1: CR2_pep1 optimization scheme

Peptide	Residue Range[†]	Sequence	Optimization	$K_D \pm SE$ (µM)
1.1	12-31	Ac-[CRISYYSTPIAVGTVIRYSC]-NH$_2$	Rational	Insoluble
1.2	12-31	Ac-[CRISYYSTPTATGTVIRYSC]-NH$_2$	Rational	Insoluble
1.3	13-18	Ac-RISYYS-NH$_2$	Rational	No binding
1.4	24-28	Ac-GTVIR-NH$_2$	Rational	No binding

[†] Residue numbering based on crystallographic structure (PDB: 3OED)[14]
[−] Square brackets indicate location of cyclization through disulfide bridge after G12 is mutated to C
[x] Amino acids highlighted in green indicate that they are newly introduced as a result of optimization.

CR2-based Peptide 2

The proposed peptide 2 sequence was extracted from CR2 SCR1 domain residues 24-45 and the iterative design and optimization of the peptide is outlined in Table 5.2. Peptide 2.1 introduces two minor mutations to the native sequence of peptide 2: mutation of G24 to C to cyclize the peptide through a disulfidge bridge with C45 and mutation of C31 to A to prevent the formation of unintended disulfide bridges. Experimental testing of peptide 2.1 through MST demonstrated binding, albeit weak. Peptide 2.2 iterated on the previous peptide by introducing two conservative mutations to introduce more polarity to the peptide: F35 to H (retaining the aromatic characteristic) and I38 to T. Peptide 2.2 demonstrated improved binding with a K_D of 269.8 µM (as seen in Figure 5.2) but could be improved further. As these iterations of peptide 2 are fairly long (22 amino acids), peptide 2.2_1 truncated the sequence to exclude amino acids with minor or no contributions to binding observed in prior studies. However the truncated peptide 2.2_1 did not demonstrate

binding. At this juncture, combinatorial optimization through the Rosetta application Sequence Tolerance was explored. Initially, a divergent approach targeting optimization of amino acids exhibiting intermolecular contacts and amino acids demonstrating a lack of intermolecular contacts was utilized. Peptide 2.3 is the result of combinatorial optimization of peptide 2.1 amino acids with a lack of intermolecular contacts while peptide 2.4 is the result of combinatorial optimization of peptide 2.1 amino acids demonstrating intermolecular contacts. Peptide 2.5 takes a hybrid approach by optimizing amino acids exhibiting intermolecular contacts using the peptide 2.3 sequence thus combining optimization of both intermolecular non-contacts and contacts. This approach resulted in a significant improvement in binding for peptide 2.5 with a K_D of 87.9 µM as seen in Figure 5.3. Peptide 2.6 continued the pattern by attempting to optimize peptide 2.5 amino acids exhibiting a lack of intermolecular contacts but this approach did not retain the binding capabilities of peptide 2.5.

Table 5.2: CR2_pep2 optimization scheme

Peptide	Residue Range[†]	Sequence	Optimization	$K_D \pm SE$ (µM)
2.1	24-45	Ac-[CTVIRYSASGTFRLIGEKSLLC]-NH₂	Rational	≥615.4±50
2.2	24-45	Ac-[CTVIRYSASGTHRLTGEKSLLC]-NH₂	Rational	269.8±22.3
2.2_1	24-32, 39-45	Ac-[CTVIRYSPG GEKSLLC]-NH₂	Rational	No binding
2.3	24-45	Ac-[CTVIRYSASGTQRVNGEKSQLC]-NH₂	Combinatorial	Untested
2.4	24-45	Ac-[CMAKAYSASYTFRLIGKKNLTC]-NH₂	Combinatorial	Untested
2.5	24-45	Ac-[CINRAYSASYTQRVNGKKHQSC]-NH₂	Combinatorial	87.9±10
2.6	24-45	Ac-[CINRAYSASGTQRANGEKHQSC]-NH₂	Rational	No binding

[†] Residue numbering based on crystallographic structure (PDB: 3OED) [14]
[⁻] Square brackets indicate location of cyclization through disulfide bridge after G24 is mutated to C and C31 is mutated to A to prevent interference with the disulfide bridge
[x] Amino acids highlighted in green indicate that they are newly introduced as a result of optimization.
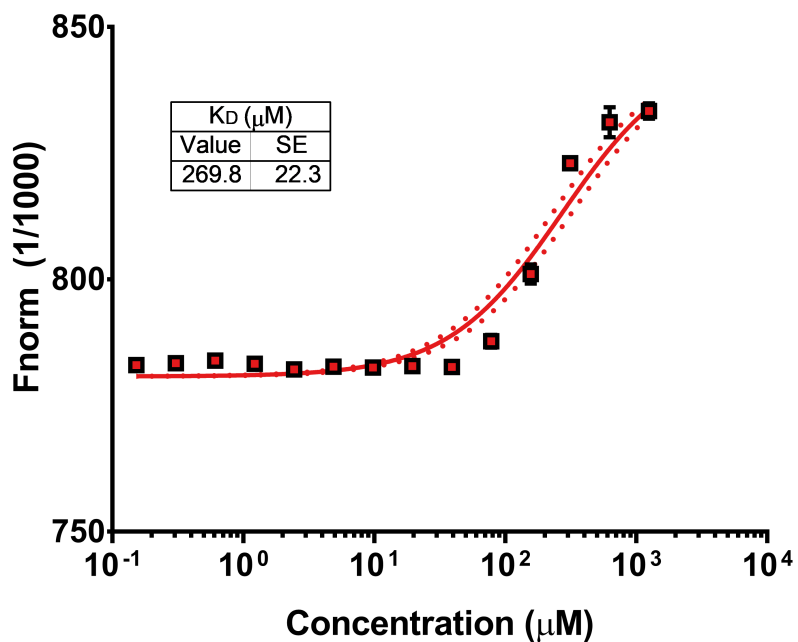
Figure 5.2: Concentration-dependent binding curve of peptide 2.2 to C3d. Thermophoretic data is plotted as mean ± standard error (as error bars) from three replicate experiments, together with the fitted binding curve in red and the 95% confidence interval of the fitted binding curve represented as red dots.

CR2-based Peptide 3

The proposed peptide 3 sequence from CR2 SCR2 residues 81-95 was chosen to include amino acids with significant intermolecular contacts and the various optimizations can be found in Table 5.3. Peptides 3.1, 3.2_1, and 3.2_2 implement side-chain to backbone, side-chain to side-chain, and head-to-tail cyclization respectively with peptides 3.2_1 and 3.2_2 containing the lone mutation to the parent sequence: T95 to E to facilitate the formation of a peptide bond. Peptides 3.2_1 and 3.2_2 were tested experimentally and did not exhibit any binding possibly due to the cyclization schemes disrupting the overall structure and native contacts of the CR2 β-sheet. This prompted a return to structural analysis of peptide 3 and peptide 3.3 was designed to extend the sequence to CR2 residues 79-98,
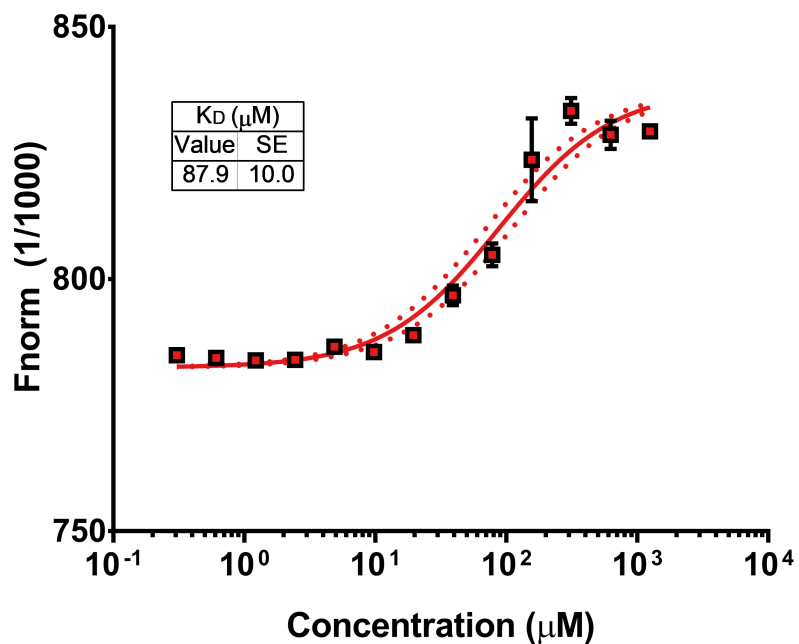
Figure 5.3: Concentration-dependent binding curve of peptide 2.5 to C3d. Thermophoretic data is plotted as mean ± standard error (as error bars) from three replicate experiments, together with the fitted binding curve in red and the 95% confidence interval of the fitted binding curve represented as red dots.

allowing for cyclization through a disulfide bridge after mutation of G79 to C. Peptide 3.3 demonstrated binding with a K_D of 228.8 µM (as seen in Figure 5.4), which is close to that of peptide 2.2. Just as with peptide 2.2, a divergent combinatorial approach targeting optimization of peptide 3.3 amino acids exhibiting intermolecular contacts and amino acids demonstrating a lack of intermolecular contacts was applied. Combinatorial optimization of peptide 3.3 amino acids with a lack of intermolecular contacts led to peptide 3.4 while combinatorial optimization of peptide 3.3 amino acids demonstrating intermolecular contacts led to peptide 3.5. Peptide 3.6 optimizes amino acids exhibiting intermolecular contacts using the peptide 3.4 sequence to incorporate optimization of both contacts and non-contacts but when had solubility issues when tested experimentally. Although the optimization steps

between peptide 2.2 to 2.5 are the same as that of 3.3 to 3.6, the physicochemical properties of the resulting sequences are much different from the input sequence in the latter optimization. Net charge of peptide 2.2 is approximately +2 with the optimized peptide 2.5 having an approximate net charge of +4, a change of 2. Peptide 3.3 however has an approximate net charge of 2 while the optimized peptide 3.6 has an approximate net charge of -4, an overall change of 6. Peptides 3.7 and 3.8 returned to optimization of the peptide 3.3 sequence by focusing on amino acids with few or no intermolecular contacts. Peptide 3.9 iterated on the sequences of 3.7 and 3.8 through rational design and although this resolved solubility issues, there was no binding observed.

Table 5.3: CR2_pep3 optimization scheme

Peptide	Residue Range[†]	Sequence	Optimization	$K_D \pm SE$ (μM)
3.1	81-95	Ac-|KIRGSTPYRHGDSVT|-NH$_2$	Rational	Untested
3.2_1	81-95	Ac-(KIRGSTPYRHGDSVE)-NH$_2$	Rational	No binding
3.2_2	81-95	Ac-{KIRGSTPYRHGDSVE}-NH$_2$	Rational	No binding
3.3	79-98	Ac-[CYKIRGSTPYRHGDSVTFAC]-NH$_2$	Rational	228.8±11.4
3.4	79-98	Ac-[CYTIRGSTPYRDGDTATYSC]-NH$_2$	Combinatorial	Untested
3.5	79-98	Ac-[CYKTDGITEYDHGDSVEFAC]-NH$_2$	Combinatorial	Untested
3.6	79-98	Ac-[CYKTDGQTEYTHGDSVYFAC]-NH$_2$	Combinatorial	Insoluble
3.7	79-98	Ac-[CVKIRGSTPMRHGDSEEYAC]-NH$_2$	Combinatorial	Untested
3.8	79-98	Ac-[CVKIRGSTPMRHGDSEEYSC]-NH$_2$	Combinatorial	Untested
3.9	79-98	Ac-[CTKIRGSTPYRHGDSATWAC]-NH$_2$	Rational	No binding

[†] Residue numbering based on crystallographic structure (PDB: 3OED)[14]
|–| Vertical bars indicate location of cyclization between side chain of K81 and backbone of T95
(–) Parenthesis indicate location of cyclization through sidechains
{–} Curly braces indicate location of head to tail cyclization
[–] Square brackets indicate location of cyclization through disulfide bridge after G79 is mutated to C
[x] Amino acids highlighted in green indicate that they are newly introduced as a result of optimization.
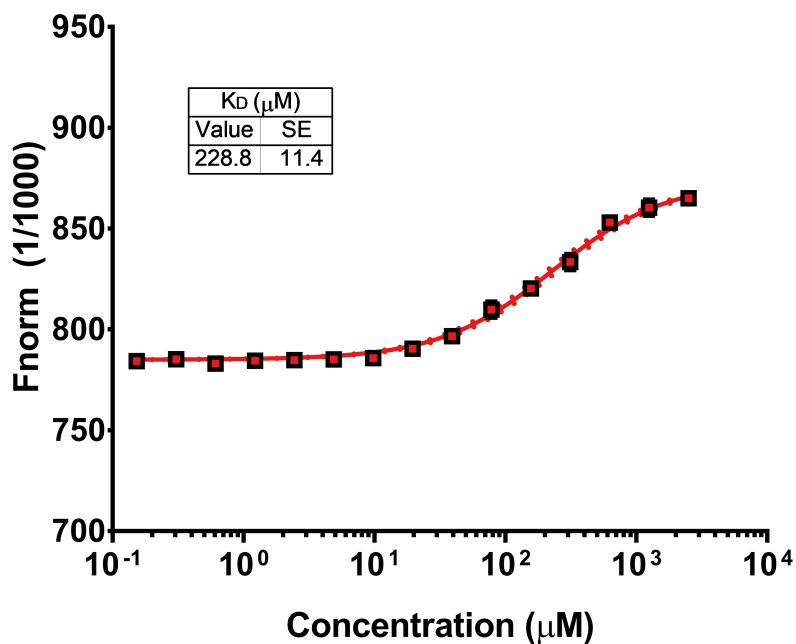
Figure 5.4: Concentration-dependent binding curve of peptide 3.3 to C3d. Thermophoretic data is plotted as mean $\pm$ standard error (as error bars) from three replicate experiments, together with the fitted binding curve in red and the 95% confidence interval of the fitted binding curve represented as red dots.

5.3.2 Optimization Limits and Future Directions

We have only evaluated peptide 1 through rational optimization and it would be worthwhile to incorporate the combinatorial approaches to see if that can help resolve some of the solubility issues, perhaps by only allowing mutations to polar amino acids. One potential cause of the lack of combinatorial optimization success with peptide 3.3 could be due to the usage of the crystallographic structure of the C3d:CR2 complex[14], in which the SCR2 domain (location of peptide 3) is further away from C3d than the SCR1 domain where (location of peptide 1 and 2). As the sequence tolerance application in Rosetta is influenced by the input structure, it's possible that this may have played a role in influencing the optimization algorithm and biased it towards unnecessary mutations. A future direction

104

to explore would be to extract representative structures of the C3d:CR2 complex from the MD trajectory or even just a few structures where the SCR2 domain is closer in proximity to C3d and use that as the basis for combinatorial optimization through the sequence tolerance application in Rosetta. This may help in further optimization of peptide 2.5 as well. Additionally, further parametrization of the genetic algorithm, the fitness weights in particular, may aid in the identification of mutants that optimize the C3d:peptide interface more and increase binding affinity.

5.4   Conclusion

Implementing rational and combinatorial design on foundational knowledge of the mechanistic properties of intermolecular interactions driving the stability of the C3d:CR2 complex allowed us to identify novel CR2-based peptides with C3d-binding capabilities. We explored both linear and cyclization schemes as well as optimizing for different physicochemical properties in mutagenesis to arrive at the eventual design of peptide 2.5 and 3.3. Although there is potential to improve the binding affinity of these peptides through further iterative design by revisiting the workflow of combinatorial design, these peptides may still amenable for potential biomarker applications in their current iteration. As the objective is a C3d-binding peptide primarily for theranostic biomarker purposes, a weak binder isn't necessarily suboptimal as quick clearance is ideal. There is a distinct lack of complement-targeted therapeutics or theranostics and those currently on the market or in development are primarily biopharmaceuticals[20,21] which suffer from prohibitive production costs, scalability issues, and delivery difficulty. Peptides are an amenable alternative due

to their balance in ease of production, specificity, and multiple delivery options (depending on the peptide). We provide two peptides which, subsequent to evaluation of stability post-fluorophore attachment, can serve as the basis for further development into eventual clinical applications for the diagnosis of AMD.
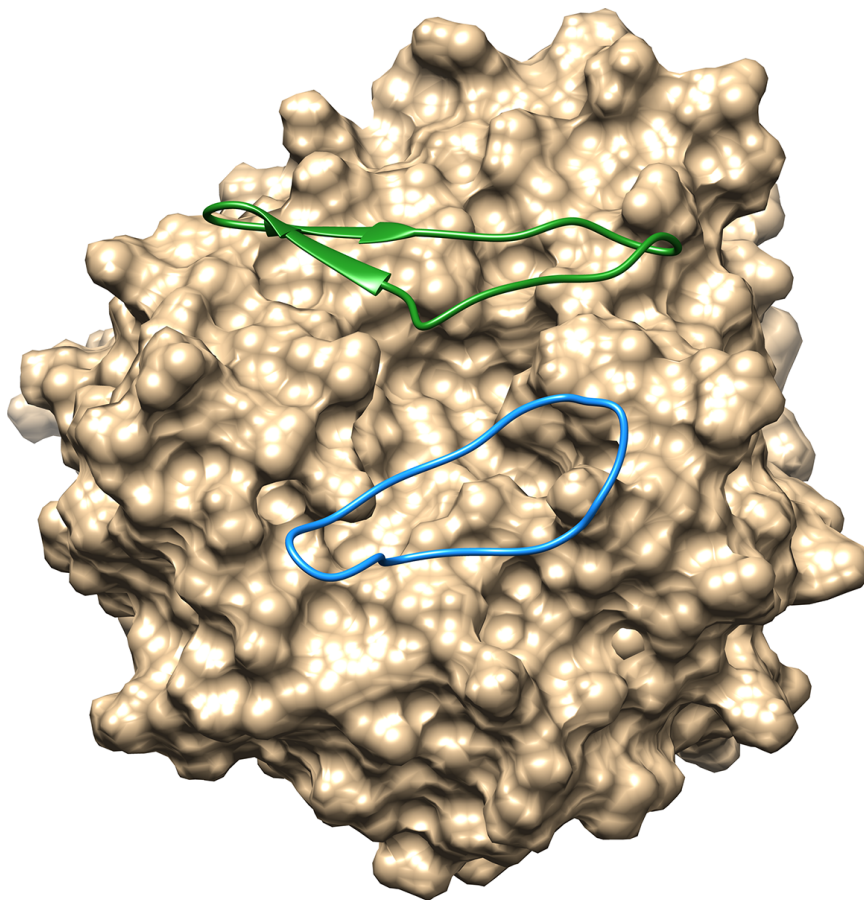
Figure 5.5: Molecular graphics showing peptide 2.5 (in green) and peptide 3.3 (in blue) at their CR2-based positions in relation to C3d. THe structures used for molecular graphics are outputs from Rosetta Sequence Tolerance [18,19]

5.5   References

[1] Liszewski, M. K. and Atkinson, J. P. Complement regulators in human disease: Lessons from modern genetics. *Journal of Internal Medicine*, 277(3):294–305, March 2015. ISSN 1365-2796. doi: 10.1111/joim.12338.

[2] Schramm, E. C., Clark, S. J., Triebwasser, M. P., Raychaudhuri, S., Seddon, J. M., and Atkinson, J. P. Genetic variants in the complement system predisposing to age-related macular degeneration: A review. *Molecular Immunology*, 61(2):118–125, October 2014. ISSN 01615890. doi: 10.1016/j.molimm.2014.06.032. 00029.

[3] Anderson, D. H., Radeke, M. J., Gallo, N. B., Chapin, E. A., Johnson, P. T., Curletti, C. R., Hancox, L. S., Hu, J., Ebright, J. N., Malek, G., Hauser, M. A., Rickman, C. B., Bok, D., Hageman, G. S., and Johnson, L. V. The Pivotal Role of the Complement System in Aging and Age-related Macular Degeneration: Hypothesis Re-visited. *Prog Retin Eye Res*, 29(2):95–112, March 2010. ISSN 1350-9462. doi: 10.1016/j.preteyeres. 2009.11.003.

[4] Anderson, D. H., Mullins, R. F., Hageman, G. S., and Johnson, L. V. A role for local inflammation in the formation of drusen in the aging eye. *American Journal of Ophthalmology*, 134(3):411–431, September 2002. ISSN 0002-9394. doi: 10.1016/S0002-9394(02)01624-0.

[5] Mohan, R. R., Gorham Jr., R. D., and Morikis, D. A theoretical view of the C3d:CR2 binding controversy. *Molecular Immunology*, 64(1):112–122, March 2015. ISSN 0161-5890. doi: 10.1016/j.molimm.2014.11.006. 00002.

[6] Mohan, R. R., Huber, G. A., and Morikis, D. Electrostatic Steering Accelerates C3d:CR2 Association. *J. Phys. Chem. B*, April 2016. ISSN 1520-6106. doi: 10.1021/acs.jpcb.6b02095. 00000.

[7] Harrison, R. E. S., Mohan, R. R., Gorham, R. D., Kieslich, C. A., and Morikis, D. AESOP: A Python Library for Investigating Electrostatics in Protein Interactions. *Biophysical Journal*, 112(9):1761–1766, May 2017. ISSN 0006-3495. doi: 10.1016/j. bpj.2017.04.005.

[8] Isenman, D. E., Leung, E., Mackay, J. D., Bagby, S., and van den Elsen, J. M. H. 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. *The Journal of Immunology*, 184(4):1946–1955, January 2010. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.0902919.

[9] Clemenza, L. and Isenman, D. E. Structure-guided identification of C3d residues essential for its binding to complement receptor 2 (CD21). *The Journal of Immunology*, 165(7):3839–3848, 2000.

[10] Hannan, J. P., Young, K. A., Guthridge, J. M., Asokan, R., Szakonyi, G., Chen, X. S., and Holers, V. M. Mutational Analysis of the Complement Receptor Type 2 (CR2/CD21)–C3d Interaction Reveals a Putative Charged SCR1 Binding Site for C3d. *Journal of Molecular Biology*, 346(3):845–858, February 2005. ISSN 0022-2836. doi: 10.1016/j.jmb.2004.12.007.

[11] Molina, H., Perkins, S. J., Guthridge, J., Gorka, J., Kinoshita, T., and Holers, V. M. Characterization of a complement receptor 2 (CR2, CD21) ligand binding site for C3. An initial model of ligand interaction with two linked short consensus repeat modules. *The Journal of Immunology*, 154(10):5426–5435, May 1995. ISSN 0022-1767, 1550-6606.

[12] Guthridge, J. M., Young, K., Gipson, M. G., Sarrias, M.-R., Szakonyi, G., Chen, X. S., Malaspina, A., Donoghue, E., James, J. A., Lambris, J. D., and others. Epitope mapping using the X-ray crystallographic structure of complement receptor type 2 (CR2)/CD21: Identification of a highly inhibitory monoclonal antibody that directly recognizes the CR2-C3d interface. *The Journal of Immunology*, 167(10):5758–5766, 2001.

[13] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. UCSF Chimera?A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, October 2004. ISSN 0192-8651, 1096-987X. doi: 10.1002/jcc.20084.

[14] van den Elsen, J. M. H. and Isenman, D. E. A Crystal Structure of the Complex Between Human Complement Receptor 2 and Its Ligand C3d. *Science*, 332(6029): 608–611, April 2011. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1201954.

[15] Šali, A. and Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology*, 234(3):779–815, December 1993. ISSN 0022-2836. doi: 10.1006/jmbi.1993.1626.

[16] Henikoff, S. and Henikoff, J. G. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A*, 89(22):10915–10919, November 1992. ISSN 0027-8424.

[17] Hutchinson, E. G. and Thornton, J. M. A revised set of potentials for beta-turn formation in proteins. *Protein Sci*, 3(12):2207–2216, December 1994. ISSN 0961-8368.

[18] Smith, C. A. and Kortemme, T. Structure-Based Prediction of the Peptide Sequence Space Recognized by Natural and Synthetic PDZ Domains. *Journal of Molecular Biology*, 402(2):460–474, September 2010. ISSN 0022-2836. doi: 10.1016/j.jmb.2010.07.032.

[19] Smith, C. A. and Kortemme, T. Predicting the Tolerated Sequences for Proteins and Protein Interfaces Using RosettaBackrub Flexible Backbone Design. *PLOS ONE*, 6 (7):e20451, July 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0020451.

[20] Thurman, J. M., Kulik, L., Orth, H., Wong, M., Renner, B., Sargsyan, S. A., Mitchell, L. M., Hourcade, D. E., Hannan, J. P., Kovacs, J. M., Coughlin, B., Woodell, A. S., Pickering, M. C., Rohrer, B., and Holers, V. M. Detection of complement activation using monoclonal antibodies against C3d. *Journal of Clinical Investigation*, 123(5): 2218–2230, May 2013. ISSN 0021-9738. doi: 10.1172/JCI65861. 00022.

[21] Ricklin, D. and Lambris, J. D. Complement in Immune and Inflammatory Disorders: Therapeutic Interventions. *J Immunol*, 190(8):3839–3847, April 2013. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.1203200. 00094.

Chapter 6

Dynamics of C3d:CR3 interaction

6.1 Introduction

The TED domain of C3b and iC3b is the opsonization site that acts through covalent attachment to the activating surface[1–3]. The standalone TED domain fragment, known as C3d, remains on surfaces and is a biomarker of complement activation. Standalone C3d or as the TED domain of C3b and iC3b, called hereafter indiscriminately C3d, participates in a variety of complement processes such as enhanced antibody production, phagocytosis, and regulation through interactions with complement proteins compelement receptor 2 (CR2)[4,5], complement receptor 3 (CR3)[6,7], and Factor H (FH)[8] respectively. C3d's interaction with CR3 is of particular importance due to its role in phagocytosis in which CR3-presenting macrophages transport C3d-opsonized complexes to B cells for augmented antibody production through CR2[4,9].

The interaction between C3d and CR3 is mediated by a Mg^{2+} ion at the interface through the metal ion-dependent adhesion site (MIDAS) of CR3[10]. The prevalence
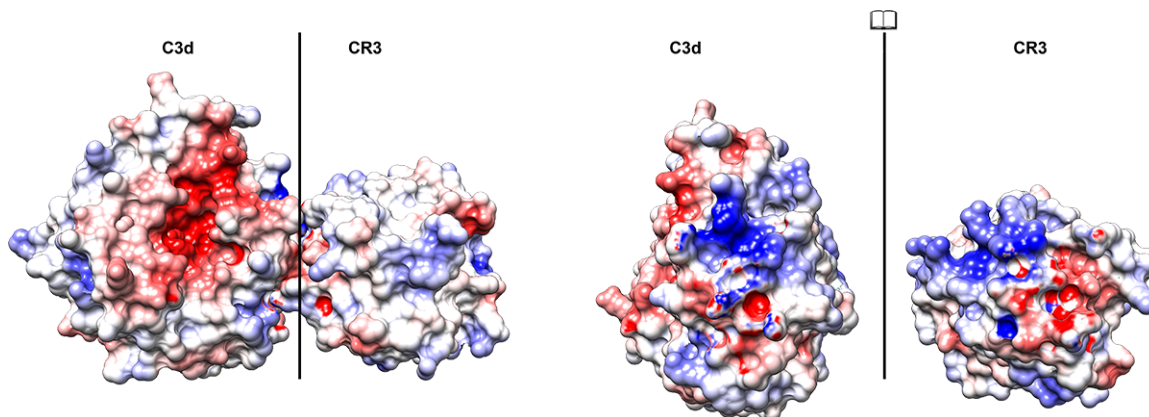
Figure 6.1: Electrostatic potentials mapped onto the protein surfaces of C3d and CR3 in closed book (left) and open book (right) representation. Electrostatic potentials were calculated at ionic strength corresponding to 150 mM monovalent counterion concentration. The color transitions from red to white to blue represent electrostatic potential values of $-5kT/e$ to $0kT/e$ to $5kT/e$.

of electrostatics in the complement system[11–19] and the charged interface (Figure 6.1) provide an opportunity to characterize the role of electrostatics at the C3d:CR3 interface. We performed detailed characterization of the role of electrostatics at the C3d:CR3 interface, using the crystal structure of the C3d:CR3 complex[10]. This study adds to our knowledge database on the prevalent role of electrostatics on the function and regulation of complement system. In this study, we incorporate electrostatic analysis and molecular dynamics simulations (both explicit-solvent and steered) to characterize the amino acids and dynamics driving the C3d:CR3 interaction as well as the role of the Mg^{2+} ion at the interface. The presence of the metal ion at the interface makes the complex particularly amenable to study through steered molecular dynamics simulations[20]. Considering the role of CR3 in antibody-based cancer therapies[21], inflammatory and anti-inflammatory response[22], and bacterial evasion of complement response[23], it is crucial to elucidate the mechanisms of the C3d:CR3 interaction and we provide insights into these mechanisms in this study.

6.2 Methods

6.2.1 Structure preparation and analysis

The crystallographic structure of the CR3 I domain in complex with C3d was utilized for analysis (PDB: 4M76)[10]. Ni^{2+} ions were used in the MIDAS site of the original crystallographic structure due to incompatibility of Mg^{2+} with the electron density in the structure. The Ni^{2+} ion directly at the complex interface in the crystallographic structure was replaced by Mg^{2+} in our study to more accurately reflect physiological conditions[10] while the other Ni^{2+} ion was removed and excluded from analysis. In order to comparatively evaluate the presence or absence of the Mg^{2+} ion in the interaction, a second version of the C3d:CR3 complex was used where the Mg^{2+} ion was removed from the structure. These structures will be referred to as C3d:CR3^{Mg2+} and C3d:CR3$^{no\ Mg2+}$ from here on. Crystallographic waters were retained in the structures. Structural analysis and visualization of the structures were performed in UCSF Chimera[24].

6.2.2 Electrostatic analysis

Electrostatic analysis was performed using the AESOP (Analysis of Electrostatic Structures Of Proteins) python package[25] similar to previous studies of complement system protein-protein interactions[15–18]. The Alanine scan class in AESOP was utilized to perform a computational alanine scan to evaluate the electrostatic contributions of C3d:CR3 amino acids and to evaluate the role of the Mg^{2+} ion at the interface.

In the AESOP workflow, initially PDB2PQR[26] is utilized to assign charges and atomic radii for each atom according to the CHARMM27 force field[27] and to convert the

PDB file format to PQR. Next, single alanine mutants of all ionizable amino acids of the C3d:CR3 complex are generated. Subsequently the Adaptive Poisson-Boltzmann Solver (APBS)[28] is used to calculate grid-based electrostatic potentials which are used in the calculation of electrostatic free energies of association.

Electrostatic free energies of association (ΔG_{assoc}) are calculated according to a thermodynamic cycle to[25] and through Coulombic potentials generated by the APBS Coulomb module to account for Coulombic and solvation effects and to avoid grid artifacts, as previously described[15,17,18,25,29,30] and shown in Eq.6.2

$$\Delta\Delta G_{solv} = \Delta G_{solv}^{C3d:CR3} - \Delta G_{solv}^{C3d} - \Delta G_{solv}^{CR3} \qquad (6.1)$$

$$\Delta G_{assoc} = \Delta G_{coul} + \Delta\Delta G_{solv} \qquad (6.2)$$

Electrostatic free energies of association of each mutant are described with respect to that of the parent protein as seen in Eq. 6.3

$$\Delta G_{binding} = \Delta G_{assoc}^{mutant} - \Delta G_{assoc}^{parent} \qquad (6.3)$$

Dielectric coefficients of 20 and 78.54 were used for the protein and solvent respectively (as previously described[29,30]) in the APBS calculations. Ionic strengths of 150 mM and 0 mM were utilized for the solvated and reference state respectively in the thermodynamic cycle. The number of grid points and mesh dimensions were set to $129 \times 129 \times 129$ and $117\,\text{Å} \times 119\,\text{Å} \times 99\,\text{Å}$ respectively. The alanine scan was carried out on both the C3d:CR3$^{\text{Mg2+}}$ and C3d:CR3$^{\text{no Mg2+}}$ structures.

6.2.3 Explicit-solvent molecular dynamics (MD) simulations

Explicit-solvent molecular dynamics (MD) simulations were carried out using both the C3d:CR3^{Mg2+} and C3d:CR3$^{no\ Mg2+}$ structures. Initial minimization of the structure in the absence of water was carried out using NAMD[31] and the CHARMM36 forcefield[27]. The structures were then solvated in TIP3P water boxes with dimensions of 97 Å × 97 Å × 86 Å and neutralized with sodium and chloride counterions at an ionic strength of 150 mM. Subsequent to addition of solvent, the structures underwent 25,000 steps of conjugate gradient energy minimization followed by heating from 0 to 300 K in 62 ps with all protein atoms constrained to their post-minimization positions. Following heating, the two systems were equilibrated through five stages for 1 ns/stage with force constants 41.84, 20.92, 8.368 and 4.184 kJ/mol/Å^2 applied during the first four stages respectively to harmonically constrain all protein atoms to their post-minimization positions. During the final stage of equilibration a force constant of 4.184 kJ/mol/Å^2 was applied to constrain only protein backbone atoms to their post-minimization positions. Following equilibration, production runs were carried out using AMBER16[32] with: periodic boundary conditions, Langevin dynamics, nonbonded interaction cutoff of 12 Å, SHAKE algorithm, and an integration timestep of 2 fs.

6.2.4 Steered MD simulations

Steered molecular dynamics simulations (SMD) were carried out for both C3d:CR3^{Mg2+} and C3d:CR3$^{no\ Mg2+}$ in NAMD using representative structures extracted from the simulation trajectories. Initially, UCSF Chimera[24] was used to calculate the

normal vectors of the solvent accessible surface of C3d atoms within 8 Å of CR3 in each structure. Subsequently the mean of the normal vectors was calculated and each complex rotated so that the mean normal vector was aligned with the $+z$ axis, allowing the $+z$ axis to be used as the direction of induced dissociation of CR3 from C3d.

The structures were solvated in TIP3P water boxes with boundaries from 12 Å from the protein in both the x and y directions and in. In order to allow for sufficient room for induced dissociation in the $+z$ direction, the boundary in the $+z$ direction is 70 Å from the protein. Subsequent to solvation, the system was minimized and heated in the same manner as the explicit-solvent MD simulations as discussed above. Following heating, the system equilibrated for 125 ps while all proteins atoms were harmonically constrained with a force constant of $41.84 \, \text{kJ/mol/Å}^2$. The SMD simulations were carried out using the equilibrated structures as input. During the SMD simulation, C3d residues >12 Å from CR3 were harmonically constrained to their post-equilibration positions with a force constant of $41.84 \, \text{kJ/mol/Å}^2$, while the center of mass of all CR3 atoms was constrained and pulled in the $+z$ direction at a constant velocity of 10 Å/ns.

Nine simulations were performed for each system for a total of 18 trajectories.

6.2.5   MD simulation analysis

RMSD analysis and visualization of the simulation trajectories was carried out using MDTraj[33] and MSMExplorer[34]. Analysis and visualization of intermolecular hydrogen bonds and salt bridges (cutoff of 5 Å) was performed using MDTraj, pandas, matplotlib and seaborn[33,35–37]. PCA decomposition was performed on the phi and psi angles observed throughout the trajectory using MSMBuilder[38]. The MiniBatchKMeans method

in MSMBuilder was utilized to cluster the principal components to nine distinct clusters and cluster centers were extracted as representative structures for both C3d:CR3^{Mg2+} and C3d:CR3$^{no\ Mg2+}$. Representative structures for C3d:CR3$^{no\ Mg2+}$ trajectories were chosen so as to avoid structures where CR3 has dissociated from C3d. Visualization of clustering and decomposition was performed using MSMExplorer.

6.3   Results and Discussion

6.3.1   Role of electrostatics in C3d:CR3 interaction

In the following analysis, the $\Delta G_{binding}$ values are represented as described in Eq. 6.3. A positive $\Delta G_{binding}$ value corresponds to loss of binding due to mutation of the corresponding residue to alanine while a negative $\Delta G_{binding}$ value corresponds to a gain of binding due to mutation of the corresponding residue alanine.

Figure 6.2 shows the results of AESOP analysis of the C3d:CR3^{Mg2+} structure. Only mutations resulting in a $\Delta G_{binding}$ value outside the thermal fluctuation range ($>2.5\,\mathrm{kJ/mol}$ or $<-2.5\,\mathrm{kJ/mol}$) are displayed. Only three C3d residues and four CR3 residues demonstrated $\Delta G_{binding}$ values outside this range and all of these residues are at the C3d:CR3 interface. All of the identified mutations were loss of binding mutations suggesting the importance of their electrostatic contributions to binding. C3d mutant D1247A in particular shows a relatively high $\Delta G_{binding}$ value of 23.6 kJ/mol. This is expected as the Mg^{2+} ion at the C3d:CR3 interface directly mediates the interaction between C3d and CR3 through C3d D1247[10]. Additionally D1247 is suggested to be essential to the interaction,
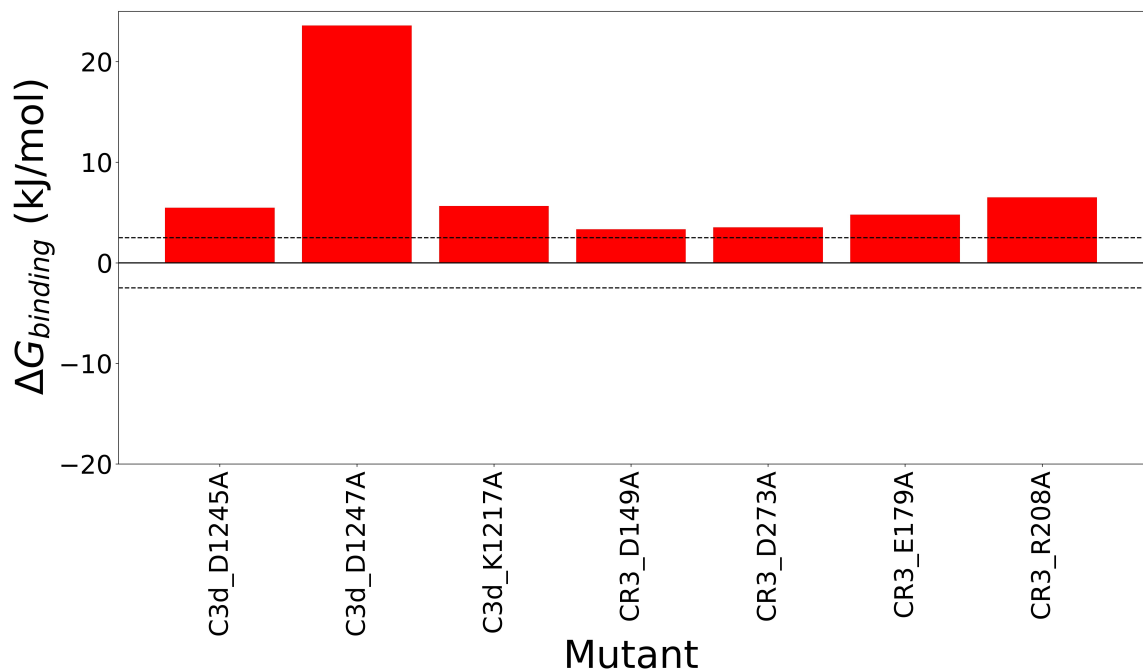
Figure 6.2: Results of computational alanine scan of C3d:CR3$^{\text{Mg2+}}$ with only mutants demonstrating $\Delta G_{binding}$ value outside the thermal fluctuation range ($>2.5\,\text{kJ/mol}$ or $<-2.5\,\text{kJ/mol}$) displayed. Changes in free energies of binding upon mutation are calculated using Eq. 6.3. A positive value (colored in red) denotes loss of binding, indicating that the mutated residue favors binding. A negative value (colored in blue) denotes gain of binding, indicating that the mutated residue opposes binding. The thermal fluctuation range is marked by dashed black lines.

along with K1217[10] which was also identified as a significant loss of binding mutation in Figure 6.2.

Figure 6.3 shows the results of AESOP analysis of the C3d:CR3$^{\text{no Mg2+}}$ structure and similar to above, only mutations resulting in a $\Delta G_{binding}$ value outside the thermal fluctuation range are displayed. In comparison to the analysis of C3d:CR3$^{\text{Mg2+}}$, here only two C3d residues demonstrated $\Delta G_{binding}$ values outside this range while seven CR3 residues were outside of this range. Both of the C3d residues in Figure 6.3 have overlap with Figure 6.2 albeit with different $\Delta G_{binding}$ values. C3d K1217A demonstrates a slightly

larger $\Delta G_{binding}$ value in comparison to C3d:CR3^{Mg2+} while still maintaining a loss of binding mutation. Interestingly, here C3d D1247A has a drastically different $\Delta G_{binding}$ value (-16.9 kJ/mol vs 23.6 kJ/mol) and results in a gain of binding mutation. It's possible that without the Mg^{2+} ion to mediate the interaction between C3d D1247, CR3 S142, CR3 S144, and CR3 T209 observed in the crystallographic structure the unfavorable like-charge interactions are much more amenable to optimization, resulting in D1247A as a gain of binding mutation here. CR3 E179A and CR3 R208A are the only mutants with overlap in Figure 6.2 and Figure 6.3 and both exhibit gain of binding mutations in Figure 6.3 as well. There are several more CR3 gain of binding mutations (D140A, D242A, E244A, E283A) at the C3d:CR3 interface in the absence of the Mg^{2+} ion. This is expected as these unfavorable interactions are reversed and become favorable upon Mg^{2+} binding.

Figure 6.4 shows the comparison of AESOP results of C3d:CR3^{Mg2+} and C3d:CR3$^{no\ Mg2+}$ with only mutants exhibiting a difference >2.5 kJ/mol in $\Delta G_{binding}$. There are four C3d mutants and six CR3 mutants identified in Figure 6.4. Three of the C3d mutants identified here (K1217A, R1247A, and R1254A) are considered to be crucial to to the steering of the C3d:CR3 interaction[10]. However, the role of the Mg^{2+} ion has the most drastic effect on R1247A. Despite the small interface, electrostatics still appear to be a major factor in the C3d:CR3 interface in part due to charged amino acids at the interface as seen in Figure 6.1.
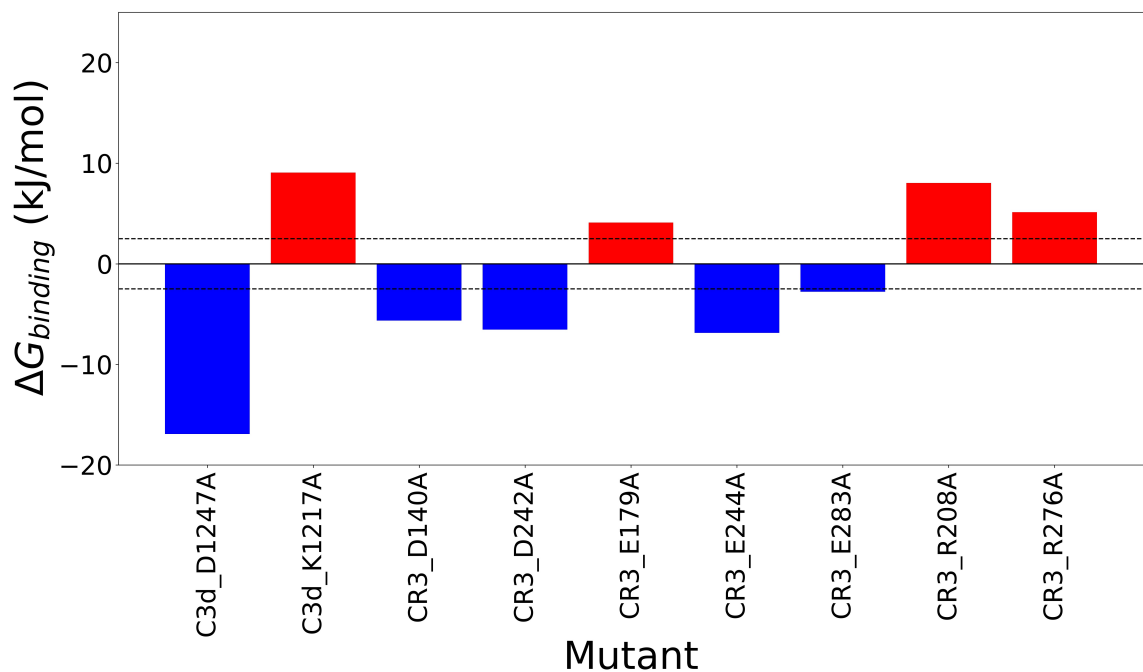
Figure 6.3: Results of computational alanine scan of C3d:CR3$^{\text{no Mg2+}}$ with only mutants demonstrating $\Delta G_{binding}$ value outside the thermal fluctuation range (>2.5 kJ/mol or <-2.5 kJ/mol) displayed. Changes in free energies of binding upon mutation are calculated using Eq. 6.3. A positive value (colored in red) denotes loss of binding, indicating that the mutated residue favors binding. A negative value (colored in blue) denotes gain of binding, indicating that the mutated residue opposes binding. The thermal fluctuation range is marked by dashed black lines.
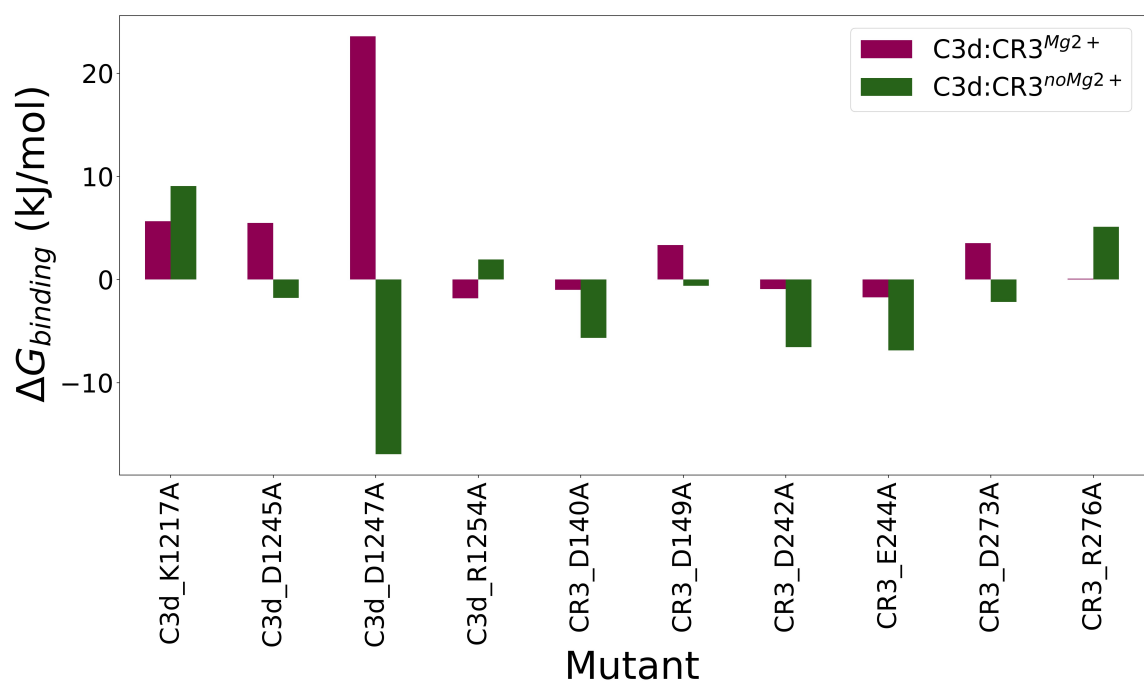
Figure 6.4: Comparison of the computational alanine scan results of C3d:CR3^{Mg2+} and C3d:CR3$^{no\ Mg2+}$ with only mutants exhibiting a difference $>2.5\,\text{kJ/mol}$ in $\Delta G_{binding}$ displayed. C3d:CR3^{Mg2+} $\Delta G_{binding}$ values are colored in purple while C3d:CR3$^{no\ Mg2+}$ $\Delta G_{binding}$ values are colored in green.
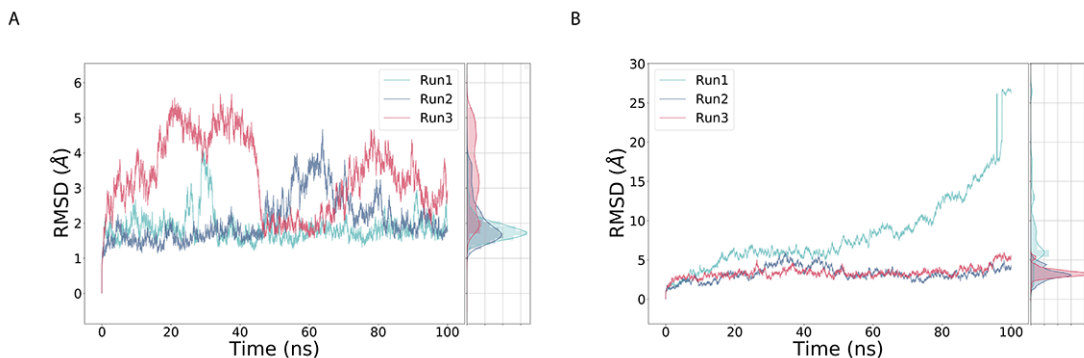
Figure 6.5: RMSD timeseries plots of (A) C3d:CR3$^{\text{Mg2+}}$ and (B) C3d:CR3$^{\text{no Mg2+}}$ simulation trajectories. Histograms of RMSD values are provided to the right of timeseries plots.

6.3.2 Stabilizing effect of Mg^{2+} on C3d:CR3 complex dynamics

Characterizing the dynamics of C3d:CR3$^{\text{Mg2+}}$ and C3d:CR3$^{\text{no Mg2+}}$ through MD simulations provides a clearer understanding of the importance of the Mg^{2+} ion to the interaction. One of the key findings was that in the absence of the Mg^{2+} ion, C3d and CR3 dissociated in at least one trajectory as the trajectory progressed whereas that was not the case when the Mg^{2+} ion was present (Figure 6.5).

Analysis of intermolecular salt bridges (Figure 6.6) and hydrogen bonds (Figure 6.7) shows that an overall weakening of interactions at the C3d:CR3 interface. In 6.6, the intermolecular salt bridge occupancy comparison between C3d:CR3$^{\text{Mg2+}}$ (Figure 6.6A) and C3d:CR3$^{\text{no Mg2+}}$ (Figure 6.6B) shows that there is an overall reduction in the number of salt bridges observed as well as in their occupancy when Mg^{2+} is removed. The salt bridge between C3d D1245 and CR3 R208 in particular experiences a drastic reduction in occupancy. The exceptions are the salt bridges between C3d D1247 and CR3 R208 and C3d K1217 and CR3 E179 which show slight increases in their occupancy. Similarly, the absence
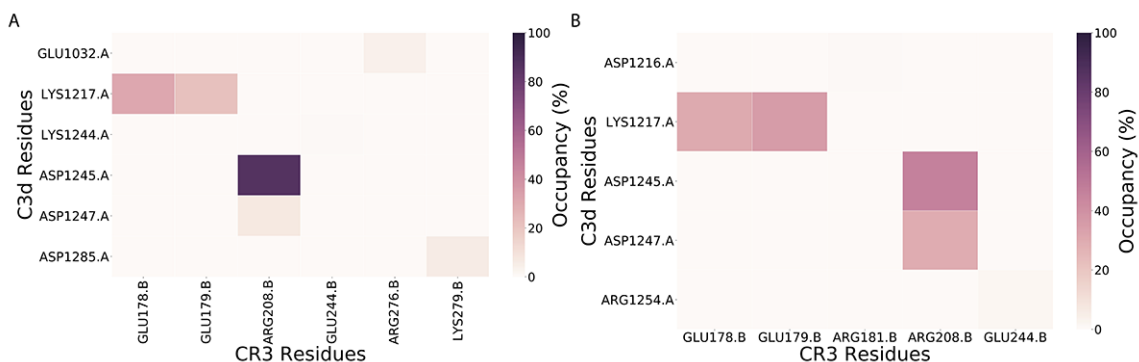
Figure 6.6: Intermolecular salt bridge occupancies for (A) C3d:CR3^{Mg2+} and (B) C3d:CR3$^{no\ Mg2+}$ trajectories.
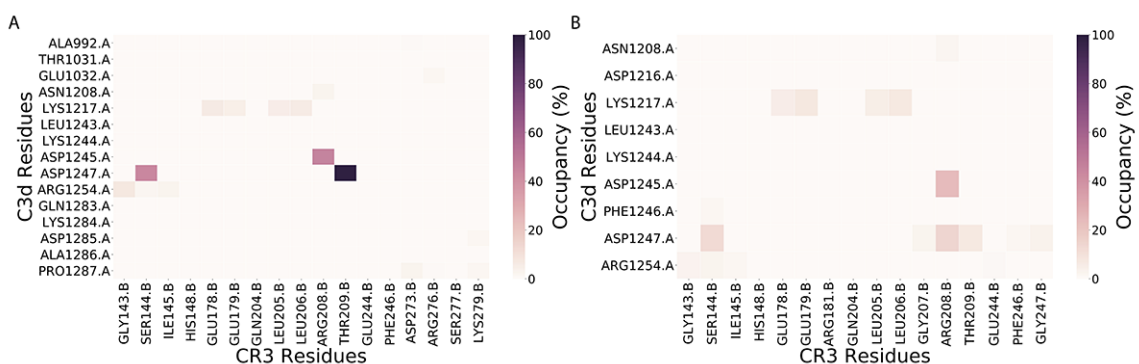
Figure 6.7: Intermolecular hydrogen bond occupancies for (A) C3d:CR3^{Mg2+} and (B) C3d:CR3$^{no\ Mg2+}$ trajectories.

of Mg^{2+} reduces the number of observed hydrogen bonds as well as their occupancy as seen in Figure 6.7. In particular, the hydrogen bonds that C3d D1247 forms with CR3 S144 and CR3 T209 are drastically reduced in occupancy. Considering the importance of D1247 to the C3d:CR3 interaction[10], this may be one of the causes for ation of the C3d:CR3 complex in the absence of Mg^{2+}

Timeseries analysis of C3d residues closest to the Mg^{2+} ion using C3d:CR3^{Mg2+} simuatlion trajectories show that D1247 and F1246 are the only residues with reasonable
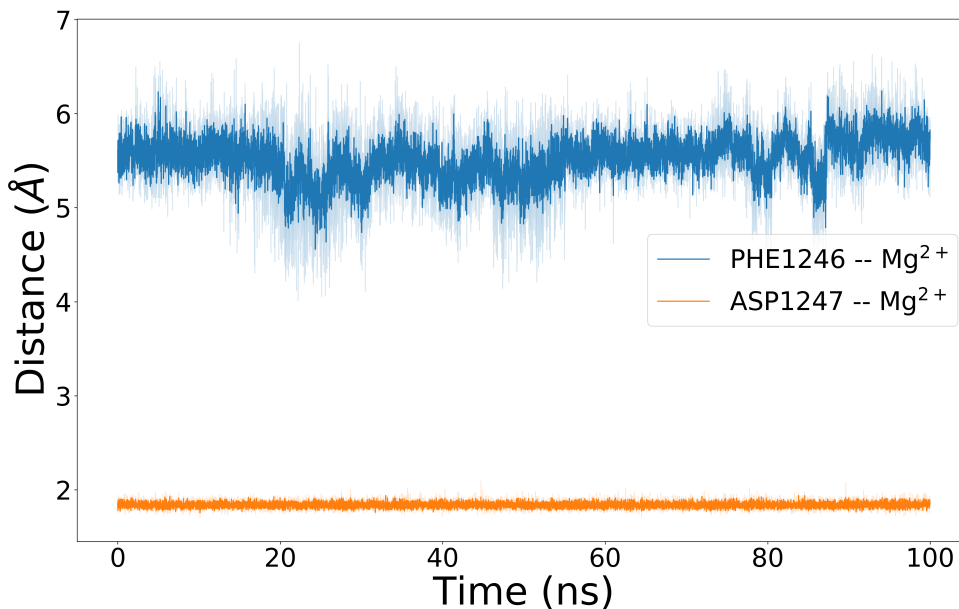
Figure 6.8: Timeseries analysis of C3d:CR3$^{\text{Mg2+}}$ trajectories to explore C3d residues observed within $5\,\text{Å}$ of Mg^{2+} at an occupancy $>30\%$. Distances are calculated based on the closest heavy atom of the residue and the Mg^{2+} ion.

occupancy $<5\,\text{Å}$ (Figure 6.8). D1247 in particular maintains close proximity to Mg^{2+} in all simulation trajectories. Analysis of CR3 residues in close proximity to Mg^{2+} using C3d:CR3$^{\text{Mg2+}}$ trajectories show several residues with high occupancy (Figure 6.9). CR3 S142, S144, and T209 show the highest occupancy at the lowest distance which corroborates their proposed role in coordinating the Mg^{2+} ion[10]. Other residues maintaining a close proximity to Mg^{2+} include CR3 D140, G143, I145, and R208 which are likely due to the coordination by S142, S144, and T209. D242 and E244 demonstrate high occupancy at a small distance as well despite not exhibiting any significant hydrogen bonds or salt bridges.
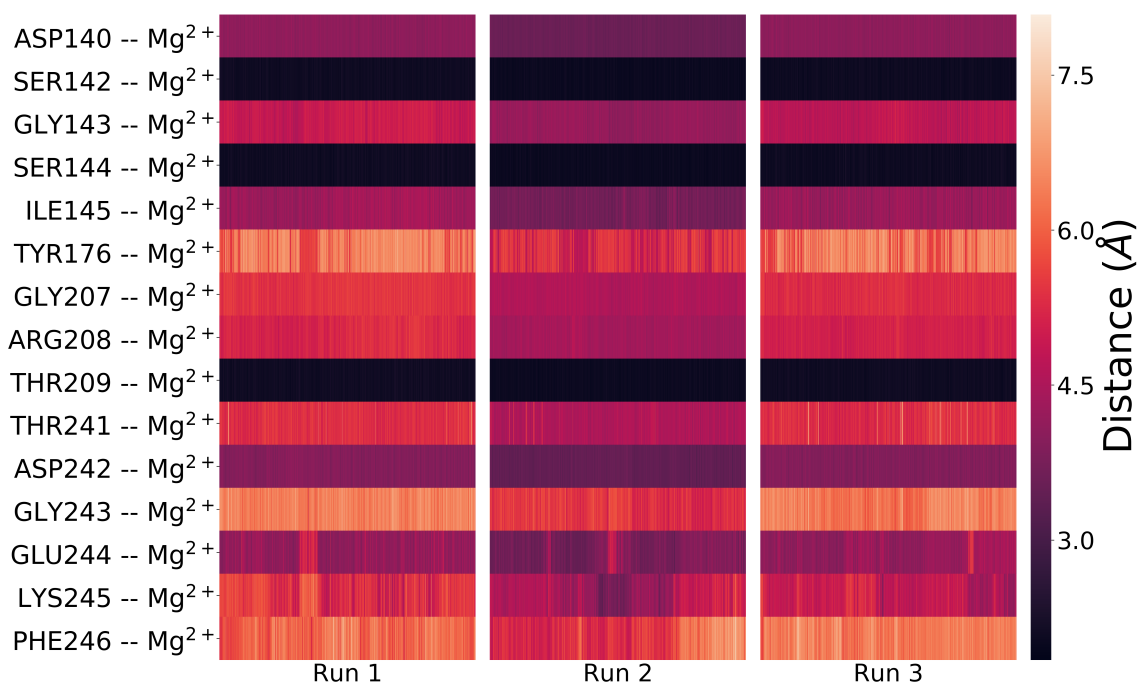
Figure 6.9: Timeseries heatmap analysis of C3d:CR3$^{\mathrm{Mg2+}}$ trajectories to characterize CR3 residues observed within 5 Å of Mg^{2+} at an occupancy >30%. Heatmaps for each simulaton trajectory are labeled Run 1, Run 2 and Run 3 respectively are plotted from left to right as a timeseries. Distances are calculated based on the closest heavy atom of the residue and the Mg^{2+} ion.
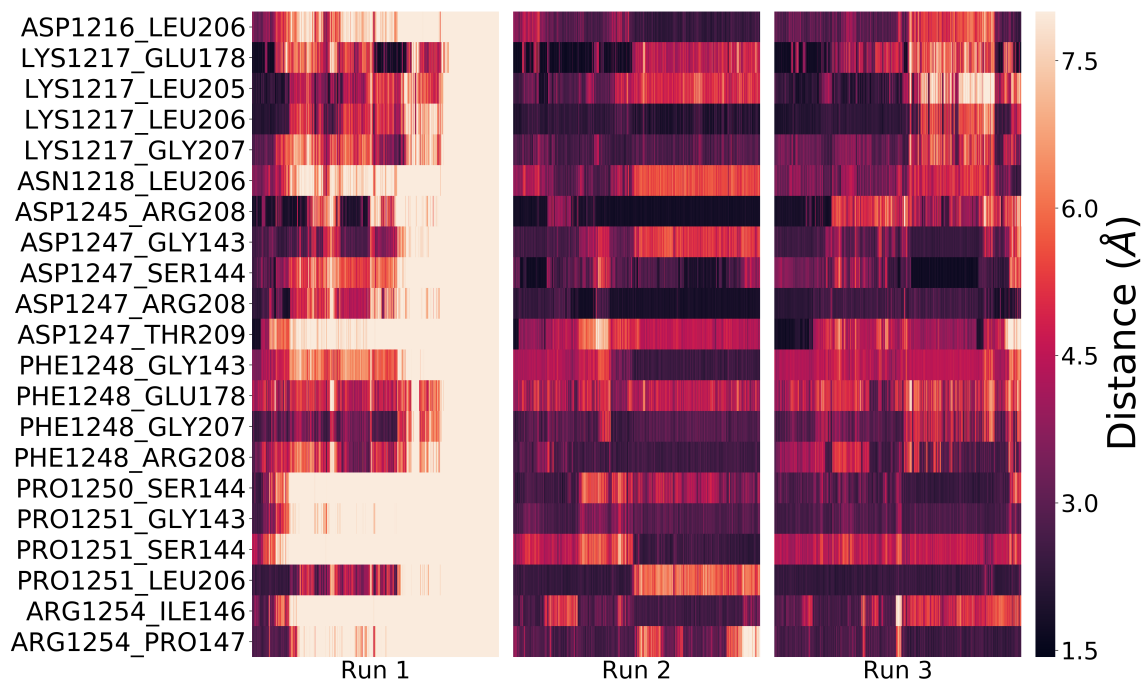
Figure 6.10: Timeseries heatmap analysis of C3d:CR3$^{\text{no Mg2+}}$ trajectories to characterize C3d and CR3 inter-residue distance timeseries in the absence of Mg^{2+}. Heatmaps for each simulaton trajectory are labeled Run 1, Run 2 and Run 3 respectively are plotted from left to right as a timeseries. Chosen inter-residue pairs were observed within 5 Å of Mg^{2+} at an occupancy >50% in Run 2 and Run 3. Distances are calculated based on the closest heavy atom of the residue and the Mg^{2+} ion. Distance values are capped at 8 Å to maintain discretization in the colorbar at lower distance values as dissociation occurs in Run 1 and skews distance values.

Exploring inter-residue distances between C3d and CR3 in the C3d:CR3$^{\text{no Mg2+}}$ simulation trajectories shows fewer consistently conserved interactions (as seen in Figure 6.10) in comparison to the C3d:CR3$^{\text{Mg2+}}$ trajectories (Figure 6.8 and 6.9). It is likely that this may be a sign of potential dissociation (as is seen in Run 1) or tendency for dissociation. In fact, the most conserved inter-residue distances in Run 2 and Run 3 (C3d K1217 - CR3 E178, C3d D1245 - CR3 R208, and C3d D1247 - CR3 R208) are conserved in the early
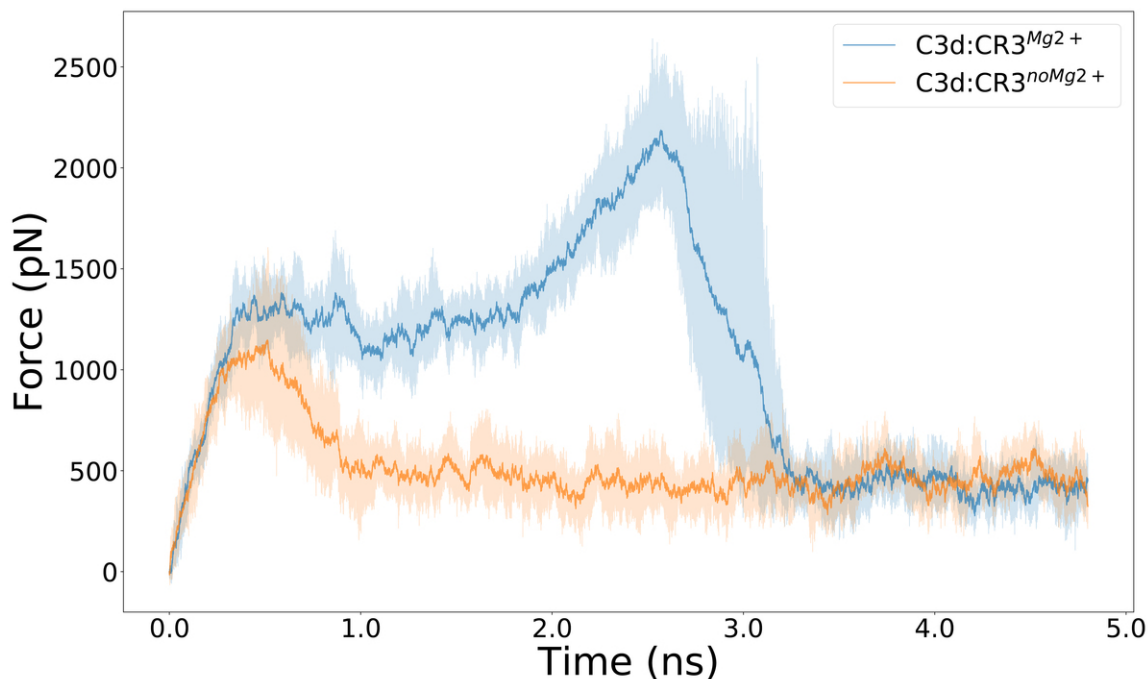
Figure 6.11: Force-time plots from the SMD simulations. Force-time curves for the nine trajectories from each system were averaged and plotted with the 95% confidence intervals.

stages of Run 1 prior to eventual dissociation. It is likely that the lack of the Mg^{2+} ion to steer the interaction results in dissociation given enough conformational sampling.

Examining the force-time plots from SMD trajectories sheds further light on the role of the Mg^{2+} ion at the interface (Figure 6.11). In all C3d:CR3^{Mg2+} SMD trajectories, complete dissociation occurs approximately between 2.8 to 3.2 ns as indicated by the sharp decline in the force-time plot and the corresponding confidence intervals. In the case of C3d:CR3$^{no\ Mg2+}$, dissociation occurs earlier around the 1 ns range. These results clearly indicate the significance of the Mg^{2+} ion in the dissociative stability of the complex. Additionally, through our analysis we confirm the importance of residues C3d D1247, CR3

S142, CR3 S144, and CR3 T209 based on their persistence in maintaining the dissociative stability of the complex.

6.4 Conclusion

Through a combination of electrostatic analysis and molecular dynamics simulations (both explicit solvent and steered), we have characterized the C3d:CR3 interaction and provided insight into the role of the Mg^{2+} ion in stabilizing the interaction. Amino acids identified to be crucial to the C3d:CR2 interaction (such as C3d D1247) exhibit reduced intermolecular interaction occupancies in the absence of Mg^{2+}, emphasizing its importance in steering the complex formation. The capability of C3d to simultaneously bind to CR2 and CR3 (as seen in Figure 6.12) has previously been explored experimentally[10] and has been suggested as the basis for a mechanism of how opsonized immune complexes may be transported from macrophages to B cells. Through our electrostatic and molecular dynamics analysis we conclude that amino acids shown to be significant to the C3d:CR3 interaction are distinct from those significant to the C3d:CR2 interaction[17,19]. The crucial role of C3d as a central fixture in several complement processes, through interactions with CR2, CR3, and FH[18], is reiterated by our results as well. These results provide a basis for potential therapeutic design as the proximity of these C3d-binding sites may be amenable for peptide design incorporating for example, key amino acids from CR2 and CR3.
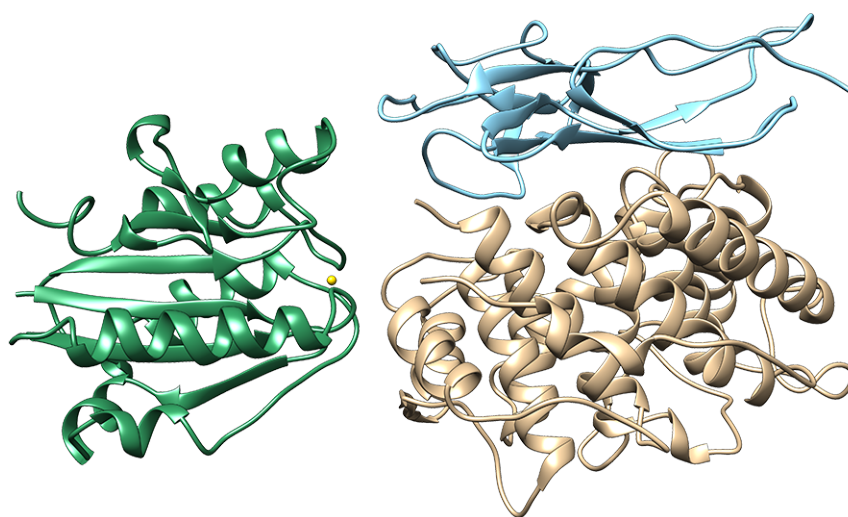
Figure 6.12: Molecular graphics representation of C3d:CR3 complex and C3d:CR2 complex superimposed. C3d, CR2, and CR3 are represented in tan, light blue, and green respectively. The Mg^{2+} ion at the interface is represented in yellow.

6.5 References

[1] Wang, Z., Thinn, A. M. M., and Zhu, J. A pivotal role for a conserved bulky residue at the a1-helix of the αI integrin domain in ligand binding. *Journal of Biological Chemistry*, 292(50):20756–20768, December 2017. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.M117.790519.

[2] Papanastasiou, M., Koutsogiannaki, S., Sarigiannis, Y., Geisbrecht, B. V., Ricklin, D., and Lambris, J. D. Structural Implications for the Formation and Function of the Complement Effector Protein iC3b. *The Journal of Immunology*, 198(8):3326–3335, April 2017. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.1601864.

[3] Xu, S., Wang, J., Wang, J.-H., and Springer, T. A. Distinct recognition of complement iC3b by integrins $\alpha_X \beta_2$ and $\alpha_M \beta_2$. *Proceedings of the National Academy of Sciences*, 114(13):3403–3408, March 2017. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1620881114.

[4] Carter, R. H. and Fearon, D. T. CD19: Lowering the Threshold for Antigen Receptor Stimulation of B Lymphocytes. *Science*, 256(5053):105–107, April 1992.

[5] Fearon, D. T. The complement system and adaptive immunity. *Seminars in Immunology*, 10(5):355–361, October 1998. ISSN 1044-5323. doi: 10.1006/smim.1998.0137.

[6] Dupuy, A. G. and Caron, E. Integrin-dependent phagocytosis – spreading from microadhesion to new concepts. *Journal of Cell Science*, 121(11):1773–1783, June 2008. ISSN 0021-9533, 1477-9137. doi: 10.1242/jcs.018036.

[7] Underhill, D. M. and Ozinsky, A. Phagocytosis of Microbes: Complexity in Action. *Annual Review of Immunology*, 20(1):825–852, 2002. doi: 10.1146/annurev.immunol.20.103001.114744.

[8] Morgan, H. P., Schmidt, C. Q., Guariento, M., Blaum, B. S., Gillespie, D., Herbert, A. P., Kavanagh, D., Mertens, H. D. T., Svergun, D. I., Johansson, C. M., Uhrín, D., Barlow, P. N., and Hannan, J. P. Structural basis for engagement by complement factor H of C3b on a self surface. *Nature Structural & Molecular Biology*, 18(4):463–470, April 2011. ISSN 1545-9985. doi: 10.1038/nsmb.2018.

[9] Phan, T. G., Grigorova, I., Okada, T., and Cyster, J. G. Subcapsular encounter and complement-dependent transport of immune complexes by lymph node B cells. *Nature Immunology*, 8(9):992–1000, September 2007. ISSN 1529-2916. doi: 10.1038/ni1494.

[10] Bajic, G., Yatime, L., Sim, R. B., Vorup-Jensen, T., and Andersen, G. R. Structural insight on the recognition of surface-bound opsonins by the integrin I domain of complement receptor 3. *PNAS*, 110(41):16426–16431, October 2013. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1311261110.

[11] Morikis, D. and Lambris, J. D. The Electrostatic Nature of C3d-Complement Receptor 2 Association. *J Immunol*, 172(12):7537–7547, June 2004. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.172.12.7537.

[12] Pyaram, K., Kieslich, C. A., Yadav, V. N., Morikis, D., and Sahu, A. Influence of electrostatics on the complement regulatory functions of Kaposica, the complement inhibitor of Kaposi's sarcoma-associated herpesvirus. *J. Immunol.*, 184(4):1956–1967, February 2010. ISSN 1550-6606. doi: 10.4049/jimmunol.0903261.

[13] Zhang, L. and Morikis, D. 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(9):3106–3119, May 2006. ISSN 0006-3495. doi: 10.1529/biophysj.105.068130.

[14] Kieslich, C. A., Vazquez, H., Goodman, G. N., de Victoria, A. L., and Morikis, D. The effect of electrostatics on factor H function and related pathologies. *Journal of Molecular Graphics and Modelling*, 29(8):1047–1055, August 2011. ISSN 1093-3263. doi: 10.1016/j.jmgm.2011.04.010.

[15] Gorham, R. D., Rodriguez, W., and Morikis, D. Molecular Analysis of the Interaction between Staphylococcal Virulence Factor Sbi-IV and Complement C3d. *Biophysical Journal*, 106(5):1164–1173, March 2014. ISSN 00063495. doi: 10.1016/j.bpj.2014.01.033.

[16] Kieslich, C. A. and Morikis, D. The Two Sides of Complement C3d: Evolution of Electrostatics in a Link between Innate and Adaptive Immunity. *PLoS Computational Biology*, 8(12):e1002840, December 2012. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002840.

[17] Mohan, R. R., Gorham Jr., R. D., and Morikis, D. A theoretical view of the C3d:CR2 binding controversy. *Molecular Immunology*, 64(1):112–122, March 2015. ISSN 0161-5890. doi: 10.1016/j.molimm.2014.11.006. 00002.

[18] E. S. Harrison, R., Gorham, R. D., and Morikis, D. Energetic evaluation of binding modes in the C3d and Factor H (CCP 19-20) complex. *Protein Science*, 24(5):789–802, May 2015. ISSN 1469-896X. doi: 10.1002/pro.2650. 00000.

[19] Mohan, R. R., Huber, G. A., and Morikis, D. Electrostatic Steering Accelerates C3d:CR2 Association. *J. Phys. Chem. B*, April 2016. ISSN 1520-6106. doi: 10.1021/acs.jpcb.6b02095. 00000.

[20] Chen, W., Lou, J., and Zhu, C. Molecular Dynamics Simulated Unfolding of von Willebrand Factor A Domains by Force. *Cel. Mol. Bioeng.*, 2(1):75–86, March 2009. ISSN 1865-5033. doi: 10.1007/s12195-009-0051-0.

[21] Meyer, S., Leusen, J. H., and Boross, P. Regulation of complement and modulation of its activity in monoclonal antibody therapy of cancer. *mAbs*, 6(5):1133–1144, September 2014. ISSN 1942-0862, 1942-0870. doi: 10.4161/mabs.29670.

[22] Schmidt, C. Q., Lambris, J. D., and Ricklin, D. Protection of host cells by complement regulators. *Immunological Reviews*, 274(1):152–171, November 2016. ISSN 01052896. doi: 10.1111/imr.12475.

[23] Hovingh, E. S., van den Broek, B., and Jongerius, I. Hijacking Complement Regulatory Proteins for Bacterial Immune Evasion. *Frontiers in Microbiology*, 7, December 2016. ISSN 1664-302X. doi: 10.3389/fmicb.2016.02004.

[24] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. UCSF Chimera?A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, October 2004. ISSN 0192-8651, 1096-987X. doi: 10.1002/jcc.20084.

[25] Harrison, R. E. S., Mohan, R. R., Gorham, R. D., Kieslich, C. A., and Morikis, D. AESOP: A Python Library for Investigating Electrostatics in Protein Interactions. *Biophysical Journal*, 112(9):1761–1766, May 2017. ISSN 0006-3495. doi: 10.1016/j.bpj.2017.04.005.

[26] Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res*, 32(Web Server issue):W665–W667, July 2004. ISSN 0305-1048. doi: 10.1093/nar/gkh381.

[27] MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B*, 102(18): 3586–3616, April 1998. ISSN 1520-6106. doi: 10.1021/jp973084f.

[28] Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *PNAS*, 98(18):10037–10041, August 2001. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.181342398.

[29] Gorham, R. D., Kieslich, C. A., and Morikis, D. Electrostatic Clustering and Free Energy Calculations Provide a Foundation for Protein Design and Optimization. *Ann Biomed Eng*, 39(4):1252–1263, April 2011. ISSN 0090-6964. doi: 10.1007/s10439-010-0226-9.

[30] Kieslich, C. A., Morikis, D., Yang, J., and Gunopulos, D. Automated computational framework for the analysis of electrostatic similarities of proteins. *Biotechnology Progress*, 27(2):316–325, March 2011. ISSN 87567938. doi: 10.1002/btpr.541.

[31] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16):1781–1802, December 2005. ISSN 0192-8651, 1096-987X. doi: 10.1002/jcc.20289.

[32] Case, D. A., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., Iii, V. W. D. C., Darden, T. A., Duke, R. E., Ghoreishi, D., Gilson, M. K., Gohlke, H., Goetz, A. W., Greene, D., Harris, R., Homeyer, N., Izadi, S., Kovalenko, A., Kurtzman, T., Lee, T. S., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Mermelstein, D. J., Merz, K. M., Miao, Y., Monard, G., Nguyen, C., Nguyen, H., Omelyan, I., Onufriev, A., Pan, F., Qi, R., Roe, D. R., Roitberg, A., Sagui, C., Schott-Verdugo, S., Shen, J., Simmerling, C. L., Smith, J., Salomon-Ferrer, R., Swails, J., Walker, R. C., Wang, J., Wei, H., Wolf, R. M., Wu, X., Xiao, L., York, D. M., and Kollman, P. A. *AMBER 2018*. University of California, San Francisco, 2018.

[33] McGibbon, R. T., Beauchamp, K. A., Harrigan, M. P., Klein, C., Swails, J. M., Hernández, C. X., Schwantes, C. R., Wang, L.-P., Lane, T. J., and Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal*, 109(8):1528–1532, October 2015. ISSN 0006-3495. doi: 10.1016/j.bpj.2015.08.015.

[34] Hernández, C. X., Harrigan, M. P., Sultan, M. M., and Pande, V. S. MSMExplorer: Data Visualizations for Biomolecular Dynamics. http://joss.theoj.org, April 2017.

[35] McKinney, W. Data Structures for Statistical Computing in Python. page 6, 2010.

[36] Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science Engineering*, 9(3):90–95, May 2007. ISSN 1521-9615. doi: 10.1109/MCSE.2007.55.

[37] Waskom, M., Botvinnik, O., drewokane, Hobson, P., David, Halchenko, Y., Lukauskas, S., Cole, J. B., Warmenhoven, J., Julian de Ruiter, Hoyer, S., Vanderplas, J., Villalba, S., Kunter, G., Quintero, E., Martin, M., Miles, A., Meyer, K., Augspurger, T., Yarkoni, T., Bachant, P., Williams, M., Evans, C., Fitzgerald, C., Brian, Wehner, D., Hitz, G., Ziegler, E., Qalieh, A., and Lee, A. Seaborn: V0.7.1 (June 2016). Zenodo, June 2016.

[38] Beauchamp, K. A., Bowman, G. R., Lane, T. J., Maibaum, L., Haque, I. S., and Pande, V. S. MSMBuilder2: Modeling Conformational Dynamics on the Picosecond to Millisecond Scale. *J. Chem. Theory Comput.*, 7(10):3412–3419, October 2011. ISSN 1549-9618. doi: 10.1021/ct200463m.

Chapter 7

Peptide redesign for inhibition of the complement system: Targeting age-related macular degeneration

7.1 Introduction

The complement system has been implicated as a major factor in the development and progression of age-related macular degeneration (AMD)[1,2]. Genome-wide associated studies (GWASs) have shown that single nucleotide polymorphisms (SNPs) in complement regulators Factor H and Factor I and complement proteins C3, C2, and Factor B are genetic risk factors for AMD[2–4]. An important GWAS finding is the Y402H SNP of Factor H, in which a tyrosine in position 402 is replaced by a histidine, resulting in the H402 risk variant[5–8]. It has been hypothesized that in the presence of the risk variant the complement

system is under-regulated, thus contributing to inflammation when activated locally in association with drusen deposits at the RPE-Bruch's membrane interface[4,9]. Although drusen formation, a characteristic accumulation of protein and membranous debris in AMD tissues, may not be initiated by the complement system, an over-activated (under-regulated) complement system has been shown to contribute to drusen accumulation and exacerbation of AMD pathology[4,10]. Therefore, inhibition of the complement system is a promising strategy to slow the progression of AMD pathogenesis.

Currently, AMD is treated using monoclonal antibody–based therapies targeting vascular endothelial growth factor (VEGF), which stimulates choroidal neovascularization and induces vascular leakage[11]. However, such therapies are effective in the wet (neovascular) form of AMD, associated with vessel rupture and local bleeding, but not in dry (atrophic) form of AMD that is characterized by the accumulation of drusen deposits and RPE atrophy. Compstatin family peptides were initially developed as inhibitors of complement-mediated autoimmune and inflammatory diseases, using phage display, functional, structural, computational studies (see review[12], and references therein). They became attractive low-molecular mass complement inhibitors for the treatment of AMD soon after the 2005 genomics studies implicated complement in AMD (see review[13], and references therein). Compstatin family peptides function by binding to complement protein C3 and sterically inhibiting the cleavage of C3 to C3a and C3b by convertase, thus impeding the formation of the chemotactic fragment C3a, the opsoninizing fragment C3b, and the propagation of the complement system through the common pathway that ultimately results in the assembly

of C5b-9n (also known as the membrane attack complex, MAC), a protein complex that forms pores on cell membranes.

One compstatin analog underwent clinical trials for AMD, and although the analog did not raise safety concerns, it did not show therapeutic efficacy. It is postulated that was likely the effects of molecular aggregation that resulted in the formation of gel-like structures[14,15] and an associated loss of functionality. This analog had been optimized over several years to have higher binding affinity than the original compstatin analogs by introducing a replacement of valine at position 4 with an aromatic amino acid, tyrosine[16,17] or tryptophan[18], and subsequently with methylated tryptophan[19]. The latter modification also increased the hydrophobic character of the peptide and presumably contributed to its aggregation in the aqueous ocular environment. Additional compstatin analogs are currently in clinical trials for various complement-mediated diseases[20].

Recent studies have focused on increasing the solubility of compstatin peptides, using structure-based rational design, computational modeling, and optimization[21,22]. These studies have identified several analogs with N-terminal extensions that have inhibitory activities similar to those of the most potent analogs and have higher aqueous solubilities. Increased solubility was made possible by introducing two polar amino acid extensions at the N-terminus. In one analog, an arginine at sequence position -1 not only contributed to solubility but was also shown by molecular dynamics simulations to form a salt bridge with a glutamic acid in C3, thus contributing to binding affinity as well[21,22]. In this study, we used a potent analog with arginine at position -1 and serine at position 0[22], Peptide 1 here (or Peptide 9 in ref[22]), as a template to further increase solubility by incorporating two polar

amino acid extensions and polyethylene glycol (PEG) blocks at the C-terminus. We demonstrate the high potency and high solubility of the new analog, using in vitro functional and solubility assays. We also demonstrate the efficacy of this peptide to inhibit complement activation in a human RPE cell–based assay that mimics AMD pathophysiology. We show that this PEGylated compstatin analog has significant promise as a therapeutic for AMD.

7.2 Methods

7.2.1 Peptide synthesis

Compstatin Peptides 1-3 (Table 7.1) were synthesized by WuXi AppTec (Shanghai, China). Peptide 2 has eight PEG blocks attached at the backbone of the C-terminal amino acid. A Competition peptide for use in the thermophoresis experiments was synthesized by ELIM Biopharm (Hayward, CA, USA) in two versions. One version was labeled with the cyanine fluorophore CY5, which was attached at the side chain of the preceding lysine, and had sequence Ac-I[CVWQDWGAHRC]TAGK-(CY5)-NH2. Another version was unlabeled and had sequence Ac-I[CVWQDWGAHRC]TAGK-NH2. All peptides were cyclized by a disulfide bridge between the two cysteine amino acids, and they were acetylated at the N-terminus and amidated at the C-terminus. The peptides had >95% purity, as determined with high-performance liquid chromatography (HPLC) and mass spectrometry (MS).

7.2.2 Hemolytic assay

Rabbit erythrocytes (Complement Technology, Inc., Tyler, TX) were washed with PBS (1X; 3.8 mM monobasic NaH_2PO_4, 16.2 mM dibasic Na_2HPO_4, 150 mM NaCl, pH 7.4)

Table 7.1: Peptide sequences

Peptide	Sequence	Molecular mass
1	Ac-RSI[CVWQDWGAHRC]T-NH$_2$	1856
2	Ac-RSI[CVWQDWGAHRC]T-PEG$_8$-NH$_2$	2279

Peptide 1 is a positive control with a two-polar amino acid N-terminal extension[22]. Peptide 2 is a new design that contains an 8-PEG block backbone extension. Brackets denote disulfide bridge cyclization between the two cysteine amino acids. Ac: acetylation blocking group; NH$_2$: amidation blocking group.

and then resuspended in a veronal-buffered saline solution (VBS 1X; 72.8 mM NaCl, 0.9 mM sodium barbital, 1.5 mM barbituric acid, pH 7.4) containing 5 mM MgCl$_2$ and 10 mM EGTA (VBS-MgEGTA). Twofold serial dilutions of the compstatin analogs were performed in round-bottom 96-well plates and then further diluted in VBS-MgEGTA. Normal human serum (NHS; Complement Technology, Inc.) diluted in VBS-MgEGTA was added to each well followed by incubation at room temperature for 15 min. Subsequently, 30 µL of rabbit erythrocytes at a concentration of 1.25×10^8 cells/ml were added to each well. Positive controls for lysis consisted of erythrocytes in deionized water and erythrocytes in NHS diluted with VBS-MgEGTA. Negative controls for lysis consisted of erythrocytes in VBS-MgEGTA and erythrocytes in NHS diluted in VBS-EDTA (20 mM EDTA). Next, plates were incubated at 37 °C for 20 min, and then ice-cold VBS containing 50 mM EDTA was added to each well to quench hemolytic reactions. The plates were centrifuged at 1000 $\times g$ for 5 min, and the supernatant was diluted 1:1 with deionized water in flat-bottom 96-well plates. Absorbance was measured spectrophotometrically at 405 nm to quantify lysis.

7.2.3 Apparent solubility measurements

Compstatin analogs were dissolved in PBS at pH 7.4 to concentrations of 10, 7.5, and 5 mg/ml. At each concentration point, the peptide solutions were shaken on a vortex mixer for 30 s and then centrifuged at 13,000 $\times g$ for 5 min. The supernatant was collected and measured 5 times spectrophotometrically at 280 nm. Optical densities were converted into concentrations according to the Beer-Lambert law. An extinction coefficient of $11\,125\,\mathrm{M^{-1}\,cm^{-1}}$ was used for each compstatin analog as each peptide contains two tryptophan amino acids (tryptophan extinction coefficient being $5562.5\,\mathrm{M^{-1}\,cm^{-1}}$).

7.2.4 Other experimental analysis

Other experimental studies such as RPE cell culture, immunofluorescence of sub-RPE deposits, and confocal imaging and analysis were performed by our collaborators for this study[23].

7.2.5 Structural modeling

A structural model of the eight-linked PEG blocks was generated using MOL-DRAW[24] [28] and then attached to a structural model of Peptide 1 using structure editing tools in Chimera[25] [29], to generate a structural model of Peptide 2. The structural model of Peptide 1 was derived from molecular dynamics simulations, based on the crystal structure of bound compstatin [30]. CHARMM parameters and topologies[26] [31] were used and modified to incorporate the peptide-like bond between PEG_8 and Peptide 1. Modifications were rationally chosen based on existing amino acid parameters and topologies. Angles

and dihedral angles that did not have a counterpart in existing CHARMM parameters and topologies were generated using SwissParam[27] [32].

7.2.6 Molecular dynamics simulation

An explicit-solvent molecular dynamics simulation was performed, for 90 ns, using as the initial structure the modeled Peptide 2 structure. The explicit-solvent environment consisted of a water box and counterions to represent the solvated environment of the biomolecular system. The peptide was solvated in a TIP3P water box with dimensions of $75 \text{ Å} \times 57 \text{ Å} \times 67 \text{ Å}$, and charges were neutralized with sodium and chloride counterions at 150 mM. The TIP3P water model is a standard 3-site water model used in explicit solvent MD simulations that describes the configuration of the water molecules (rigidity/flexibility), polarization, and interactions in the simulation. Preparation steps (minimization, heating, and equilibration) were carried out to remove strain in the system, to heat the system to the desired temperature, and to relax the system prior to the production simulation dynamics, as described in a previous study[28] [33]. Following 25,000 steps of conjugate gradient energy minimization, the system was heated from 0 to 300 K in 62 ps with protein atoms constrained to post-minimization positions. Subsequently, the system was equilibrated through five stages for 50 ps per stage. Force constants of 41.83, 20.92, 8.368, and 4.184 kJ/mol/Å^2 were applied during the first four stages, respectively, to harmonically constrain all protein atoms to their post-minimization positions. During the fifth stage of equilibration, only the backbone atoms were harmonically constrained using a force constant of 4.184 kJ/mol/Å^2. Following equilibration, a production run was performed for 90 ns with periodic boundary conditions, SHAKE algorithm, 2 fs time steps, Langevin pressure and temperature controls,

and particle-mesh Ewald electrostatics. The molecular dynamics trajectory (9,000 frames) was clustered using the root-mean-square deviation (RMSD) of the backbone or alpha carbon atoms of Peptide 2, and a representative structure from the highest-populated clusters was identified and depicted for molecular graphics visualization.

7.2.7 Microscale Thermophoresis

The binding affinity of Peptide 2 was evaluated in a competitive microscale thermophoresis assay, using a Monolith NT.115 instrument (NanoTemper Technologies GmbH, Munich, Germany). Competition was performed against the Competition Peptide, labeled with Cy5 for signal detection. Two different 1:1 serial dilution series of Peptide 2 were performed in MST buffer (50 mM Tris-HCl, 150 mM NaCl, 10 mM MgCl$_2$, 0.05% Tween 20). The first dilution series started with a final concentration of 333 µM and ended in a final concentration of 40.69 nM, while the second dilution series started with a final concentration of 166.67 µM and ended in a final concentration of 20.35 nM. To each dilution series of Peptide 2, purified C3c (Complement Technology) and the Competition Peptide were dissolved to final concentrations of 117 nM and 50 nM, respectively. The resulting mixture was incubated for 15 minutes in the dark at room temperature. Following incubation, the samples were loaded into hydrophilic capillary tubes and the thermophoretic response of the fluorescently labeled marker was measured. Each dilution series was performed in at least triplicate and the results from both dilution series were combined for estimation of the IC$_{50}$ through nonlinear regression.

The fragment C3c was chosen for the thermophoresis assay because it binds compstatin, but it does not contain the thioester domain (TED). It is actually the co-crystal

structure of C3c with a bound compstatin analog (Ac-I[CVWQDWGAHRC]T-NH$_2$) that has been reported in literature[29]. The bound compstatin analog of the crystal structure is the parent peptide of Peptide 1, with Peptide 1 having an Arg-Ser extension at sequence positions -1 and 0. The internal thioester bond of the TED undergoes spontaneous hydrolysis in C3, and the TED is highly mobile in C3 and C3b, possibly obscuring the protein's thermodiffusion properties in a direct (non-competitive) binding assay.

Results of the analysis can be found in Figure B.1.

7.3   Results

The objective of this study is to optimize the aqueous solubility while maintaining the binding affinity of Peptide 1 (Table 7.1), our compstatin peptide that previously had the most promise to become a therapeutic for AMD. Our previous studies[21,22] focused on improving the solubility of the peptide that underwent clinical studies for AMD, with sequence Ac-I[CV(meW)QDWGAHRC]T-NH$_2$[13]. This peptide had high aggregation propensity in aqueous solution, attributed to the peptide's reduced solubility compared to other less potent compstatin analogs[21]. It was first shown that adding polar dipeptides at the N-terminus (positions -1 and 0) improved solubility while maintaining potency, with the peptide Ac-RSI[CV(meW)QDWGAHRC]T-NH$_2$ was the most efficacious in the human RPE cell–based assay described in Methods (Peptide VI in ref[21]). In a subsequent study, it was deemed necessary to eliminate the methyl group from tryptophan at position 4 to further improve solubility, without compromising potency, resulting in the sequence of Peptide 1 (Peptide 9 in ref[22]).

To achieve our objective, we redesigned Peptide 1 by adding PEG block extensions at the C-terminus (Peptide 2; Table 7.1). The choice of the extension at the C-terminus of Peptide 2 was guided by the results of molecular dynamics simulations, which had shown that the C-terminus of compstatin points away from the C3-binding site toward the solvent[21,30]. Thus, we reasoned that such extensions would not interfere with the binding interface between the compstatin analog and C3. Peptide 2 contains eight PEG blocks attached at the peptide backbone in the C-terminus. The choice of PEG blocks was guided by earlier surface plasmon resonance (SPR) and enzyme-linked immunosorbent assay (ELISA) data, which had shown that PEGylated compstatin peptides had higher solubility compared to non-PEGylated peptides with the same sequence[31]. The addition of a spacer of eight PEG blocks to compstatin analogs was deemed necessary for the SPR binding experiments to increase the space between the peptides and attachment to the streptavidin sensor chip via lysine–biotin binding. This spacer aimed to increase the mobility of the peptides, enhance their accessibility to C3, and decrease non-specific interactions, thus emulating unbound ligand states as closely as possible within the experimental constraints. The inhibitory activities of peptides with the same sequences, but without the PEG blocks, were tested using ELISAs in the same study[31]. PEGylation is an established procedure in drug design and delivery, as this procedure has been shown to increase aqueous solubility and bioavailability, including enhanced structural and chemical stability and circulation lifespan, and reduced renal clearance[32]. In addition, PEGylation has a shielding effect on drugs, typically reducing drug immunogenicity, antigenicity, and toxicity[32]. Currently, there are several PEGylated drugs in the clinic, including pegaptanib, which has been used

for the treatment of age-related macular degeneration[33]. PEGylation has also been used in a compstatin variant with modified backbone, but in that case, a large (40 kDa) Y-shaped PEG structure was attached either at the N- or C-terminus[34], compared to the small (423 Da) linear 8-PEG block structure attached at the C-terminus in our study.

Figure 7.1 shows the dose–response curves of the complement hemolytic assay for the PEGylated Peptide 2 and the parent Peptide 1 (positive control). Peptides 1 and 2 have similar IC_{50} values within the confidence intervals from four replicate experiments, and therefore similar potencies (Table 7.2).

Table 7.2: IC_{50} Values from hemolytic assay

Peptide	Mean IC_{50}(µM)	95% Confidence interval	
		Upper	Lower
1	0.97	1.04	0.90
2	0.96	1.14	0.80
Competition	1.40	1.50	1.20

IC_{50}: Ligand concentration at 50% maximal inhibition. Data are from four replicate experiments (n=4). This table is an updated version of the one presented in ref[35].

We then tested our solubility objective. Figure 7.2A shows the difference between the calculated and spectrophotometrically measured (observed) concentrations in the range of 5–10 mg/ml. The concentration was experimentally measured using absorption spectroscopy at 280 nm and calculated using the weight per volume values of the dilution series, as described in Methods. Peptide 2 is much more soluble than Peptide 1 at 5 mg/ml, as the concentration difference is close to zero, the expected difference for nearly perfect solubility. In addition, Peptide 2 remains soluble up to 10 mg/ml with a slight deviation from the difference of zero, whereas the difference of Peptide 1 significantly deviates from zero and
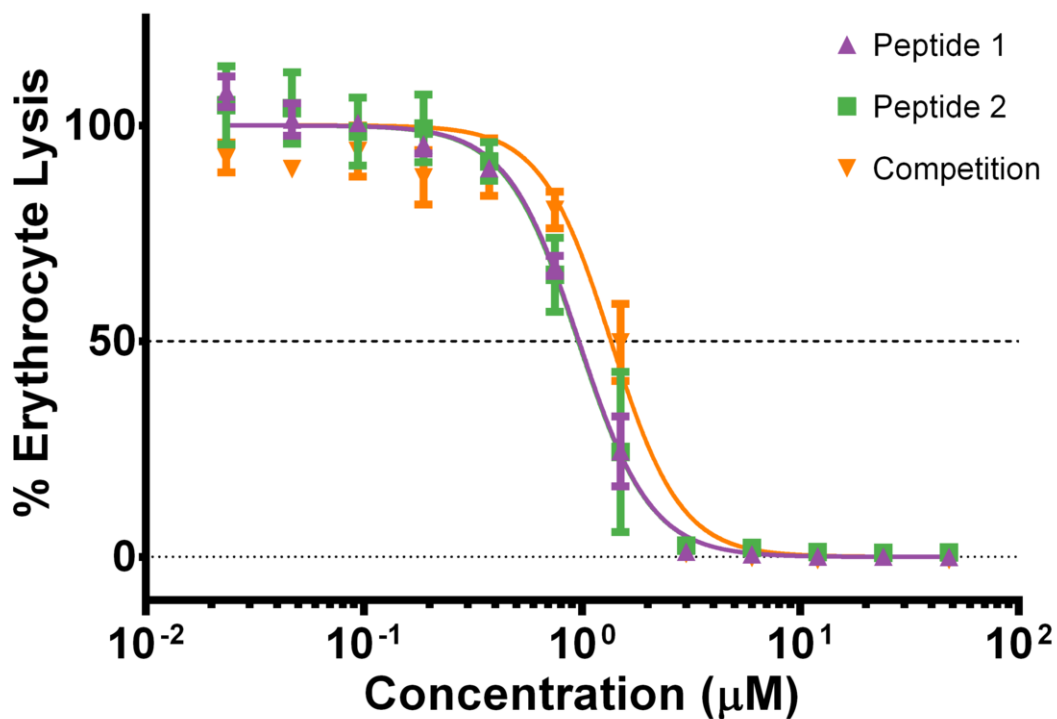
Figure 7.1: Concentration-dependent inhibition curves of compstatin peptides in four replicate hemolytic assay experiments. The plotted data represent the mean percent inhibition $\pm$ standard error of the mean (SEM). The dashed line intersects each inhibition curve at the IC_{50}. Peptide 1 is the parent peptide (positive control), Peptide 2 is the PEGylated form of the parent peptide, and third peptide is the unlabeled form of the Competition Peptide. This figure is an updated version of the one presented in ref[23].

from constancy as the concentration increases. Figure 7.2B shows a different presentation of the same data in the form of a correlation plot between the observed and calculated peptide concentrations in millimolar. The data for Peptide 2 show much higher correlation than those for Peptide 1. In addition, the fitted straight line of the data for Peptide 2 is closer to a straight line with slope 1 that passes through the origin (Figure 7.2B). A straight line with slope 1 represents perfect correlation, denoting the highest solubility and the lowest aggregation. These data demonstrate that Peptide 2 has significantly higher apparent solubility, or significantly lower tendency to aggregate, than the parent Peptide 1. Figure B.2 also shows that both Peptide 1 and 2 demonstrate better solubility than the Competition Peptide.

We measured direct binding of Peptide 2 to C3c using the microscale thermophoresis assay, described in Methods. Figure B.1 shows competitive binding of Peptide 2 to the C3c-Competition Peptide complex, where the Competition Peptide was labeled with the fluorophore Cy5 for detection of the thermophoresis signal. The unlabeled Competition Peptide has reduced potency compared to Peptide 2 (Table 7.2 and Figure 7.1), and for this reason was chosen for competitive replacement in the thermophoresis assay. This competitive binding experiment shows that Peptide 2 binds to C3c with a dissociation constant (K_D) of 0.69 ± 0.24 µM.

We continued our study by testing the efficacy of Peptides 1 and 2 in a human RPE cell–based assay[36], which was used in previous optimizations of compstatin family peptides[21,22]. The RPE cell-based assay, along with immunofluorescence and confocal imaging,

145
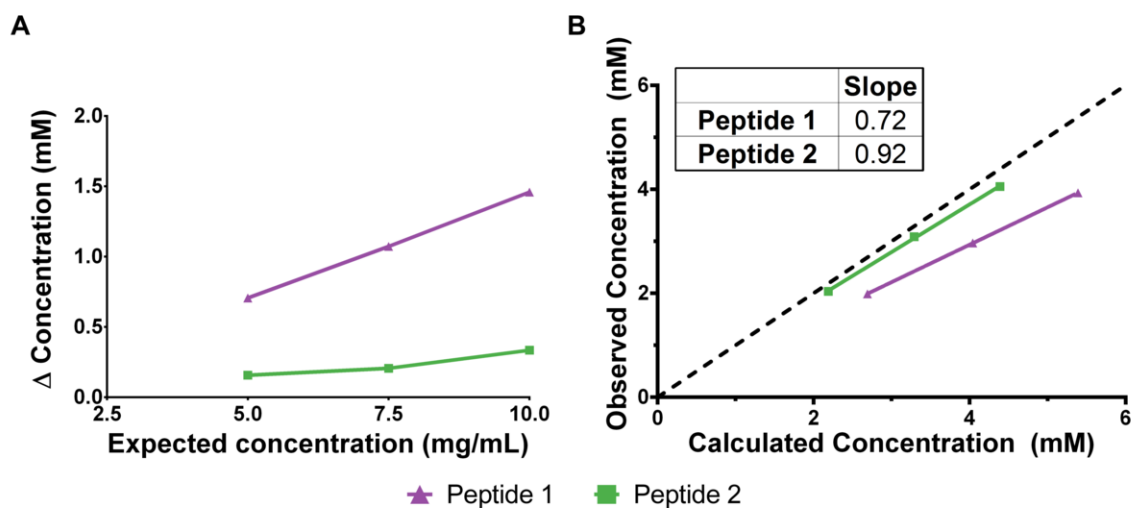
Figure 7.2: Apparent solubility of compstatin peptides. A: The difference in concentration, calculated – observed (measured), in mM, plotted against the expected concentration (in mg/ml). The actual observables are plotted, the expected concentration (in mg/ml) of the dilution series on the horizontal axis, and the observed (in mM, measured using tryptophan absorbance) minus the calculated (in mM, from the mg/ml expected concentration values) on the vertical axis. Peptide 2 is the most soluble, as the concentration difference is close to zero and remains nearly constant in the dilution series. B: Correlation of the observed concentration with the calculated concentration, with the concentrations presented in mM. Peptide 2 is the most soluble, as indicated by the high correlation between the observed and calculated concentrations (slope of 0.92). Each data point represents the mean measured concentration and the corresponding expected concentration with a linear regression fit to the data. A straight line of slope 1 passing through the origin is inserted to indicate the closeness of the data to perfect correlation.

as performed by our collaborators[23] demonstrated that the inhibitory effect of Peptide 2 was significantly greater (p<0.001) than that achieved by Peptide 1.

In combination, the hemolytic assay, apparent solubility, microscale thermophoresis, and human RPE cell–based assay data indicate that Peptide 2 is a more promising compstatin analog for further optimization and potential clinical translation, compared to Peptide 1. The parent Peptide 1 had emerged to be the best analog until now in previous studies, in terms of solubility and affinity balance and efficacy of complement inhibition in the human RPE cell–based assay (see refs[21,22] and references therein for earlier optimization studies).

To gain insight into the molecular features that contribute to the structural stability and solubility of Peptide 2, we performed an extended molecular dynamics simulation. Figure 7.3A shows representative conformations from the top (highest occupancy) five structural clusters derived from the molecular dynamics trajectory, using backbone atom RMSD-based clustering. These five clusters represent 79% of the conformations spanned by the peptide. The PEG_8 C-terminal extension demonstrates high local flexibility and global mobility, in essence forming a dynamic polar shell around Peptide 2. Flexibility and mobility of the PEG_8 extension are expected, given its polar character and interactions with water molecules of the solvent. This dynamic polar shell perhaps functions as a shield from self-association and aggregation of Peptide 2 owed to peptide's inherent hydrophobic features, thus contributing to the solubility of the peptide.

Figure 7.3B shows the conformation of the representative peptide from the structural cluster with highest occupancy, depicting key amino acid side chains for the optimiza-
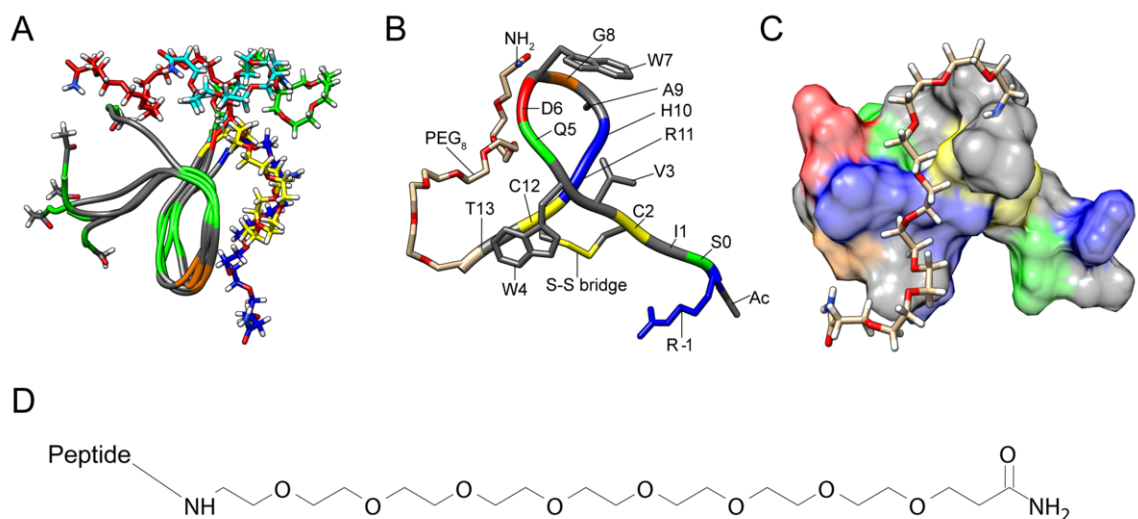
Figure 7.3: Molecular structure of Peptide 2. A: Representative conformers of Peptide 2 from the five highest-populated root-mean-square deviation (RMSD) clusters of the molecular dynamics trajectory, with occupancies of 38%, 12%, 10%, 10%, and 9%. The peptide backbone is shown in a tube representation, colored by the amino acid property type: gray for hydrophobic, green for polar, and brown for glycine. Cys2 and Cys12 are marked as hydrophobic because they form a disulfide bridge. The PEG_8 C-terminal extensions are shown in stick models of different colors for each conformer. The acetyl and amide terminal blocks are shown in stick models, colored by atom type: gray for carbon, white for hydrogen, blue for nitrogen, and red for oxygen. B: The major conformer of Peptide 2 (38% occupancy) in backbone tube representation, with key side chains in compstatin peptide optimization in stick representation. The location of all amino acids is marked. The following color code is used: gray for hydrophobic, green for polar neutral, blue for basic, red for acidic, yellow for cysteines and the disulfide bridge, and brown for glycine. The PEG_8-NH_2 C-terminal extension is shown in stick representation without hydrogens, with carbons depicted in brown, oxygens in red, and nitrogen in blue. C: Surface representation of Peptide 2, with the PEG_8-NH_2 extension shown in stick representation (including hydrogens). The color code is as in Panel B, with the added hydrogens of PEG_8-NH_2 shown in white. D: The chemical structure of PEG_8-NH_2.

148

tion of compstatin from research in the past 20 years. These amino acids are Arg(-1), Val3, Trp4, Trp7, Ala9, and the disulfide bridge Cys2-Cys12. The Arg(-1) addition corresponds to the Arg(-1)/Ser0 extension of Peptide 1[22], the parent analog of Peptide 2. Our latest addition in this work is the PEG_8-NH_2 (Figure 7.3D) extension at the C-terminus (Figure 7.3A–C), which results in Peptide 2, our most promising, in terms of affinity and solubility properties, lead peptide of the compstatin family until now.

7.4 Discussion

We report the design of a new compstatin peptide that has superior aqueous solubility and comparable complement inhibitory activity characteristics, compared to previously known peptides of the compstatin family. The new peptide, Peptide 2 (Table 7.1), is a PEGylated form of our previously most promising peptide in terms of inhibitory activity and aqueous solubility, Peptide 1 (Table 7.1)[22]. Peptide 2 has eight PEG blocks attached to the backbone C-terminus, accounting for an additional molecular mass of 423 Da compared to Peptide 1 (Table 7.1). Since the original discovery of compstatin using a phage-displayed random peptide library[37], there have been many benchmarks in the optimization of the sequence of compstatin. Initial structure-activity studies had derived a sequence template with seven amino acids being indispensable for inhibitory activity and six amino acids being optimizable[17,38]. Figure 7.3B shows the side chains of essential amino acids for the optimal binding and inhibitory activity of compstatin, including benchmark residue-specific optimization steps over the period of several years. Initial NMR, alanine scan, and inhibitory activity studies indicated that Val3 and Trp7 are important for binding to C3 and inhibi-

tion of complement activation[39,40]. A subsequent crystal structure of C3c in complex with a compstatin analog confirmed these findings, showing that Val3 and Trp7 are inserted in hydrophobic cavities[29]. Benchmark optimization steps include the incorporation of (i) Ala at position 9, which introduces helicity in the sequence and shifts a structural beta-turn from the central toward the C-terminal portion of the sequence[39,40]; (ii) aromatic amino acids at position 4[16] with Trp4 being optimal[18], which was shown to participate in a hydrophobic clustering in the crystal structure[29]; (iii) dipeptide N-terminal extensions, with Arg at position -1 being optimal because it increased solubility compared to previous peptides and introduced a new intermolecular salt bridge as shown by molecular dynamics simulations[21,22]; and in this work, (iv) C-terminal extension using an eight-block PEG construct, which greatly increases aqueous solubility, unprecedented by any other compstatin analog. In addition, acetylation at the N-terminus and amidation at the C-terminus contribute to improved activity[41]. All active peptides contain a Cys2-Cys12 disulfide bridge.

Compstatin has two surfaces, a hydrophobic and a polar one, as was pointed out by the original NMR studies[39,40], but it is the hydrophobic surface that makes the main contacts with the C3, as it was pointed out by the crystal structure[29]. It is likely that the hydrophobic surface is responsible for the aggregation properties of the compstatin analog that underwent the early clinical trials. This analog contained a methylated Trp4 residue, in which the hydrophobic methyl group had replaced the polar hydrogen of the indole amide group, making the peptide even more hydrophobic. We had reasoned in earlier optimization studies that incorporation of polar amino acid extensions at the termini would increase solubility without perturbing binding properties, and this was shown to be

the case for the N-terminus[21,22] and C-terminus (Mohan RR, Gorham RD Jr, Morikis D, unpublished data). However, the analog with the best aqueous solubility is Peptide 2 of this work that incorporates an 8-PEG block construct at the C-terminus. Our earlier molecular dynamics studies had shown that the C-terminus is mobile pointing outward from the C3-peptide interface toward the solvent. Thus, we reasoned that the 8-PEG block construct would not sterically interfere with the binding interface of the C3-peptide complex. This is evident in Figure 7.3, where the 8-PEG construct shows mobility around the non-binding site of Peptide 2, without specific contacts with the peptide. Therefore, PEGylation acts a solubilizer of compstatin.

To determine the potential of Peptide 2 and its parent Peptide 1 as AMD therapeutics, we evaluated the peptides in the human RPE cell–based assay (as performed by our collaboratiors[23]). Our collaborators demonstrated that Peptides 1 and 2 significantly inhibit the formation of C5b-9-rich sub-RPE deposits, with the PEGylated Peptide 2 exhibiting a twofold greater inhibitory effect than the parent Peptide 1. This difference in anti-C5b-9 effect may be attributed to lack of aggregation of Peptide 2, owing to its greater solubility compared to Peptide 1, which is expected to result in a higher "effective" concentration in the RPE culture. At the molecular level, the optimal solubility of Peptide 2 results from the formation of a dynamic polar shell, introduced by PEGylation, which predominantly surrounds and enhances the polar surface, and to a lesser extent the nonpolar surface of the peptide. This polar shell is designed to be a solubilizer, shielding Peptide 2 from self-association and higher-order aggregation, while leaving it unobstructed for binding the nonpolar surface.

In conclusion, we report the design of a new peptide analog of compstatin that combines an arginine-serine N-terminal polar amino acid extension and an 8-PEG block C-terminal extension. This peptide demonstrates significantly improved aqueous solubility and efficacy in a human RPE cell–based assay that mimics the pathobiology of AMD, compared to its parent peptide, while retaining comparable inhibitory activity against complement activation as its parent peptide. The new peptide can lead to a therapeutic treatment of dry AMD, as the peptide overcomes the aggregation limitation of a previous compstatin analog that underwent clinical trials.

7.5 References

[1] Liszewski, M. K. and Atkinson, J. P. Complement regulators in human disease: Lessons from modern genetics. *Journal of Internal Medicine*, 277(3):294–305, March 2015. ISSN 1365-2796. doi: 10.1111/joim.12338.

[2] Schramm, E. C., Clark, S. J., Triebwasser, M. P., Raychaudhuri, S., Seddon, J. M., and Atkinson, J. P. Genetic variants in the complement system predisposing to age-related macular degeneration: A review. *Molecular Immunology*, 61(2):118–125, October 2014. ISSN 01615890. doi: 10.1016/j.molimm.2014.06.032. 00029.

[3] Black, J. R. M. and Clark, S. J. Age-related macular degeneration: Genome-wide association studies to translation. *Genet Med*, 18(4):283–289, April 2016. ISSN 1098-3600. doi: 10.1038/gim.2015.70.

[4] Anderson, D. H., Radeke, M. J., Gallo, N. B., Chapin, E. A., Johnson, P. T., Curletti, C. R., Hancox, L. S., Hu, J., Ebright, J. N., Malek, G., Hauser, M. A., Rickman, C. B., Bok, D., Hageman, G. S., and Johnson, L. V. The Pivotal Role of the Complement System in Aging and Age-related Macular Degeneration: Hypothesis Re-visited. *Prog Retin Eye Res*, 29(2):95–112, March 2010. ISSN 1350-9462. doi: 10.1016/j.preteyeres. 2009.11.003.

[5] Hageman, G. S., Anderson, D. H., Johnson, L. V., Hancox, L. S., Taiber, A. J., Hardisty, L. I., Hageman, J. L., Stockman, H. A., Borchardt, J. D., Gehrs, K. M., Smith, R. J. H., Silvestri, G., Russell, S. R., Klaver, C. C. W., Barbazetto, I., Chang, S., Yannuzzi, L. A., Barile, G. R., Merriam, J. C., Smith, R. T., Olsh, A. K., Bergeron, J., Zernant, J., Merriam, J. E., Gold, B., Dean, M., and Allikmets, R. A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *PNAS*, 102(20):7227–7232, May 2005. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0501536102.

[6] Klein, R. J., Zeiss, C., Chew, E. Y., Tsai, J.-Y., Sackler, R. S., Haynes, C., Henning, A. K., SanGiovanni, J. P., Mane, S. M., Mayne, S. T., Bracken, M. B., Ferris, F. L., Ott, J., Barnstable, C., and Hoh, J. Complement Factor H Polymorphism in Age-Related Macular Degeneration. *Science*, 308(5720):385–389, April 2005. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1109557.

[7] Haines, J. L., Hauser, M. A., Schmidt, S., Scott, W. K., Olson, L. M., Gallins, P., Spencer, K. L., Kwan, S. Y., Noureddine, M., Gilbert, J. R., Schnetz-Boutaud, N., Agarwal, A., Postel, E. A., and Pericak-Vance, M. A. Complement Factor H Variant Increases the Risk of Age-Related Macular Degeneration. *Science*, 308(5720):419–421, April 2005. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1110359.

[8] Edwards, A. O., Ritter, R., Abel, K. J., Manning, A., Panhuysen, C., and Farrer, L. A. Complement Factor H Polymorphism and Age-Related Macular Degeneration.

Science, 308(5720):421–424, April 2005. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1110189.

[9] Hollyfield, J. G. Age-Related Macular Degeneration: The Molecular Link between Oxidative Damage, Tissue-Specific Inflammation and Outer Retinal Disease: The Proctor Lecture. *Invest. Ophthalmol. Vis. Sci.*, 51(3):1276–1281, March 2010. ISSN 1552-5783. doi: 10.1167/iovs.09-4478.

[10] Anderson, D. H., Mullins, R. F., Hageman, G. S., and Johnson, L. V. A role for local inflammation in the formation of drusen in the aging eye. *American Journal of Ophthalmology*, 134(3):411–431, September 2002. ISSN 0002-9394. doi: 10.1016/S0002-9394(02)01624-0.

[11] Lai, K. and Landa, G. Current choice of treatments for neovascular AMD. *Expert Review of Clinical Pharmacology*, 8(1):135–140, January 2015. ISSN 1751-2433. doi: 10.1586/17512433.2015.990379.

[12] Morikis, D. and Lambris, J. Structure, Dynamics, Activity, and Function of Compstatin and Design of More Potent Analogues. In *Structural Biology of the Complement System*, pages 317–340. CRC Press, March 2005. ISBN 978-0-8247-2540-2. 00000.

[13] Ricklin, D. and Lambris, J. D. Compstatin: A Complement Inhibitor on its Way to Clinical Application. *Adv Exp Med Biol*, 632:273–292, 2008. ISSN 0065-2598. 00134.

[14] Zarbin, M. A. and Rosenfeld, P. J. PATHWAY-BASED THERAPIES FOR AGE-RELATED MACULAR DEGENERATION: An Integrated Survey of Emerging Treatment Alternatives. *RETINA*, 30(9):1350, October 2010. ISSN 0275-004X. doi: 10.1097/IAE.0b013e3181f57e30.

[15] Yehoshua, Z., Rosenfeld, P. J., and Albini, T. A. Current Clinical Trials in Dry AMD and the Definition of Appropriate Clinical Outcome Measures. *Seminars in Ophthalmology*, 26(3):167–180, May 2011. ISSN 0882-0538. doi: 10.3109/08820538.2011.577132.

[16] Klepeis, J. L., Floudas, C. A., Morikis, D., Tsokos, C. G., Argyropoulos, E., Spruce, L., and Lambris, J. D. Integrated Computational and Experimental Approach for Lead Optimization and Design of Compstatin Variants with Improved Activity. *J. Am. Chem. Soc.*, 125(28):8422–8423, July 2003. ISSN 0002-7863. doi: 10.1021/ja034846p. 00101.

[17] Morikis, D., Soulika, A. M., Mallik, B., Klepeis, J. L., Floudas, C. A., and Lambris, J. D. Improvement of the anti-C3 activity of compstatin using rational and combinatorial approaches. *Biochemical Society Transactions*, 32(1):28–32, 2004.

[18] Mallik, B., Katragadda, M., Spruce, L. A., Carafides, C., Tsokos, C. G., Morikis, D., and Lambris, J. D. Design and NMR Characterization of Active Analogues of

Compstatin Containing Non-Natural Amino Acids. *J. Med. Chem.*, 48(1):274–286, January 2005. ISSN 0022-2623. doi: 10.1021/jm0495531. 00072.

[19] Katragadda, M., Magotti, P., Sfyroera, G., and Lambris, J. D. Hydrophobic Effect and Hydrogen Bonds Account for the Improved Activity of a Complement Inhibitor, Compstatin. *J. Med. Chem.*, 49(15):4616–4622, July 2006. ISSN 0022-2623. doi: 10.1021/jm0603419.

[20] Morgan, B. P. and Harris, C. L. Complement, a target for therapy in inflammatory and degenerative diseases. *Nature Reviews Drug Discovery*, 14(12):857–877, October 2015. ISSN 1474-1776, 1474-1784. doi: 10.1038/nrd4657. 00002.

[21] Gorham, R. D., Forest, D. L., Tamamis, P., López de Victoria, A., Kraszni, M., Kieslich, C. A., Banna, C. D., Bellows-Peterson, M. L., Larive, C. K., Floudas, C. A., Archontis, G., Johnson, L. V., and Morikis, D. Novel compstatin family peptides inhibit complement activation by drusen-like deposits in human retinal pigmented epithelial cell cultures. *Experimental Eye Research*, 116:96–108, November 2013. ISSN 00144835. doi: 10.1016/j.exer.2013.07.023. 00009.

[22] Gorham, R. D., Forest, D. L., Khoury, G. A., Smadbeck, J., Beecher, C. N., Healy, E. D., Tamamis, P., Archontis, G., Larive, C. K., Floudas, C. A., Radeke, M. J., Johnson, L. V., and Morikis, D. New Compstatin Peptides Containing N-Terminal Extensions and Non-Natural Amino Acids Exhibit Potent Complement Inhibition and Improved Solubility Characteristics. *J. Med. Chem.*, 58(2):814–826, January 2015. ISSN 0022-2623. doi: 10.1021/jm501345y. 00002.

[23] Mohan, R. R., Cabrera, A. P., Harrison, R. E., Gorham, R. D., Johnson, L. V., Ghosh, K., and Morikis, D. Peptide redesign for inhibition of the complement system: Targeting age-related macular degeneration. *Molecular Vision*, 22:1280–1290, 2016. 00000.

[24] Ugliengo, P., Viterbo, D., and Chiari, G. MOLDRAW: Molecular graphics on a personal computer. *Zeitschrift für Kristallographie-Crystalline Materials*, 208(1-2):383–384, 1993. 00209.

[25] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. UCSF Chimera?A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, October 2004. ISSN 0192-8651, 1096-987X. doi: 10.1002/jcc.20084.

[26] Brooks, B., Brooks, C., MacKerell, A., Nilsson, L., Petrella, R., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R., Post, C., Pu, J., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D., and Karplus, M. CHARMM: The Biomolecular Simulation Program. *J Comput Chem*, 30(10):1545–1614, July 2009. ISSN 0192-8651. doi: 10.1002/jcc.21287.

[27] Zoete, V., Cuendet, M. A., Grosdidier, A., and Michielin, O. SwissParam: A fast force field generation tool for small organic molecules. *J. Comput. Chem.*, 32(11):2359–2368, August 2011. ISSN 1096-987X. doi: 10.1002/jcc.21816. 00229.

[28] Mohan, R. R., Gorham Jr., R. D., and Morikis, D. A theoretical view of the C3d:CR2 binding controversy. *Molecular Immunology*, 64(1):112–122, March 2015. ISSN 0161-5890. doi: 10.1016/j.molimm.2014.11.006. 00002.

[29] Janssen, B. J. C., Halff, E. F., Lambris, J. D., and Gros, P. Structure of Compstatin in Complex with Complement Component C3c Reveals a New Mechanism of Complement Inhibition. *J. Biol. Chem.*, 282(40):29241–29247, May 2007. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.M704587200.

[30] Tamamis, P., López de Victoria, A., Gorham, R. D., Bellows-Peterson, M. L., Pierou, P., Floudas, C. A., Morikis, D., and Archontis, G. Molecular Dynamics in Drug Design: New Generations of Compstatin Analogs: New Generations of Compstatin Analogs. *Chemical Biology & Drug Design*, 79(5):703–718, May 2012. ISSN 17470277. doi: 10.1111/j.1747-0285.2012.01324.x.

[31] López de Victoria, A., Gorham, R. D., Bellows-Peterson, M. L., Ling, J., Lo, D. D., Floudas, C. A., and Morikis, D. A New Generation of Potent Complement Inhibitors of the Compstatin Family. *Chemical Biology & Drug Design*, 77(6):431–440, June 2011. ISSN 1747-0285. doi: 10.1111/j.1747-0285.2011.01111.x. 00021.

[32] Kang, J. S., DeLuca, P. P., and Lee, K. C. Emerging PEGylated drugs. *Expert Opinion on Emerging Drugs*, 14(2):363–380, June 2009. ISSN 1472-8214. doi: 10. 1517/14728210902907847.

[33] Li, W., Zhan, P., De Clercq, E., Lou, H., and Liu, X. Current drug research on PEGylation with small molecular agents. *Progress in Polymer Science*, 38(3):421–444, March 2013. ISSN 0079-6700. doi: 10.1016/j.progpolymsci.2012.07.006.

[34] Risitano, A. M., Ricklin, D., Huang, Y., Reis, E. S., Chen, H., Ricci, P., Lin, Z., Pascariello, C., Raia, M., Sica, M., Vecchio, L. D., Pane, F., Lupu, F., Notaro, R., Resuello, R. R. G., DeAngelis, R. A., and Lambris, J. D. Peptide inhibitors of C3 activation as a novel strategy of complement inhibition for the treatment of paroxysmal nocturnal hemoglobinuria. *Blood*, 123(13):2094–2101, March 2014. ISSN 0006-4971, 1528-0020. doi: 10.1182/blood-2013-11-536573.

[35] Mohan, R. R., Huber, G. A., and Morikis, D. Electrostatic Steering Accelerates C3d:CR2 Association. *J. Phys. Chem. B*, April 2016. ISSN 1520-6106. doi: 10.1021/acs.jpcb.6b02095. 00000.

[36] Johnson, L. V., Forest, D. L., Banna, C. D., Radeke, C. M., Maloney, M. A., Hu, J., Spencer, C. N., Walker, A. M., Tsie, M. S., Bok, D., Radeke, M. J., and Anderson, D. H. Cell culture model that mimics drusen formation and triggers complement activation associated with age-related macular degeneration. *Proceedings of the National Academy*

of Sciences, 108(45):18277–18282, November 2011. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1109703108. 00062.

[37] Sahu, A., Kay, B. K., and Lambris, J. D. Inhibition of human complement by a C3-binding peptide isolated from a phage-displayed random peptide library. *J Immunol*, 157(2):884–891, July 1996. ISSN 0022-1767, 1550-6606. 00219.

[38] Morikis, D. and Lambris, J. D. Structural aspects and design of low-molecular-mass complement inhibitors. *Biochemical Society Transactions*, 30(6):1026–1036, November 2002. ISSN 0300-5127, 1470-8752. doi: 10.1042/bst0301026. 00060.

[39] Morikis, D., Assa-Munt, N., Sahu, A., and Lambris, J. D. Solution structure of Compstatin, a potent complement inhibitor. *Protein science: a publication of the Protein Society*, 7(3):619, 1998.

[40] Morikis, D., Roy, M., Sahu, A., Troganis, A., Jennings, P. A., Tsokos, G. C., and Lambris, J. D. The Structural Basis of Compstatin Activity Examined by Structure-Function-based Design of Peptide Analogs and NMR. *J. Biol. Chem.*, 277(17):14942–14953, April 2002. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.M200021200. 00060.

[41] Sahu, A., Soulika, A. M., Morikis, D., Spruce, L., Moore, W. T., and Lambris, J. D. Binding Kinetics, Structure-Activity Relationship, and Biotransformation of the Complement Inhibitor Compstatin. *The Journal of Immunology*, 165(5):2491–2499, September 2000. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.165.5.2491. 00115.

Chapter 8

Virtual Screening of Chemical Compounds for Discovery of Complement C3 Ligands

8.1 Introduction

The complement system, consisting of over 40 soluble and cell-bound proteins, is integral to innate immunity[1–5]. In the event of pathogen exposure or injury, cascading complement response occurs resulting in opsonization, chemotaxis, phagocytosis, and lysis[6,7]. Complement activity is also double-edged as a lack of regulation or balance in its response can be observed in numerous autoimmune and inflammatory diseases, including age-related macular degeneration, lupus, rheumatoid arthritis, multiple sclerosis, Sjögren syndrome, scleroderma, chronic obstructive pulmonary disease, ischemia reperfusion injuries, and rare

diseases such as paroxysmal nocturnal hemoglobinuria, atypical hemolytic uremic syndrome, and C3 glomerulopathy, among others[8–10]. Currently, there are clinically available drugs for only two targets within the complement cascade, variations of the natural protein inhibitor C1-INH and a C5 inhibiting monoclonal antibody eculizumab, both of them being protein-based therapeutics[9,10].

Compstatin[11] is a cyclic peptide capable of inhibiting complement response through C3[12], originally discovered using a phage-displayed peptide library screening[13], subsequently reaching clinical trials for age-related macular degeneration and other complement-related diseases[14,15]. We have been involved in structure- and computation-based optimization of compstatin family peptides, originally using a major structural conformer of free compstatin from solution NMR studies (reviewed in refs[16–21]) and subsequently using bound structures from computational de novo design studies and molecular dynamics (MD) simulations, on the basis of the crystal structure of a compstatin analogue bound to C3c[22] (e.g., see refs[23–28]). Although our most recent design has led to overcoming solubility/aggregation issues of the previously most potent compstatin analogues[26,28], peptides in general suffer from low stability and bioavailability in vivo and often require intravenous administration. Chemical compounds are typically orally administered and are more cost effective for scaled-up industrial production. Therefore, there is a need for the development of nonpeptidic low-molecular-mass inhibitors. Toward this goal, we launched a pharmacophore-based virtual screening study, described here, to identify druglike chemical compounds with the geometric and physicochemical characteristics of compstatin, which are capable of binding to the compstatin binding site of target protein C3.

Virtual screening has proven to be a valuable methodology for identifying potential therapeutic candidates [29] and an alternative to fragment-based chemical compound design [30] or rational peptide design [31–33]. Virtual screening has the benefit of being a high-throughput method and can be used to screen millions of chemical compounds, while being more time and resource efficient than experimental high-throughput screening methods, thanks to the advances in computer hardware architecture and drug design-related algorithms [34–36]. Small druglike compounds are desirable because they typically exhibit better pharmacological properties, but at the expense of lower specificity, compared with peptide- or protein-based therapeutics. In this study, we utilize virtual screening with the objective to identify novel druglike compounds capable of binding in the compstatin binding site of C3 and potentially inhibiting complement response. We use a molecular dynamics structure of a potent compstatin analogue bound to C3c as the basis for the development of pharmacophore models and docking of molecules. Our virtual screening framework is similar to that used in a recent identification of 11 druglike ligands of complement fragment C3d, 10 of which are fluorescent markers of complement activation [37].

8.2   Methods

8.2.1   Pharmacophore Models

A pharmacophore model is represented by a framework of features corresponding to the spatial distribution of physicochemical properties (aromaticity, hydrophobicity, hydrogen bond donor/acceptor capability, and positive/negative charge) of an active ligand. During screening, a database of molecules is compared against the pharmacophore

model and if the molecule contains matching pharmacophore features, then it is considered a positive hit.

The workflow of the primary approach procedure is shown in Figure 8.1. Pharmacophore models were developed using a molecular dynamics (MD) trajectory of the complex between C3c and the RSI-compstatin analogue with sequence Ac-RSI[CVWQDWGAHRC]T-NH$_2$ (brackets denote cyclization through a disulfide bridge)[26]. Pharmacophore features were primarily selected from earlier molecular dynamics data of C3c-bound compstatin analogues[27,38], with the aid of prior knowledge from MD studies[24,39] and optimization studies[23–25,27] of C3c-bound compstatin analogues and optimization studies of free compstatin analogues[16]. All MD simulations were based on the crystal structure of C3c with the W4A9 analogue of compstatin[22].

Mechanistic binding analysis of compstatin structure throughout the MD trajectory was performed using the R package Bio3D, the Python library MDTraj, and Chimera[40–43] to identify significant physicochemical properties and nonpolar contacts at the binding site. In addition, free-energy contributions of individual amino acids and hydrogen bond occupancies from previous studies[27,38] informed the selection of pharmacophore features. Mean positions of centers of mass were calculated for the atoms identified in each selected feature. The tolerance radii of each pharmacophore feature were defined by calculating the conformational flexibility of the feature from the MD trajectory. In the first round of screening, 473 pharmacophore models were developed using subsets of 3–5 features identified to be of interest. The purchasable subset of the ZINC 12 database[44], consisting at the time of the study of ~19 million molecules (190 million conformers) in stock or to
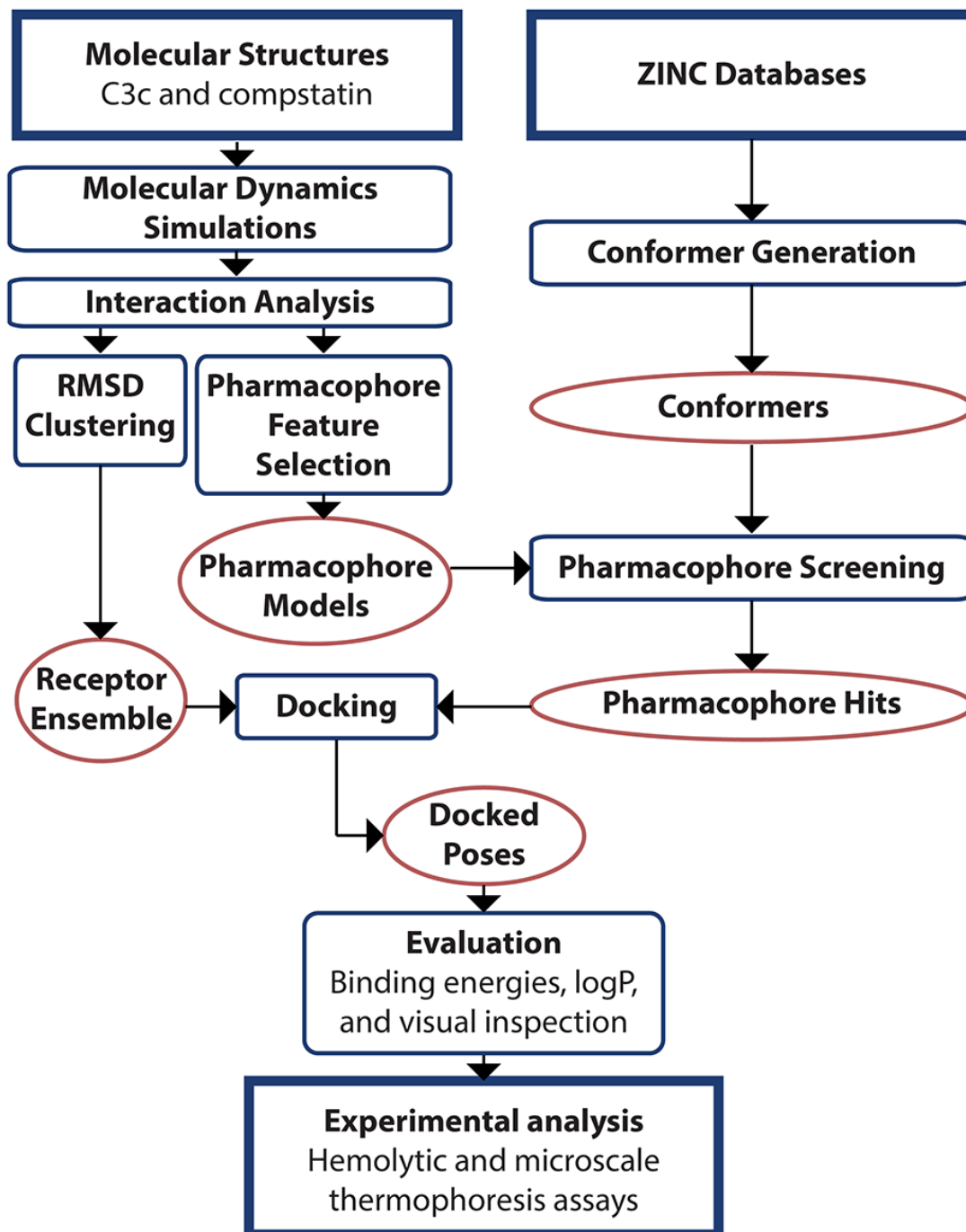
Figure 8.1: Flowchart of the primary virtual screening approach.

be made-to-order, was screened using ZINCPharmer[45] with each of the 473 pharmacophore models. A molecule was a positive hit during screening if its chemical moieties were spatially distributed such that there was overlap with the tolerance radii of the specific features of the pharmacophore model.

In the second round of screening, 40 new pharmacophore models consisting of subsets of 3-6 features were developed with iterative improvements over the initial 473 models. A set of ~7 million molecules with ~1.1 billion conformers that was used in one of our previous virtual screening studies[37] was used for screening of these 40 pharmacophore models using Phase[46,47]. This set of molecules was from the Drugs Now subset of the ZINC 12 database, consisting of molecules that were in stock at the time of the study, and the conformers were generated using Phase.

8.2.2   Docking

Molecules identified to fulfill the selection criteria of the pharmacophore models as positive hits during the pharmacophore screening were docked to C3c in the binding region of RSI-compstatin. In the first round of pharmacophore screening and docking, the molecules were docked to a single conformation of C3c (acquired from the final frame of the MD trajectory of the C3c/RSI-compstatin complex). In the second round of screening, representative conformational states of the binding site of C3c were extracted from the MD trajectory of the C3c/RSI-compstatin complex to accurately capture the conformational variations of the binding site. The conformational states of C3c observed in each frame of the MD trajectory were superimposed on the basis of C_α atoms and hierarchical clustering was performed on the basis of the root-mean-square deviation (RMSD) of C3c amino

acids identified to be involved in the interaction with RSI-compstatin[27]. Five clusters were calculated, corresponding to five representative structures of C3c. Molecular docking was performed using AutoDock Vina[48] with preprocessing of structures using AutoDock Tools. Each molecule was docked to the representative structures of C3c within a box (with dimensions of $28\,\text{Å} \times 28\,\text{Å} \times 28\,\text{Å}$), encompassing the entire RSI-compstatin binding site of C3c. The exhaustiveness parameter of AutoDock Vina was set to 20 to improve docking accuracy, and the top 20 docked poses of each molecule, based on predicted binding energies, were returned.

8.2.3 Scoring

Each of the docked poses were scored using the Vina scoring function, and the predicted binding energies were reported. Mean predicted binding energies were calculated for docked poses to each of the five representative structures of C3c. The predicted solubility of the molecules was calculated using the partition coefficient ($\log P$) using the ChemmineR package in R[49]. Molecules were also evaluated using ChemmineR, visually inspected with Chimera, and verified through the ZINC 12 database for adherence to Lipinski's Rule of Five[50]. A combination of the above scoring methods, in addition to visual inspection of the molecules for geometric properties and occurrence of specific pharmacophore features, were utilized to select 58 compounds for ordering and experimental testing.

8.2.4 Alternative Approach

An alternative virtual screening approach was performed by our collaborators[51]. The top 6 compounds from this approach were selected for experimental validation.

8.2.5 Experimental Validation

Hemolytic Assays

Selected compounds were obtained from ChemBridge (compounds 1–55), Asinex (compounds 57 and 58), MolPort (compounds 56, A–C, E, and F), and Specs (compound D). Stock solutions were prepared by dissolving each compound in dimethyl sulfoxide (DMSO) to concentrations of 4 mM.

Rabbit erythrocytes (Complement Technology, Inc.) were washed in phosphate-buffered saline and resuspended in a veronal-buffered saline solution containing 5 mM $MgCl_2$ and 10 mM ethylene glycol tetraacetic acid (EGTA) (VBS-MgEGTA). Each compound was diluted in VBS-MgEGTA to end up with a final concentration of 1% DMSO and was added to round-bottom 96-well plates. Normal human serum (NHS) diluted in VBS-MgEGTA was added to each well, and the plates were incubated at room temperature for 15 min. Thirty microliters of rabbit erythrocytes at a concentration of 1.25×10^8 cells/mL were then added to each well. Positive controls for lysis included rabbit erythrocytes in deionized water and rabbit erythrocytes in NHS diluted in VBS-MgEGTA, whereas negative controls included rabbit erythrocytes in VBS-MgEGTA and VBS-EDTA (20 mM EDTA). Following the addition of rabbit erythrocytes, the plates were incubated at 37 °C for 20 min and then quenched with ice-cold VBS with 50 mM EDTA. After centrifugation at $1000g$ for 5 min, the supernatant from the plates was diluted 1:1 with deionized water in flat-bottom 96-well plates and the lysis was quantified spectrophotometrically at 405 405 nm. The assays were performed in triplicate to ensure reproducibility.

Microscale Thermophoresis

The binding affinity for C3c of the top 10 compounds identified in the primary approach and the top 6 compounds identified in the secondary approach were evaluated in a competitive microscale thermophoresis (MST) assay, using a Monolith NT.115 instrument (NanoTemper Technologies GmbH, Munich, Germany). Competition was performed against a competition peptide, synthesized by ELIM Biopharm (Hayward, CA). The competition peptide was labeled with the cyanine fluorophore CY5, which was attached at the side chain of the preceding lysine, and had sequence Ac-I[CVWQDWGAHRC]TAGK-(CY5)-NH$_2$. The peptide was cyclized by a disulfide bridge between the two cysteine amino acids and acetylated at the N-terminus and amidated at the C-terminus. A 1:1 serial dilution series of each of the selected compounds was performed in MST buffer (50 mM Tris–HCl, 150 mM NaCl, 10 mM MgCl2, and 0.05% Tween 20) with 5% DMSO. Each dilution series started with a final concentration of 500 µM and ended in a final concentration of 15.3 nM. To each dilution series of a compound, purified C3c (Complement Technology) and the competition peptide (labeled with Cy5) were dissolved to a final concentration of 20.4 nM (C3c) and 50 nM (competition peptide). The resulting mixture was incubated for 15 min in the dark at room temperature. Following incubation, the samples were loaded into standard capillary tubes and the thermophoretic response of the fluorescently labeled marker was measured. Each dilution series was performed in triplicate and estimation of the IC$_{50}$ was performed through nonlinear regression.

The fragment C3c that was chosen for the MST assay binds compstatin but it does not contain the thioester domain. The choice of C3c for the MST assay is because the only

cocrystal structure of a compstatin family peptide bound to C3/C3 fragment reported in the literature is with C3c[22], and the sequence of the peptide is Ac-I[CVWQDWGAHRC]T-NH$_2$. The bound compstatin analogue of the crystal structure is the parent peptide of the competition peptide.

We utilized MST to determine protein–ligand binding affinities[52,53].(59,60) MST measures binding that occurs in a bulk solution, avoiding artifacts encountered in methods similar to surface plasmon resonance where one binding partner must be immobilized onto a chip. Although other methods such as isothermal titration calorimetry (ITC) and fluorescent polarization (FP) can also be used to measure binding in solution, these methods are not as suitable for our application as MST. ITC is a low-throughput method that compensates for low sensitivity by increasing the amount of sample measured. This strategy is prohibitive in cost for evaluating multiple protein–ligand pairs. FP, on the other hand, can be performed in a high-throughput manner but would require fluorescent labeling of small molecules to detect changes in polarization of emitted light when binding occurs. As our molecules have small molecular weight, less than 500 Da, labeling is likely to perturb protein–ligand interactions. In contrast to ITC and FP, MST is more sensitive, requiring minimal sample volume (20 µL for a dilution series), and detects changes in surface area, hydration entropy, and net charge. Our MST instrument requires fluorescent labeling of one binding partner; labeling is less likely to perturb binding of a small molecule as the target protein with a large molecular weight can be labeled. Although not strictly a high-throughput device, our MST device is relatively fast and can determine a dissociation constant for a single protein–ligand interaction within the period of an hour.

8.3 Results and Discussion

8.3.1 Virtual High-Throughput Screening

The objective for our study was to identify low-molecular mass molecules that are capable of binding complement protein C3 or its activation fragments, C3b/C3c, in the binding site of compstatin. Our study was based on previous knowledge of the structure of free and bound compstatin and many computational and experimental studies, which pointed to key structural and physicochemical features of compstatin that are important for binding to C3/C3b/C3c and for inhibiting the complement system. Our approach involved the development of pharmacophore models, pharmacophore-based virtual screening of conformationally flexible molecules, docking of pharmacophore-matched molecules to multiple conformations of the C3/C3b/C3c binding site, and scoring of docking poses using energetics, lipophilicity, and Lipinski's rule of five criteria. Figure 8.2 shows a schematic flowchart of our approach.

In the first round of virtual screening with the initial 473 pharmacophore models (Appendices S1 and S2 of ref[51]), we were able to identify specific combinations of features that were likely to yield molecules of interest. The screening of the purchasable subset of the ZINC 12 database with the pharmacophore models suggested that features corresponding to R(-1), V3, W4, Q5, W7, A9, and H10 on RSI-compstatin were necessary as those features resulted in positive hits that had favorable predicted binding affinities as well as a proper docked fit when visually inspected. A positive charge pharmacophore feature was chosen at the position of R(-1). The R(-1)-S0 N-terminal modification was chosen to improve
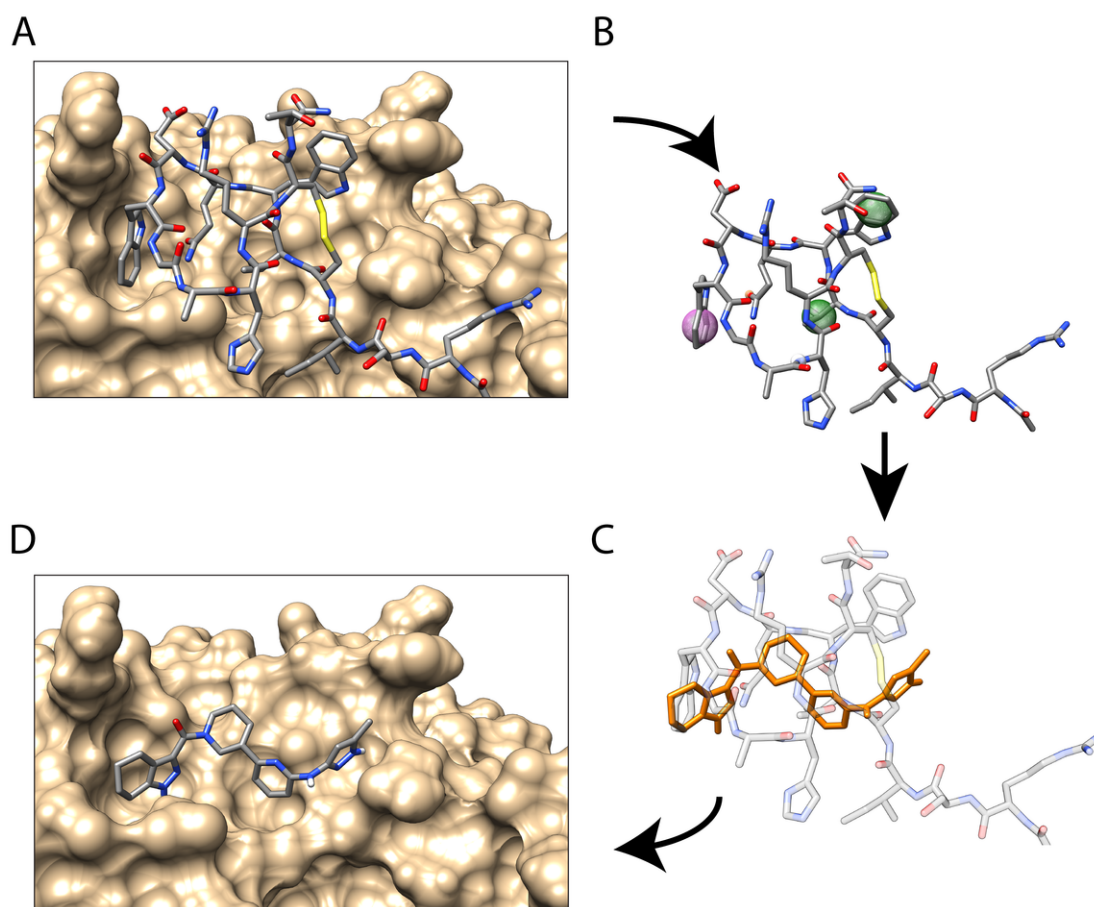
Figure 8.2: Diagram outlining an example of the pharmacophore model generation and docking output. (A) Molecular graphics showing the binding interaction between compstatin and C3c. (B) A pharmacophore model generated through selection of features identified from the C3c–compstatin interaction. (C) One of the molecules identified through the pharmacophore model superimposed on the structure of compstatin. (D) The same molecule as in (C) docked on C3c.

solubility and was found to retain inhibitory activity[26,27]. The addition of the R(-1) side chain introduced an ionic interaction with E372 of C3c (Figure 8.3A), which contributed in the binding affinity and stability of the C3/RSI-compstatin complex[26,27]. The hydrophobic character of V3 was included as a pharmacophore feature, as V3 inserts into a hydrophobic subcavity in C3c (Figure 8.3A). The amino acids W4 and W7 were observed in the MD trajectory to participate in highly conserved hydrogen bonds with C3c (Appendix S3 of ref[51]) and as a result, hydrogen bond donor and acceptor features at their corresponding positions were included in pharmacophore models. Additionally, the aromatic and hydrophobic properties of W4 and W7 were utilized as pharmacophore features due to the pervasiveness of these features in druglike molecules and their role in favorable interactions with C3c (Figure 8.3A). Amino acids Q5, A9, and H10 are participating in hydrogen bonds (Appendix S3 of ref[51]) and were included as pharmacophore features. The specific locations of A9 and H10 elongate the pharmacophore models, therefore increasing the screening diversity.

On the other hand, specific features or combinations of features were identified to be either too lenient as screening parameters or unlikely to exist in drug molecules. For example, pharmacophores that included more than one hydrophobic feature resulted in too many molecules matched. Features such as the hydrogen bond donor/acceptor capability of D6 in compstatin were found to result in positive hits ($\sim$10,000) during the pharmacophore screen but did not result in viable predicted binding energies (> -6.5 kcal/mol) during docking.

After screening, docking and scoring, features identified to result in potentially viable molecules were used to iteratively improve upon the initial pharmacophore mod-

170
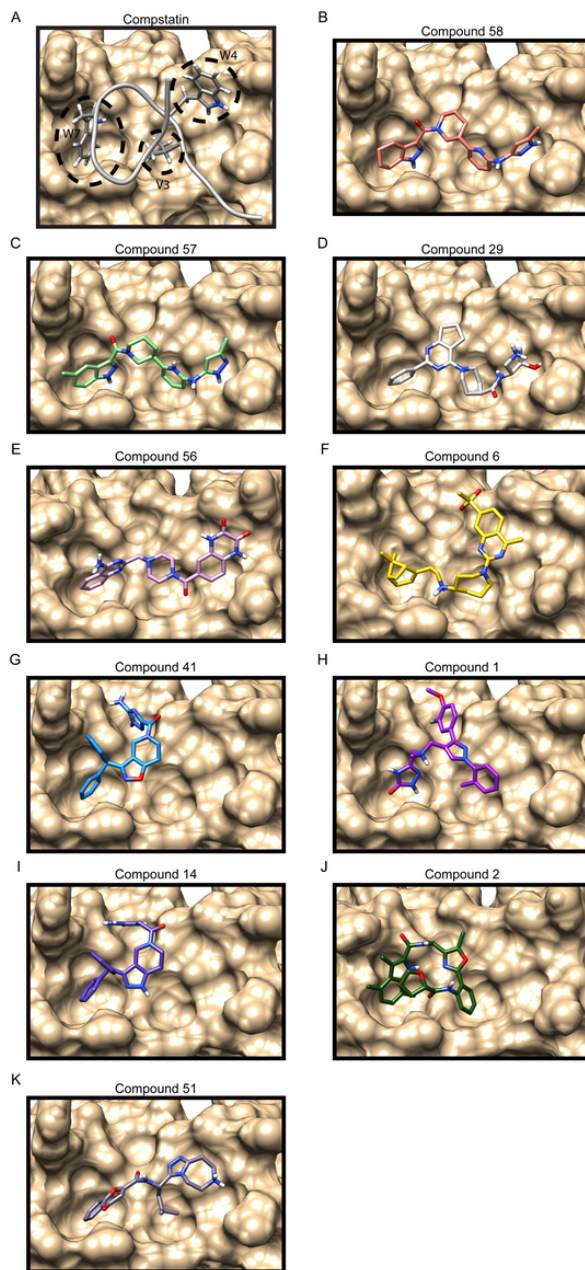
Figure 8.3: Molecular graphics of (A) compstatin bound to C3c and (B–K) the top 10 compounds (of the selected 58 from the primary approach) docked to the corresponding representative conformation of C3c. In (A), three main features of compstatin, V3, W4, and W7, important to the interaction with C3c are identified with dashed circles.

els, resulting in 40 new pharmacophore models. Upon screening the conformers generated from the Drugs Now subset of ZINC and docking the resulted hits to C3c, we identified ~81,000 docked conformer poses. We refined the list of molecules by applying a threshold value of <-7 kcal/mol in predicted binding energies and a $\log P$ threshold of 5. In a few cases, molecules exhibiting $\log P$ values >5 but significantly favorable predicted binding energies were considered as well. Further filtering was performed by evaluating the molecules using Lipinski's rule of five and visual inspection for physicochemical and geometric properties. A final list of 167 molecules were identified (Appendix S4 of ref[51]), out of which 58 (Appendices S5 and S6 of ref[51]) were purchased and tested experimentally. Figure 8.3B–K shows the top 10 compounds out of the selected 58 and displays the variety in physicochemical properties and spatial placement when docked. The top 10 compounds are ZINC72382898 (compound 58), ZINC72382894 (compound 57), ZINC67742743 (compound 29), ZINC29862046 (compound 56), ZINC14995377 (compound 6), ZINC67881194 (compound 41), ZINC12000754 (compound 1), ZINC67605047 (compound 14), ZINC12079160 (compound 2), and ZINC67974289 (compound 51). Compstatin has three main features of importance to the interaction with C3c: the V3 interaction with a smaller hydrophobic subcavity, the W4 interaction with a steric wall-like structural feature, and the W7 interaction with a larger hydrophobic subcavity (Figure 8.2A). A few of the top 10 compounds (compounds 56, 6, and 51) match all three of these features, whereas the other compounds match at least one of these features. The other compounds in the selected 58 (Appendix S7 of ref[51]) also match at least one of the main features of compstatin, except for two compounds (compounds 28 and 16) that were docked to another region.

8.3.2 Experimental Validation and Analysis

Hemolytic Assay

A standard rabbit erythrocyte assay was used to initiate complement activation and to evaluate possible inhibitory effects of the virtual screening chemical compounds. In this assay, complement, as part of normal human serum, is activated by the foreign rabbit erythrocyte cells, resulting in cell lysis by the membrane attack complex. Addition of chemical compounds with potential inhibitory activities would result in decreased lysis, compared with experiments without inhibitors, as shown before in the case of compstatin analogues[26–28].

Despite demonstrating good predicted binding properties, none of the 58 selected compounds from the primary approach and the 6 selected compounds from the alternative approach exhibited inhibitory activity in the hemolytic assays (Figure 8.4). Compstatin analogues are fairly long, cyclic, and bulky peptides (13-15 amino acids of 1500–1900 Da molecular weight), so it may be unlikely for a single druglike compound ($<$500 Da molecular weight) to retain the needed intermolecular contacts necessary for inhibition. It is possible, however, that the identified compounds are capable of binding to C3c, without demonstrating inhibition of hemolytic activity. In such a case, possible combination of compounds using chemical synthesis methods could result in a larger molecule with potential binding affinity for C3c and complement inhibitory activity. Also, compound 29, and perhaps additional experimentally untested compounds from the library, may be used as a scaffold to design new and larger molecules with favorable binding and inhibitory properties.

173
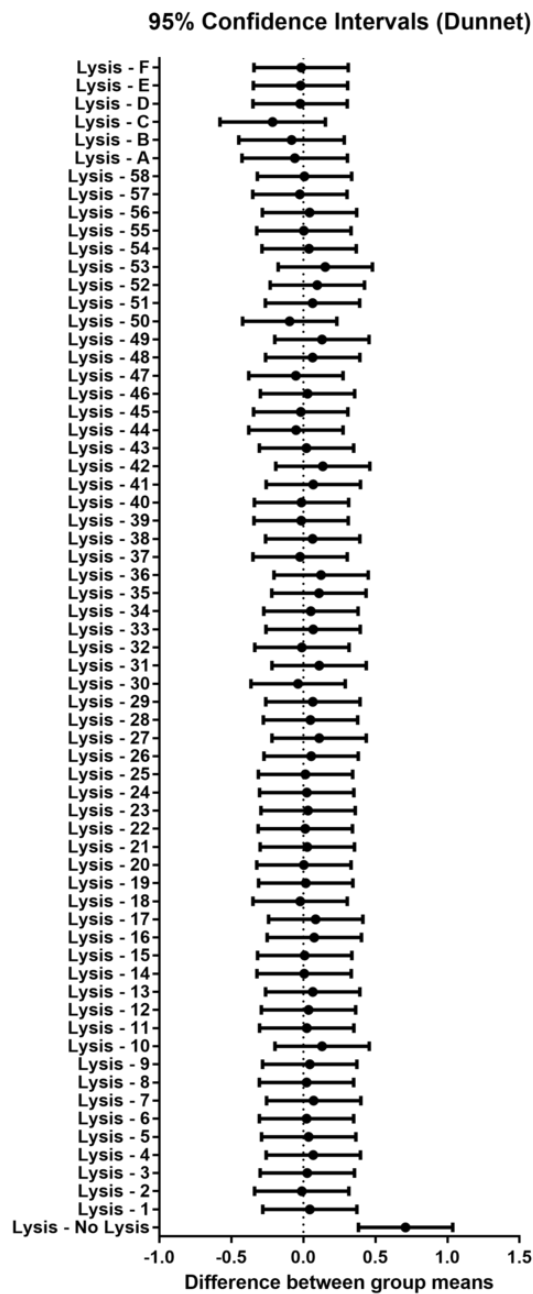
Figure 8.4: Evaluation of hemolytic assay results. The results of the hemolytic assays are represented as a comparison of means using one-way analysis of variance between the controls and the drug compounds. The bottom row of the figure shows the comparison between the two controls for lysis (averaged from both types of positive controls for lysis) and no lysis (see Methods), demonstrating a significant difference in means. The other rows show that the difference between lysis and the inhibitory effects of the compounds is negligible. The means are for data acquired through triplicate hemolytic assays.

174

Microscale Thermophoresis Binding Assay

We measured direct binding of the top 10 compounds identified in the primary approach and the top 6 compounds identified in the secondary approach 2 to C3c using the microscale thermophoresis assay, as described in Methods. Of the selected compounds, only compound 29 demonstrated binding to C3c. Figure 8.5 shows competitive binding of compound 29 to the C3c–competition peptide complex, where the competition peptide was labeled with the fluorophore Cy5 for detection of the thermophoresis signal. This competitive binding experiment shows that compound 29 binds to C3c, albeit with a low affinity. An accurate measurement K_D was not possible because the dose–response binding curve requires additional data points to reach a plateau at higher concentrations, and this was limited by the stock concentration of the compound.
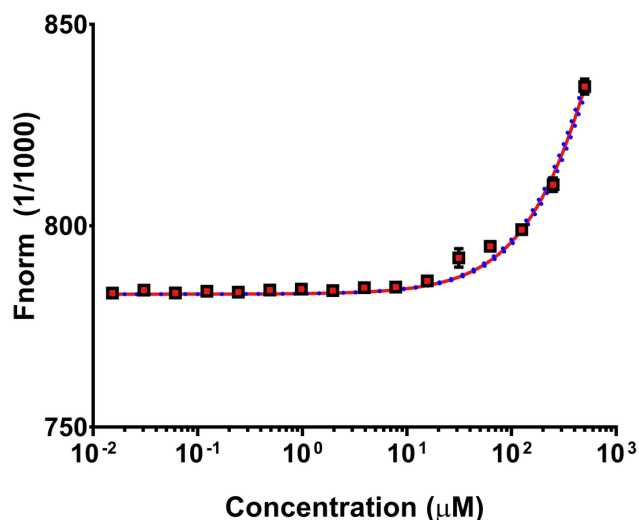
Figure 8.5: Concentration-dependent binding curve of compound 29 to C3c in competition with the competition peptide. Thermophoretic data is plotted as mean ± standard error of the mean (as error bars) from three replicate experiments, together with the fitted binding curve in red and the 95% confidence interval of the fitted binding curve represented as blue dots.

8.4 Conclusions

In this manuscript, we provide a library of 274 molecules that were identified by pharmacophore-based virtual screening and were computationally predicted to have binding affinities for the compstatin binding site of C3c. These molecules comply with pharmacophore properties that are derived from the binding characteristics of compstatin with C3c and are expected to follow a compstatin-like mechanism of binding to C3. A subset of 64 chemical compounds from these 274 molecules was experimentally tested using hemolytic assays and although they were not found to have inhibitory effects, compound 29 demonstrated binding to C3c, thus establishing a proof-of-concept for this methodology.

Despite complement being implicated in various autoimmune and inflammatory diseases, very few complement-targeted therapeutics are currently in the clinic, such as Cinryze (Shire Plc)[54] and Soliris (Alexion Pharmaceuticals, Inc.)[55]. Both of them are protein-based therapeutics, and they are two of the most expensive drugs in the market, an aspect that makes them cost-prohibitive for most patients. Protein-based therapeutics are typically administered intravenously and often suffer from low stability and bioavailability and high production costs. On the other hand, low-molecular-mass chemical compound therapeutics are typically orally available, have better stability and bioavailability properties, and are less costly for industrial production. Although the virtual screening chemical compounds selected for experimental evaluation in this study did not demonstrate complement inhibitory activity, it is possible that several of them retain binding capabilities. Future binding and inhibitory activity studies are needed to further explore the entirety of the library of 274 virtual screening molecules. If binding can be experimentally observed in

future studies, chemical combination of two or more C3-binding compounds may be suffi-cient to reproduce the binding contacts of the long and bulky compstatin family peptides and produce a molecule with compstatin-like complement inhibitory activity.

8.5   References

[1] Walport, M. J. Advances in Immunology: Complement (First of Two Parts). *New England Journal of Medicine*, 344(14):1058–1066, April 2001. ISSN 0028-4793. doi: 10.1056/NEJM200104053441406. 02220.

[2] Walport, M. J. Advances in Immunology: Complement (Second of Two Parts). *New England Journal of Medicine*, 344(15):1140–1144, April 2001. ISSN 0028-4793. doi: 10.1056/NEJM200104123441506.

[3] Merle, N. S., Church, S. E., Fremeaux-Bacchi, V., and Roumenina, L. T. Complement system part I – molecular mechanisms of activation and regulation. *Front. Immunol.*, page 262, 2015. doi: 10.3389/fimmu.2015.00262. 00057.

[4] Merle, N. S., Noe, R., Halbwachs-Mecarelli, L., Fremeaux-Bacchi, V., and Roumenina, L. T. Complement system part II: Role in immunity. *Front. Immunol.*, 6:257, 2015. doi: 10.3389/fimmu.2015.00257. 00030.

[5] Freeley, S., Kemper, C., and Le Friec, G. The "ins and outs" of complement-driven immune responses. *Immunol Rev*, 274(1):16–32, November 2016. ISSN 1600-065X. doi: 10.1111/imr.12472.

[6] Zewde, N., Jr, R. D. G., Dorado, A., and Morikis, D. Quantitative Modeling of the Alternative Pathway of the Complement System. *PLOS ONE*, 11(3):e0152337, March 2016. ISSN 1932-6203. doi: 10.1371/journal.pone.0152337. 00000.

[7] Zewde, N. and Morikis, D. A Computational Model for the Evaluation of Complement System Regulation under Homeostasis, Disease, and Drug Intervention. *bioRxiv*, page 225029, November 2017. doi: 10.1101/225029.

[8] Ricklin, D. and Lambris, J. D. Complement in Immune and Inflammatory Disorders: Pathophysiological Mechanisms. *The Journal of Immunology*, 190(8):3831–3838, April 2013. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.1203487. 00107.

[9] Morgan, B. P. and Harris, C. L. Complement, a target for therapy in inflammatory and degenerative diseases. *Nature Reviews Drug Discovery*, 14(12):857–877, October 2015. ISSN 1474-1776, 1474-1784. doi: 10.1038/nrd4657. 00002.

[10] Ricklin, D. and Lambris, J. D. Complement in Immune and Inflammatory Disorders: Therapeutic Interventions. *J Immunol*, 190(8):3839–3847, April 2013. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.1203200. 00094.

[11] Morikis, D., Assa-Munt, N., Sahu, A., and Lambris, J. D. Solution structure of Compstatin, a potent complement inhibitor. *Protein science: a publication of the Protein Society*, 7(3):619, 1998.

[12] Sahu, A., Soulika, A. M., Morikis, D., Spruce, L., Moore, W. T., and Lambris, J. D. Binding Kinetics, Structure-Activity Relationship, and Biotransformation of the Complement Inhibitor Compstatin. *The Journal of Immunology*, 165(5):2491–2499, September 2000. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.165.5.2491. 00115.

[13] Sahu, A., Kay, B. K., and Lambris, J. D. Inhibition of human complement by a C3-binding peptide isolated from a phage-displayed random peptide library. *J Immunol*, 157(2):884–891, July 1996. ISSN 0022-1767, 1550-6606. 00219.

[14] Ricklin, D. and Lambris, J. D. Compstatin: A Complement Inhibitor on its Way to Clinical Application. *Adv Exp Med Biol*, 632:273–292, 2008. ISSN 0065-2598. 00134.

[15] Mastellos, D. C., Yancopoulou, D., Kokkinos, P., Huber-Lang, M., Hajishengallis, G., Biglarnia, A. R., Lupu, F., Nilsson, B., Risitano, A. M., Ricklin, D., and Lambris, J. D. Compstatin: A C3-targeted complement inhibitor reaching its prime for bedside intervention. *Eur J Clin Invest*, 45(4):423–440, April 2015. ISSN 1365-2362. doi: 10.1111/eci.12419. 00032.

[16] Morikis, D. and Lambris, J. Structure, Dynamics, Activity, and Function of Compstatin and Design of More Potent Analogues. In *Structural Biology of the Complement System*, pages 317–340. CRC Press, March 2005. ISBN 978-0-8247-2540-2. 00000.

[17] Morikis, D., Soulika, A. M., Mallik, B., Klepeis, J. L., Floudas, C. A., and Lambris, J. D. Improvement of the anti-C3 activity of compstatin using rational and combinatorial approaches. *Biochemical Society Transactions*, 32(1):28–32, 2004.

[18] Morikis, D., Floudas, C. A., and Lambris, J. D. Structure-Based Integrative Computational and Experimental Approach for the Optimization of Drug Design. In Sunderam, V. S., van Albada, G. D., Sloot, P. M. A., and Dongarra, J. J., editors, *Computational Science – ICCS 2005*, Lecture Notes in Computer Science, pages 680–688. Springer Berlin Heidelberg, 2005. ISBN 978-3-540-32114-9.

[19] Morikis, D. and Lambris, J. D. Structural aspects and design of low-molecular-mass complement inhibitors. *Biochemical Society Transactions*, 30(6):1026–1036, November 2002. ISSN 0300-5127, 1470-8752. doi: 10.1042/bst0301026. 00060.

[20] Morikis, D., Roy, M., Sahu, A., Troganis, A., Jennings, P. A., Tsokos, G. C., and Lambris, J. D. The Structural Basis of Compstatin Activity Examined by Structure-Function-based Design of Peptide Analogs and NMR. *J. Biol. Chem.*, 277(17):14942–14953, April 2002. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.M200021200. 00060.

[21] Soulika, A. M., Morikis, D., Sarrias, M.-R., Roy, M., Spruce, L. A., Sahu, A., and Lambris, J. D. Studies of Structure-Activity Relations of Complement Inhibitor Compstatin. *The Journal of Immunology*, 171(4):1881–1890, August 2003. ISSN 0022-1767, 1550-6606. doi: 10.4049/jimmunol.171.4.1881. 00044.

[22] Janssen, B. J. C., Halff, E. F., Lambris, J. D., and Gros, P. Structure of Compstatin in Complex with Complement Component C3c Reveals a New Mechanism of Complement Inhibition. *J. Biol. Chem.*, 282(40):29241–29247, May 2007. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.M704587200.

[23] Bellows, M., Fung, H., Taylor, M., Floudas, C., López de Victoria, A., and Morikis, D. New Compstatin Variants through Two De Novo Protein Design Frameworks. *Biophys J*, 98(10):2337–2346, May 2010. ISSN 0006-3495. doi: 10.1016/j.bpj.2010.01.057. 00034.

[24] Tamamis, P., López de Victoria, A., Gorham, R. D., Bellows-Peterson, M. L., Pierou, P., Floudas, C. A., Morikis, D., and Archontis, G. Molecular Dynamics in Drug Design: New Generations of Compstatin Analogs: New Generations of Compstatin Analogs. *Chemical Biology & Drug Design*, 79(5):703–718, May 2012. ISSN 17470277. doi: 10.1111/j.1747-0285.2012.01324.x.

[25] López de Victoria, A., Gorham, R. D., Bellows-Peterson, M. L., Ling, J., Lo, D. D., Floudas, C. A., and Morikis, D. A New Generation of Potent Complement Inhibitors of the Compstatin Family. *Chemical Biology & Drug Design*, 77(6):431–440, June 2011. ISSN 1747-0285. doi: 10.1111/j.1747-0285.2011.01111.x. 00021.

[26] Gorham, R. D., Forest, D. L., Khoury, G. A., Smadbeck, J., Beecher, C. N., Healy, E. D., Tamamis, P., Archontis, G., Larive, C. K., Floudas, C. A., Radeke, M. J., Johnson, L. V., and Morikis, D. New Compstatin Peptides Containing N-Terminal Extensions and Non-Natural Amino Acids Exhibit Potent Complement Inhibition and Improved Solubility Characteristics. *J. Med. Chem.*, 58(2):814–826, January 2015. ISSN 0022-2623. doi: 10.1021/jm501345y. 00002.

[27] Gorham, R. D., Forest, D. L., Tamamis, P., López de Victoria, A., Kraszni, M., Kieslich, C. A., Banna, C. D., Bellows-Peterson, M. L., Larive, C. K., Floudas, C. A., Archontis, G., Johnson, L. V., and Morikis, D. Novel compstatin family peptides inhibit complement activation by drusen-like deposits in human retinal pigmented epithelial cell cultures. *Experimental Eye Research*, 116:96–108, November 2013. ISSN 00144835. doi: 10.1016/j.exer.2013.07.023. 00009.

[28] Mohan, R. R., Cabrera, A. P., Harrison, R. E., Gorham, R. D., Johnson, L. V., Ghosh, K., and Morikis, D. Peptide redesign for inhibition of the complement system: Targeting age-related macular degeneration. *Molecular Vision*, 22:1280–1290, 2016. 00000.

[29] Lavecchia, A. and Di Giovanni, C. Virtual screening strategies in drug discovery: A critical review. *Current medicinal chemistry*, 20(23):2839–2860, 2013. 00133.

[30] Loving, K., Alberts, I., and Sherman, W. Computational approaches for fragment-based and de novo design. *Current topics in medicinal chemistry*, 10(1):14–32, 2010. 00086.

[31] Zhou, P., Wang, C., Ren, Y., Yang, C., and Tian, F. Computational peptidology: A new and promising approach to therapeutic peptide design. *Current medicinal chemistry*, 20(15):1985–1996, 2013. 00073.

[32] Mallik, B., Katragadda, M., Spruce, L. A., Carafides, C., Tsokos, C. G., Morikis, D., and Lambris, J. D. Design and NMR Characterization of Active Analogues of Compstatin Containing Non-Natural Amino Acids. *J. Med. Chem.*, 48(1):274–286, January 2005. ISSN 0022-2623. doi: 10.1021/jm0495531. 00072.

[33] Klepeis, J. L., Floudas, C. A., Morikis, D., Tsokos, C. G., Argyropoulos, E., Spruce, L., and Lambris, J. D. Integrated Computational and Experimental Approach for Lead Optimization and Design of Compstatin Variants with Improved Activity. *J. Am. Chem. Soc.*, 125(28):8422–8423, July 2003. ISSN 0002-7863. doi: 10.1021/ja034846p. 00101.

[34] Song, C. M., Lim, S. J., and Tong, J. C. Recent advances in computer-aided drug design. *Briefings in Bioinformatics*, 10(5):579–591, September 2009. ISSN 1467-5463, 1477-4054. doi: 10.1093/bib/bbp023.

[35] Lionta, E., Spyrou, G., Vassilatis, D. K., and Cournia, Z. Structure-based virtual screening for drug discovery: Principles, applications and recent advances. *Curr Top Med Chem*, 14(16):1923–1938, 2014. ISSN 1873-4294.

[36] Michel, J. Current and emerging opportunities for molecular simulations in structure-based drug design. *Phys. Chem. Chem. Phys.*, 16(10):4465–4477, January 2014. ISSN 1463-9076, 1463-9084. doi: 10.1039/C3CP54164A.

[37] Gorham, R. D., Nuñez, V., Lin, J., Rooijakkers, S. H. M., Vullev, V. I., and Morikis, D. Discovery of Small Molecules for Fluorescent Detection of Complement Activation Product C3d. *J. Med. Chem.*, November 2015. ISSN 0022-2623. doi: 10.1021/acs.jmedchem.5b01062. 00000.

[38] Tamamis, P., Morikis, D., Floudas, C. A., and Archontis, G. Species specificity of the complement inhibitor compstatin investigated by all-atom molecular dynamics simulations. *Proteins*, 78(12):2655–2667, September 2010. ISSN 1097-0134. doi: 10.1002/prot.22780. 00027.

[39] Tamamis, P., Pierou, P., Mytidou, C., Floudas, C. A., Morikis, D., and Archontis, G. Design of a modified mouse protein with ligand binding properties of its human analog by molecular dynamics simulations: The case of C3 inhibition by compstatin. *Proteins: Structure, Function, and Bioinformatics*, 79(11):3166–3179, November 2011. ISSN 08873585. doi: 10.1002/prot.23149. 00016.

[40] Grant, B. J., Rodrigues, A. P. C., ElSawy, K. M., McCammon, J. A., and Caves, L. S. D. Bio3d: An R package for the comparative analysis of protein structures. *Bioinformatics*, 22(21):2695–2696, January 2006. ISSN 1367-4803, 1460-2059. doi: 10.1093/bioinformatics/btl461.

[41] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. UCSF Chimera?A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, October 2004. ISSN 0192-8651, 1096-987X. doi: 10.1002/jcc.20084.

[42] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.

[43] McGibbon, R. T., Beauchamp, K. A., Harrigan, M. P., Klein, C., Swails, J. M., Hernández, C. X., Schwantes, C. R., Wang, L.-P., Lane, T. J., and Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal*, 109(8):1528–1532, October 2015. ISSN 0006-3495. doi: 10.1016/j.bpj.2015.08.015.

[44] Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., and Coleman, R. G. ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.*, 52(7):1757–1768, July 2012. ISSN 1549-9596. doi: 10.1021/ci3001277.

[45] Koes, D. R. and Camacho, C. J. ZINCPharmer: Pharmacophore search of the ZINC database. *Nucl Acids Res*, 40(W1):W409–W414, July 2012. ISSN 0305-1048. doi: 10.1093/nar/gks378.

[46] Dixon, S. L., Smondyrev, A. M., and Rao, S. N. PHASE: A Novel Approach to Pharmacophore Modeling and 3D Database Searching. *Chemical Biology & Drug Design*, 67(5):370–372, May 2006. ISSN 1747-0285. doi: 10.1111/j.1747-0285.2006.00384.x.

[47] Dixon, S. L., Smondyrev, A. M., Knoll, E. H., Rao, S. N., Shaw, D. E., and Friesner, R. A. PHASE: A new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J Comput Aided Mol Des*, 20(10-11):647–671, October 2006. ISSN 0920-654X, 1573-4951. doi: 10.1007/s10822-006-9087-6.

[48] Trott, O. and Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J Comput Chem*, 31(2):455–461, January 2010. ISSN 0192-8651. doi: 10.1002/jcc.21334. 03932.

[49] Cao, Y., Charisi, A., Cheng, L.-C., Jiang, T., and Girke, T. ChemmineR: A compound mining framework for R. *Bioinformatics*, 24(15):1733–1734, August 2008. ISSN 1367-4803. doi: 10.1093/bioinformatics/btn307. 00078.

[50] Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 23(1–3):3–25, January 1997. ISSN 0169-409X. doi: 10.1016/S0169-409X(96)00423-1.

[51] Mohan, R. R., Wilson, M., Gorham, R. D., Harrison, R. E. S., Morikis, V. A., Kieslich, C. A., Orr, A. A., Coley, A. V., Tamamis, P., and Morikis, D. Virtual Screening of Chemical Compounds for Discovery of Complement C3 Ligands. *ACS Omega*, 3(6): 6427–6438, June 2018. ISSN 2470-1343. doi: 10.1021/acsomega.8b00606.

[52] Seidel, S. A., Dijkman, P. M., Lea, W. A., van den Bogaart, G., Jerabek-Willemsen, M., Lazic, A., Joseph, J. S., Srinivasan, P., Baaske, P., Simeonov, A., Katritch, I., Melo, F. A., Ladbury, J. E., Schreiber, G., Watts, A., Braun, D., and Duhr, S. Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. *Methods*, 59(3):301–315, March 2013. ISSN 10462023. doi: 10.1016/j.ymeth.2012.12.005.

[53] Wienken, C. J., Baaske, P., Rothbauer, U., Braun, D., and Duhr, S. Protein-binding assays in biological liquids using microscale thermophoresis. *Nature Communications*, 1(7):100, October 2010. ISSN 2041-1723. doi: 10.1038/ncomms1093.

[54] Cocchio, C. and Marzella, N. Cinryze, a Human Plasma-Derived C1 Esterase Inhibitor for Prophylaxis Of Hereditary Angioedema. *P T*, 34(6):293–328, June 2009. ISSN 1052-1372.

[55] Kim, J. S., Lee, J. W., Kim, B. K., Lee, J.-H., and Chung, J. The use of the complement inhibitor eculizumab (Soliris®) for treating Korean patients with paroxysmal nocturnal hemoglobinuria. *Korean J Hematol*, 45(4):269–274, December 2010. ISSN 1738-7949. doi: 10.5045/kjh.2010.45.4.269.

Chapter 9

Conclusions

Through a combination of hardware and software advances in computing power and accessibility, computational methods have been shown to complement experimental methods and even capable of accurately reproducing and predicting biological phenomena, provided appropriate parametrization is utilized[1]. Leveraging the synergy of computational and experimental methods allows us to reinforce, validate and efficiently explore solutions involving structure-dynamics-function relations and function inhibition. With the increased popularity of computational methodologies, ensuring reproducibility[2] and following best practices for scientific code (such as version control[3]) is essential to conducting science. Through the application of these principles and combining computational and experimental methodologies, we map the druggability of C3 and its derivatives for inhibitor design and biomarker discovery.

In pursuit of inhibitor and biomarker discovery targeting C3 and its derivatives, we utilized a variety of computational approaches to explore the mechanisms driving the

interaction between complement fragment C3d (a cleavage product of C3) and CR2[4]. Using computational methods such as electrostatic analysis[5], explicit solvent molecular dynamics (MD) simulations, MM-GBSA energetics analysis, and steered molecular dynamics simulations, we identified key intermolecular interactions driving the overall C3d:CR2 complex dynamics. We provide a theoretical basis on the physiological likelihood of one binding mode[6] over another[7] and provided insight into the role of nonphysiological zinc ions in crystal packing energetics. Given the role of C3d:CR2 in innate and adaptive immunity conferral, it is crucial to have a clear understanding of the underlying mechanisms of the interaction. Additionally, through our evaluation of the structure suggested to be an artifact of crystallization[7], we emphasize the importance of understanding the importance of proper parametrization and evaluation of inputs (which has implications for both experimental and computational work) while clarifying the mechanistic properties of the structure.

To further explore the significance of electrostatics in the C3d:CR2 interaction, we incorporated Brownian Dynamics (BD) simulations to evaluate the role of electrostatic steering in the formation of the encounter complex[8]. We establish the ionic-strength dependence of the C3d:CR2 interaction thus establishing the role of electrostatic steering. Additionally, we validate the importance of specific intermolecular interactions identified previously[4] through a combination of computational mutagenesis and BD simualations. These results are particularly interesting as the presence of an acidic hotspot at the CR2-binding site of C3d lead us to expect destabilization of local structure due to concentration of like-charge amino acids. However, evolution favors speed[9] and function over stability[10] and that is exemplified here. Through these two studies, we have thoroughly character-

185

ized the C3d:CR2 interaction which allows us to leverage these insights for for C3d-binding biomarker discovery through a two-pronged approach: virtual screening of small molecules and peptide design.

C3d's role as an opsonin in complement response makes it amenable as a means of tracking disease progression, specifically in the context inflammatory and autoimmune diseases where complement response is implicated in their pathology. Although there have been recent efforts to target C3d as a marker for complement-induced disease progression[11,12], these efforts have been limited to the development biopharmaceuticals which suffer from a host of limitations (bioavailability, scalability, etc.). Small molecules allow for an alternative approach as they issues in bioavailability while remaining cost-effective in production and capable of delivery through different mediums. Building off of the success of a prior study[13], we utilize a virtual screening workflow to develop robust pharmacophore models and implement exhaustive conformer generation for a thorough pharmacophore screening process. Subsequent to rigorous molecular docking with a representative conformational ensemble, we identified a subset of small molecules exhibiting physicochemical properties of interest. Our choice of specific pharmacophore features and models appear to have given us the intended effect, for example, the inclusion of a hydrophobic feature in the cavity of the CR2-binding site of C3d resulted in several small molecules docked deep into the cavity with high predicted binding affinities. To verify the biological significance of these theoretical results, further experimental testing to characterize activity and binding is needed.

Peptides serve as the middle ground to biopharmaceuticals and small molecules with their scalability in production, specificity and oral delivery capability. Leveraging

information gleaned from prior studies on the C3d:CR2 interaction[4,8] and computational methodologies, we explore peptidic design for the development of potential C3d biomarkers for theranostics. During initial rational design, we aimed to identify CR2-based peptides retaining β-turns from CR2 to leverage structural stability and further rational design focused on optimizing solubility of the peptides. Following the identification of two initial peptides exhibiting binding to C3d, we explored additional optimization and design approaches. In order to optimally and efficiently explore the mutation space possible with peptides of 20 - 22 amino acids in length, a combinatorial design approach was incorporated using a Rosetta tool[14,15]. This allowed us to improve the K_D of our peptides by an order of magnitude. To evaluate the capabilities of these peptides as potential theranostics, we plan to attach fluorophores and characterize binding and fluorescence. It is also possible that further optimization of binding may be needed as well. Through a combination of the peptidic design approach and virtual screening of small molecules approach, we hope to provide a basis for the development of theranostics and potential inhibitors targeting C3d and complement response.

To explore another facet of the mechanistic properties and structural dynamics of C3d, we investigated the C3d:CR3 interaction as well using computational methodologies as previously outlined. We characterized significant intermolecular interactions and evaluated the importance of the Mg^{2+} ion at the interface as well. Dissociative stability of the complex was evaluated using steered MD simulations as well. The close proximity of the CR2 and CR3 binding sites in addition to the lack of overlap in the binding sites suggests a potential for peptide design targeting both the CR2 and CR3 binding sites[16].

We also applied the two prong (peptidic design and virtual screening) approach to C3, the point of convergence for all three pathways of complement activation[17]. Building on our expertise of compstatin and its analogs[18–23], we improved solubility and retained binding capabilities through the development of novel compstatin analogs[24]. The top performing peptide (which incorporates PEGylation into its design) demonstrated improved solubility over the prior optimal compstatin analog while retaining its binding capabilities. Although PEGylation is a common measure for improving solubility, the application of a shorter multimer of 8-PEG block is not common and we believe this chemical moiety allows for binding of compstatin to C3 without steric hindrance while also creating a hydration shell through the PEG block. We also verified the activity of these peptides in cell-based assays through experimental assays with our collaborators[24]. Given the role of complement activation in the pathology of age-related macular degeneration (AMD)[25–28], our collaborators' work verified the efficacy of the PEGylated peptide in inhibiting activation in a human RPE cell-based assay that mimics AMD pathophysiological conditions.

We also incorporated virtual screening methodologies as a means to identify C3 ligands through the generation of pharmocophore models, molecular docking, scoring and filtering, and experimental evaluation[29]. These methodologies were informed by prior molecular dynamics studies[30,31] and intermolecular interaction analysis as well. We also worked with colloborators to explore additional virtual screening workflows to ensure a variety of small molecules were identified. Our approach yielded at least one molecule capable of binding (albeit weakly) but further testing of hits from docking as well as iterative optimization is needed as only a small subset of molecules were experimentally characterized. It may also

be prudent to investigate fragment-based screening as there is a yet to be fully explored virtual chemical library[32] to be explored but limitations of computing power, time, and experimental expertise persist.

In summary we utilized a variety of computational and experimental methodologies to explore the mechanistic and dynamic properties of C3 and its derivatives so that we can map their druggability. Using electrostatic analysis and molecular and brownian dynamics simulations, we characterized the dynamics and structural stability of the C3d:CR2 interaction and established the role of electrostatic steering in the acceleration of the interaction. We leveraged the insights gleaned from study of the C3d:CR2 interaction for the identification of potential C3d-binding biomarkers through virtual screening of small molecules and peptide design. We provided additional insight into the structural dynamics of C3d through the evaluation of the C3d:CR3 complex using molecular dynamics simulations and steered molecular dynamics simulations. We applied the two pronged approach to biomarker and inhibitor discovery in the context of C3 as well and identified potential candidates through peptide redesign and virtual drug screening. Through these studies we have helped characterize the druggability of C3 and its derivatives and provide a basis for the development of potential inhibitors and biomarkers for therapeutic and theranostic applications.

9.1 References

[1] Houk, K. N. and Liu, F. Holy Grails for Computational Organic Chemistry and Biochemistry. *Acc. Chem. Res.*, 50(3):539–543, March 2017. ISSN 0001-4842. doi: 10.1021/acs.accounts.6b00532.

[2] Peng, R. D. Reproducible Research in Computational Science. *Science*, 334(6060): 1226–1227, December 2011. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1213847.

[3] Blischak, J. D., Davenport, E. R., and Wilson, G. A Quick Introduction to Version Control with Git and GitHub. *PLOS Computational Biology*, 12(1):e1004668, January 2016. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1004668.

[4] Mohan, R. R., Gorham Jr., R. D., and Morikis, D. A theoretical view of the C3d:CR2 binding controversy. *Molecular Immunology*, 64(1):112–122, March 2015. ISSN 0161-5890. doi: 10.1016/j.molimm.2014.11.006. 00002.

[5] Harrison, R. E. S., Mohan, R. R., Gorham, R. D., Kieslich, C. A., and Morikis, D. AESOP: A Python Library for Investigating Electrostatics in Protein Interactions. *Biophysical Journal*, 112(9):1761–1766, May 2017. ISSN 0006-3495. doi: 10.1016/j. bpj.2017.04.005.

[6] van den Elsen, J. M. H. and Isenman, D. E. A Crystal Structure of the Complex Between Human Complement Receptor 2 and Its Ligand C3d. *Science*, 332(6029): 608–611, April 2011. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1201954.

[7] Szakonyi, G., Guthridge, J. M., Li, D., Young, K., Holers, V. M., and Chen, X. S. Structure of Complement Receptor 2 in Complex with Its C3d Ligand. *Science*, 292 (5522):1725–1728, January 2001. ISSN 0036-8075, 1095-9203. doi: 10.1126/science. 1059118.

[8] Mohan, R. R., Huber, G. A., and Morikis, D. Electrostatic Steering Accelerates C3d:CR2 Association. *J. Phys. Chem. B*, April 2016. ISSN 1520-6106. doi: 10.1021/acs.jpcb.6b02095. 00000.

[9] McCammon, J. A. Darwinian biophysics: Electrostatics and evolution in the kinetics of molecular binding. *PNAS*, 106(19):7683–7684, December 2009. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0902767106. 00015.

[10] Schreiber, G., Buckle, A. M., and Fersht, A. R. Stability and function: Two constraints in the evolution of barstar and other proteins. *Structure*, 2(10):945–951, October 1994. ISSN 0969-2126. doi: 10.1016/S0969-2126(94)00096-4. 00156.

[11] Serkova, N. J., Renner, B., Larsen, B. A., Stoldt, C. R., Hasebroock, K. M., Bradshaw-Pierce, E. L., Holers, V. M., and Thurman, J. M. Renal Inflammation: Targeted Iron Oxide Nanoparticles for Molecular MR Imaging in Mice. *Radiology*, 255(2):517–526, April 2010. ISSN 0033-8419. doi: 10.1148/radiol.09091134. 00041.

[12] Thurman, J. M., Kulik, L., Orth, H., Wong, M., Renner, B., Sargsyan, S. A., Mitchell, L. M., Hourcade, D. E., Hannan, J. P., Kovacs, J. M., Coughlin, B., Woodell, A. S., Pickering, M. C., Rohrer, B., and Holers, V. M. Detection of complement activation using monoclonal antibodies against C3d. *Journal of Clinical Investigation*, 123(5): 2218–2230, May 2013. ISSN 0021-9738. doi: 10.1172/JCI65861. 00022.

[13] Gorham, R. D., Nuñez, V., Lin, J., Rooijakkers, S. H. M., Vullev, V. I., and Morikis, D. Discovery of Small Molecules for Fluorescent Detection of Complement Activation Product C3d. *J. Med. Chem.*, November 2015. ISSN 0022-2623. doi: 10.1021/acs. jmedchem.5b01062. 00000.

[14] Smith, C. A. and Kortemme, T. Structure-Based Prediction of the Peptide Sequence Space Recognized by Natural and Synthetic PDZ Domains. *Journal of Molecular Biology*, 402(2):460–474, September 2010. ISSN 0022-2836. doi: 10.1016/j.jmb.2010. 07.032.

[15] Smith, C. A. and Kortemme, T. Predicting the Tolerated Sequences for Proteins and Protein Interfaces Using RosettaBackrub Flexible Backbone Design. *PLOS ONE*, 6 (7):e20451, July 2011. ISSN 1932-6203. doi: 10.1371/journal.pone.0020451.

[16] Bajic, G., Yatime, L., Sim, R. B., Vorup-Jensen, T., and Andersen, G. R. Structural insight on the recognition of surface-bound opsonins by the integrin I domain of complement receptor 3. *PNAS*, 110(41):16426–16431, October 2013. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1311261110.

[17] Zipfel, P. F. and Skerka, C. Complement: The Alternative Pathway. *eLS*, 2001.

[18] Morikis, D. and Lambris, J. Structure, Dynamics, Activity, and Function of Compstatin and Design of More Potent Analogues. In *Structural Biology of the Complement System*, pages 317–340. CRC Press, March 2005. ISBN 978-0-8247-2540-2. 00000.

[19] Klepeis, J. L., Floudas, C. A., Morikis, D., Tsokos, C. G., Argyropoulos, E., Spruce, L., and Lambris, J. D. Integrated Computational and Experimental Approach for Lead Optimization and Design of Compstatin Variants with Improved Activity. *J. Am. Chem. Soc.*, 125(28):8422–8423, July 2003. ISSN 0002-7863. doi: 10.1021/ja034846p. 00101.

[20] Morikis, D., Soulika, A. M., Mallik, B., Klepeis, J. L., Floudas, C. A., and Lambris, J. D. Improvement of the anti-C3 activity of compstatin using rational and combinatorial approaches. *Biochemical Society Transactions*, 32(1):28–32, 2004.

[21] Mallik, B., Katragadda, M., Spruce, L. A., Carafides, C., Tsokos, C. G., Morikis, D., and Lambris, J. D. Design and NMR Characterization of Active Analogues of Compstatin Containing Non-Natural Amino Acids. *J. Med. Chem.*, 48(1):274–286, January 2005. ISSN 0022-2623. doi: 10.1021/jm0495531. 00072.

[22] Gorham, R. D., Forest, D. L., Tamamis, P., López de Victoria, A., Kraszni, M., Kieslich, C. A., Banna, C. D., Bellows-Peterson, M. L., Larive, C. K., Floudas, C. A., Archontis, G., Johnson, L. V., and Morikis, D. Novel compstatin family peptides inhibit complement activation by drusen-like deposits in human retinal pigmented epithelial cell cultures. *Experimental Eye Research*, 116:96–108, November 2013. ISSN 00144835. doi: 10.1016/j.exer.2013.07.023. 00009.

[23] Gorham, R. D., Forest, D. L., Khoury, G. A., Smadbeck, J., Beecher, C. N., Healy, E. D., Tamamis, P., Archontis, G., Larive, C. K., Floudas, C. A., Radeke, M. J., Johnson, L. V., and Morikis, D. New Compstatin Peptides Containing N-Terminal Extensions and Non-Natural Amino Acids Exhibit Potent Complement Inhibition and Improved Solubility Characteristics. *J. Med. Chem.*, 58(2):814–826, January 2015. ISSN 0022-2623. doi: 10.1021/jm501345y. 00002.

[24] Mohan, R. R., Cabrera, A. P., Harrison, R. E., Gorham, R. D., Johnson, L. V., Ghosh, K., and Morikis, D. Peptide redesign for inhibition of the complement system: Targeting age-related macular degeneration. *Molecular Vision*, 22:1280–1290, 2016. 00000.

[25] Liszewski, M. K. and Atkinson, J. P. Complement regulators in human disease: Lessons from modern genetics. *Journal of Internal Medicine*, 277(3):294–305, March 2015. ISSN 1365-2796. doi: 10.1111/joim.12338.

[26] Schramm, E. C., Clark, S. J., Triebwasser, M. P., Raychaudhuri, S., Seddon, J. M., and Atkinson, J. P. Genetic variants in the complement system predisposing to age-related macular degeneration: A review. *Molecular Immunology*, 61(2):118–125, October 2014. ISSN 01615890. doi: 10.1016/j.molimm.2014.06.032. 00029.

[27] Black, J. R. M. and Clark, S. J. Age-related macular degeneration: Genome-wide association studies to translation. *Genet Med*, 18(4):283–289, April 2016. ISSN 1098-3600. doi: 10.1038/gim.2015.70.

[28] Anderson, D. H., Radeke, M. J., Gallo, N. B., Chapin, E. A., Johnson, P. T., Curletti, C. R., Hancox, L. S., Hu, J., Ebright, J. N., Malek, G., Hauser, M. A., Rickman, C. B., Bok, D., Hageman, G. S., and Johnson, L. V. The Pivotal Role of the Complement System in Aging and Age-related Macular Degeneration: Hypothesis Re-visited. *Prog Retin Eye Res*, 29(2):95–112, March 2010. ISSN 1350-9462. doi: 10.1016/j.preteyeres.2009.11.003.

[29] Mohan, R. R., Wilson, M., Gorham, R. D., Harrison, R. E. S., Morikis, V. A., Kieslich, C. A., Orr, A. A., Coley, A. V., Tamamis, P., and Morikis, D. Virtual Screening of Chemical Compounds for Discovery of Complement C3 Ligands. *ACS Omega*, 3(6):6427–6438, June 2018. ISSN 2470-1343. doi: 10.1021/acsomega.8b00606.

[30] Tamamis, P., Morikis, D., Floudas, C. A., and Archontis, G. Species specificity of the complement inhibitor compstatin investigated by all-atom molecular dynam-

ics simulations. *Proteins*, 78(12):2655–2667, September 2010. ISSN 1097-0134. doi: 10.1002/prot.22780. 00027.

[31] Tamamis, P., López de Victoria, A., Gorham, R. D., Bellows-Peterson, M. L., Pierou, P., Floudas, C. A., Morikis, D., and Archontis, G. Molecular Dynamics in Drug Design: New Generations of Compstatin Analogs: New Generations of Compstatin Analogs. *Chemical Biology & Drug Design*, 79(5):703–718, May 2012. ISSN 17470277. doi: 10.1111/j.1747-0285.2012.01324.x.

[32] Walters, W. P. Virtual Chemical Libraries. *J. Med. Chem.*, August 2018. ISSN 0022-2623. doi: 10.1021/acs.jmedchem.8b01048.

Appendices

Appendix A

Supporting Information for

Chapter 2
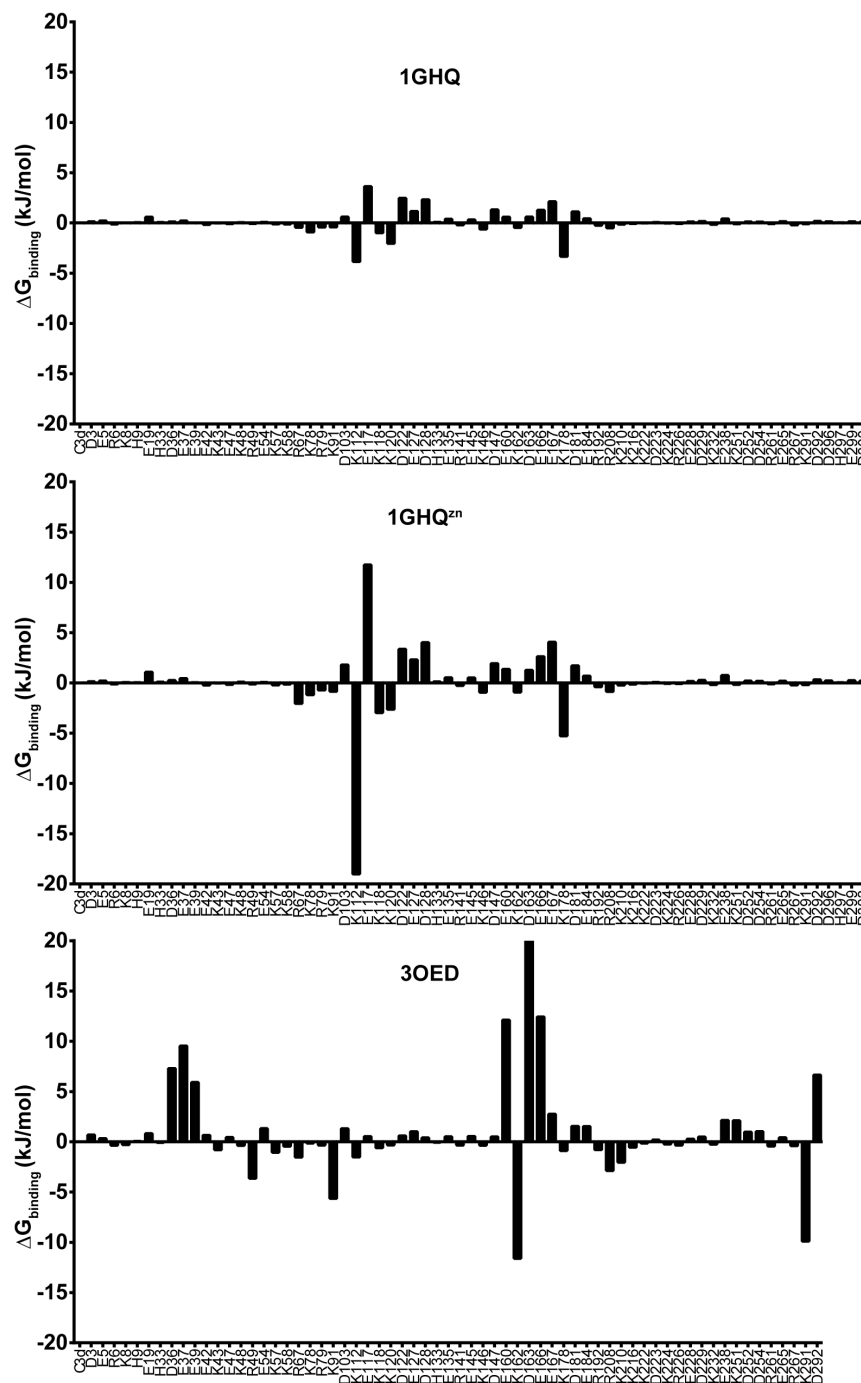
Figure A.1: Results of electrostatic analysis using mutants generated by a computational alanine scan of all ionizable amino acids on C3d. Changes in free energies of binding upon mutation are calculated using Equation 2.3. A positive value denotes loss of binding, indicating that the mutated residue favors binding. A negative value denotes gain of binding, indicating that the mutated residue opposes binding.
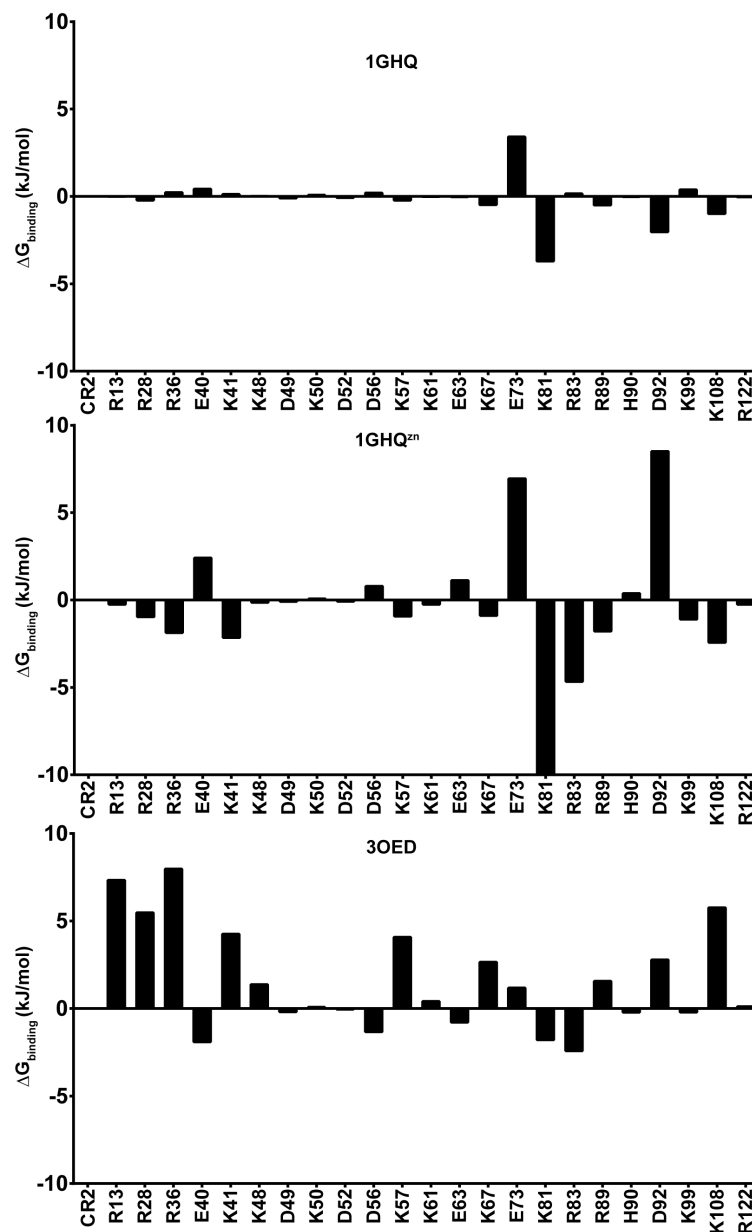
Figure A.2: Results of electrostatic analysis using mutants generated by a computational alanine scan of all ionizable amino acids on CR2. Changes in free energies of binding upon mutation are calculated using Equation 2.3. A positive value denotes loss of binding, indicating that the mutated residue favors binding. A negative value denotes gain of binding, indicating that the mutated residue opposes binding.

Appendix B

Supporting Information for

Chapter 7
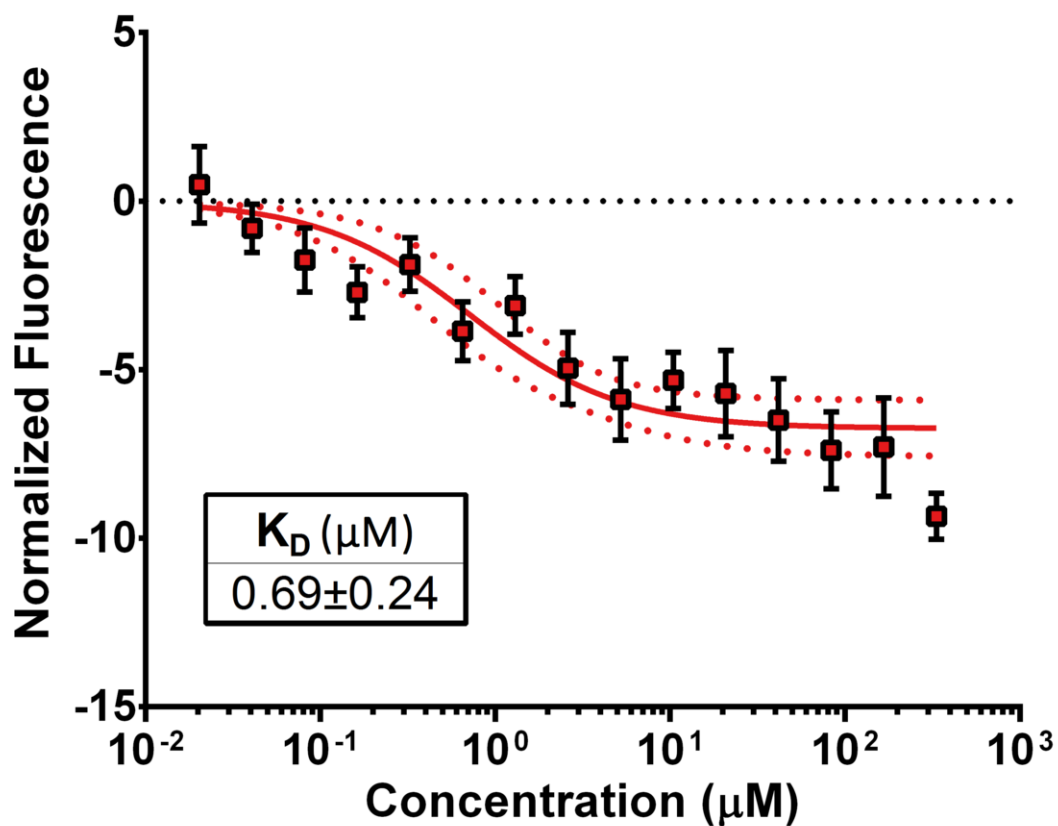
Figure B.1: Concentration dependent binding curve of Peptide 2 to C3c in competition with the Competition Peptide. The labeled form of the Competition Peptide was used. Thermophoretic data is plotted as mean $\pm$ SEM from 7 replicate experiments, together with the fitted binding curve. The dotted lines in red indicate the 95% confidence interval of the fitted binding curve. The black dotted line represents the baseline thermophoretic response in the absence of binding. The inset table indicates the $K_D \pm$ SEM.
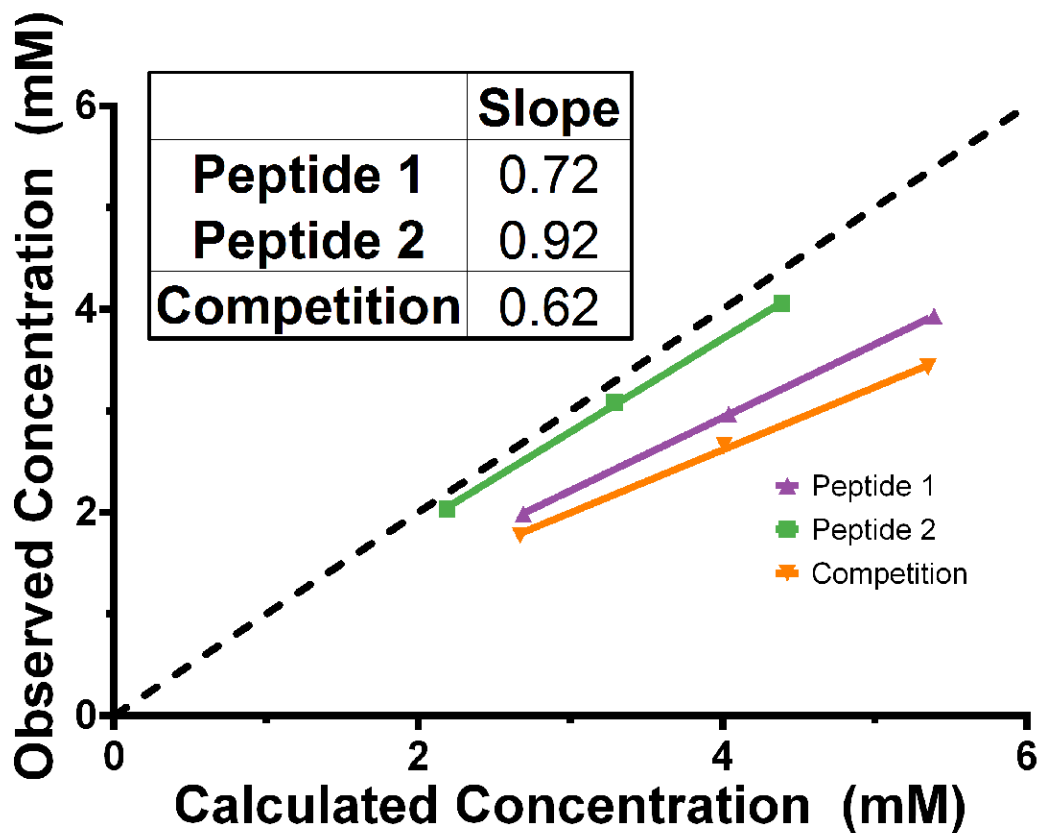
Figure B.2: Correlation of the observed concentration with the calculated concentration, with the concentrations presented in mM. Peptide 2 is the most soluble, as indicated by the high correlation between the observed and calculated concentrations (slope of 0.92). Each data point represents the mean measured concentration and the corresponding expected concentration with a linear regression fit to the data. A straight line of slope 1 passing through the origin is inserted to indicate the closeness of the data to perfect correlation.

Appendix C

AESOP: Analysis of Electrostatic Structures of Proteins

Electrostatics have been shown to play a pivotal role in guiding the association of a protein with its respective binding partner, forming the encounter complex [1–4]. Additionally, electrostatic interactions can act to thermodynamically stabilize a protein complex [4]. In attempts to improve protein activity, enhancing association rates is preferable to thermodynamic stabilization of the complex, as interactions are typically diffusion limited [1]. To aid investigations into the electrostatics of protein interactions and to facilitate attempts to re-engineer protein behavior, our lab previously developed a computational tool for analysis of electrostatic structures of proteins (AESOP) [5–8]. Here, we present an implementation of AESOP in Python 2.7 with increased functionality and the capability of parallel processing. This framework may be used to both compare electrostatic similarity of protein families and

dissect the electrostatic contribution of individual amino acids to the free energy of association[5–8].

Though several other computational tools have been developed for similar purposes, AESOP is the only platform, to our knowledge, that is focused on protein electrostatics and offers multiple computational methods for both family-based and single-structure-based analyses. In comparison, Surface Diver and PIPSA have been developed to compare electrostatic potentials across families of structurally homologous proteins. Surface Diver accomplishes this quantitative comparison without prior structural superpositioning through spherical harmonic decomposition[9], whereas PIPSA requires superpositioning and instead allows for comparisons between common structural regions that are functionally similar[10,11]. Similar to PIPSA, AESOP requires superposed structures and quantitatively compares electrostatic potentials in a common grid space. PIPSA performs this comparison via calculation of a Hodgkin similarity index, a Carbo similarity index, or the average difference between electrostatic potentials in some specified spherical region or across the entire protein shell region[10,11], whereas AESOP uses a Reynolds linear similarity measure (vide infra) for every grid point in the system[7]. A feature unique to AESOP, however, allows for comparisons to be made for a library of mutants from an input protein structure. This comparison is performed by mutating all ionizable amino acids within the protein structure to alanine such that each mutant within the library has a single mutation. The end result is a computational analysis that identifies electrostatic hotspots within protein structure. Another computational tool named DrugScore^{PPI} uses statistical potentials derived from a database of experimental mutations to predict effects of alanine mutations at the interface

of protein complexes[12]. With this method, mutations may involve any type of amino acid. In contrast, AESOP uses Poisson-Boltzmann electrostatics and a thermodynamic cycle of solvation to predict effects of mutations involving ionizable amino acids on free energies of association relative to the parent structure. Table C.1 provides a brief comparison of features available across the computational tools discussed here and illustrates the flexibility of AESOP in studying the role of electrostatics in protein structure.

Table C.1: Summary of features and methods across several computational tools

Feature	AESOP	Surface Diver	PIPSA	DrugScore PPI
Alanine scan	Y	N	N	Y
Directed mutagenesis	Y	N	N	N
Electrostatic similarity	Y	Y	Y	N
Local installation	Y	Y	Y	N
Cross-platform	Y	N	N	-
Parallelized implementation	Y	N	Y	-
Mutant generation	Y	N	N	-
Free energy predictions	Y	N	N	Y
Web server	alanine scan	N	Y	Y

AESOP implements family-based comparisons of electrostatic potentials with the ElecSimilarity class and perturbation based analyses to gauge the contribution of single amino acids with the Alascan and Directed mutagenesis classes. Compared to previous iterations of the framework, the current implementation in Python has been coded in an object-oriented manner with each computational method implemented as a unique class. Furthermore, the current version of AESOP offers several additional features including energy minimization and structural superpositioning that are made possible through other python libraries such as Modeller. The end result is a more streamlined, efficient computational method for comprehensive investigations of protein electrostatics.

$$\Delta G_C$$

$$R_1 + ... + R_n \; \overset{\Delta G_C}{\leftrightharpoons} \; R_1\text{-}...\text{-}R_n \qquad \varepsilon_p = \varepsilon_s = \text{varies}$$

$$\Big\updownarrow \Delta G_S^{(1)} \qquad \Big\updownarrow \Delta G_S^{(n)} \; \Big\updownarrow \Delta G_S^{(1\text{-}...\text{-}n)} \qquad \kappa = 0$$

$$----------------$$

$$\varepsilon_p = \text{varies}, \; \varepsilon_s = 78.5$$

$$S_1 + ... + S_n \; \overset{}{\underset{\Delta G_A}{\leftrightharpoons}} \; S_1\text{-}...\text{-}S_n \qquad \kappa = \text{varies}$$
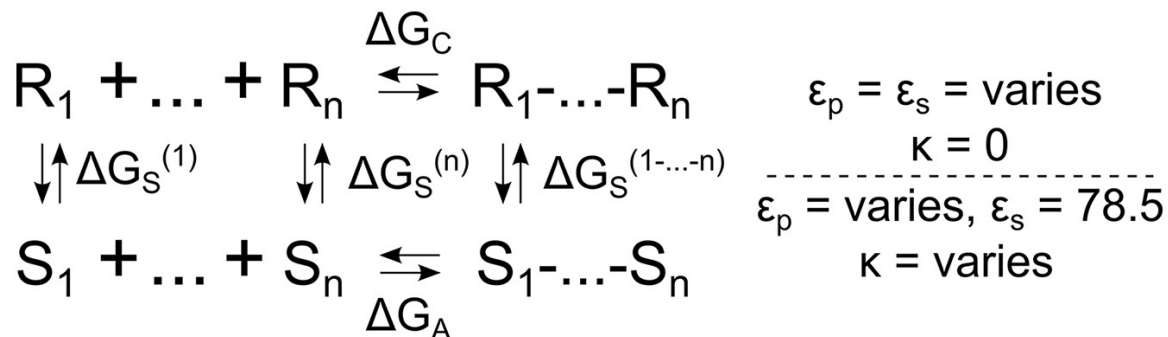
Figure C.1: Thermodynamic cycle for binding of protein subunits in a reference (top) and solvated (bottom) state. In the reference state, the dielectric constant, ϵ, is assumed to be uniform in the protein, ϵ_p, and solvent, ϵ_s; however, in the solvated state, the uniform dielectric constants may differ between the protein and solvent. For the solvent, we set the dielectric coefficient to 78.5, the value for water at room temperature. A parameter for ion accessibility, K, is held constant within each state. This variable is related to the ionic strength of the solvent.

C.1 Electrostatic potentials

Our framework utilizes the PDB2PQR[13,14] and Adaptive Poisson-Boltzmann Solver (APBS)[15] software to generate grid-based electrostatic potentials for protein systems in solvated and reference states. AESOP uses a linearized adaptation of the Poisson-Boltzmann equation from APBS to calculate electrostatic potentials from unit or partial charges located within the protein, the electric permittivity of the protein and solvent, and the ionic strength and accessibility of the solvent. Furthermore, AESOP avoids grid artifacts where possible by using consistent grid specifications, canceling out artifacts according to the thermodynamic cycle of Figure C.1.

C.2 Electrostatic similarity

Pairwise comparisons between grid-based potentials can be made according to

$$ESD(\psi_A, \psi_B) = \frac{1}{N} \sum_{k=1}^{N} \frac{|\psi_A - \psi_B|}{max(|\psi_A|, |\psi_B|)} \tag{C.1}$$

where N is the number of grid points, ψ is an electrostatic potential from system A or B at the same grid point k, and ESD is the electrostatic similarity distance[7]. An ESD of zero indicates identical electrostatic potentials at every grid point. Using this metric, we can compare a family of proteins on the basis of their electrostatic potentials in consistent grid dimensions.

C.3 Thermodynamic cycle

To predict the effect of mutations on the free energy of binding for some protein complex, our framework computes Coulombic and solvation free energies in each state of the thermodynamic cycle shown in Figure C.1[4–8]. This task is accomplished by invoking APBS multiple times for each state of the thermodynamic cycle with consistent grid-space specifications. For each calculation, the ionic strength, protein dielectric coefficient, and solvent dielectric coefficient are set according to the parameters of the solvated or reference state. Each state includes a set of protein subunits, $\{1, ..., n\}$, and the resulting protein complex, $1 - ... - n$. Each protein subunit and the resulting complex is analyzed in a reference state, $\{R_1, ..., R_n, R_{1-...-n}\}$, and a solvated state, $\{S_1, ..., S_n, S_{1-...-n}\}$. Although the end user is able to adjust ionic strength and dielectric coefficients, we suggest that the default values be used, since we have previously parameterized calculations to agree with experimental data[5].

Association free energies are calculated from mathematical operations on solvation and Coulombic free energies. First, the free energies of solvation (vertical processes in Figure C.1) are calculated according to

$$\Delta G_S = G_S - G_R \tag{C.2}$$

where ΔG_S is the free energy of solvation calculated for each vertical process. Then, $\Delta\Delta G_S$ is calculated according to

$$\Delta\Delta G_S = \Delta G_S^{1-\cdots-n} - \sum_{k=1}^{n} \Delta G_S^{(k)} \tag{C.3}$$

and it represents the difference in solvation free energy between the complex and the sum of the individual subunits. The change in Coulombic free energy of complex formation, ΔG_C (upper horizontal process in Figure C.1), is calculated similarly according to

$$\Delta G_C = G_C^{1-\cdots-n} - \sum_{k=1}^{n} \Delta G_C^{(k)} \tag{C.4}$$

where G_C represents the Coulombic free energy for the complex or an individual subunit. Finally, the free energy of association, ΔG_A (lower horizontal process in Figure C.1), is calculated according to

$$\Delta G_A = \Delta\Delta G_S + \Delta G_C \tag{C.5}$$

The resulting association free energy quantitatively describes the thermodynamic behavior of the protein system when considering only electrostatic forces, neglecting hydrophobic interactions and entropic effects.

C.4   Materials and Methods

AESOP is implemented in Python 2.7 and consists of a class for each method discussed here (three in total) and a number of accessory functions. In the spirit of open access, the latest version of the framework is available to all users through PyPi (`https://pypi.python.org`) and GitHub (`https://github.com`). Examples and walk-throughs of real-world biological analyses, along with detailed documentation, can be found on the AESOP webpage (`http://aesop.readthedocs.io`). AESOP primarily depends on APBS[15], PDB2PQR[13,14], Modeller[16,17], ProDy[18], NumPy[19], SciPy[20], NetworkX[21], and Matplotlib[22]. Notably, the ProDy library facilitates VMD-style selection strings to easily specify regions within a protein structure. In addition, AESOP currently uses the Python library multiprocessing for parallel processing. The organization of AESOP is inspired by libraries such as SciPy that encapsulate computational methods with programming objects[20].

C.4.1   Electrostatic similarity

Within the AESOP framework, the ElecSimilarity class has been implemented to set up and perform an analysis of electrostatic similarities within any family of protein structures. We advise the user to superpose these structures before running them through

AESOP analysis; however, the ElecSimilarity class can superpose or center coordinates through the Modeller and ProDy libraries[16–18]. After the user provides structural files, the ElecSimilarity class calls PDB2PQR to add hydrogen atoms, place charges, assign van der Waals radii in each structure according to the PARSE force field, and convert the Protein Data Bank file format to an APBS-readable PQR format[13]. In the case where only one structural file is provided, the ElecSimilarity class will generate a family of alanine mutants before converting structural files to PQR files. For each structure, electrostatic potentials are calculated in the same grid space and compared in a pairwise manner by calculating the ESD (Eqn. C.1). The results of this analysis can be returned as a distance matrix or a dendrogram (Figure C.2). This analysis may take a significant amount of time to complete if many structures are being compared or if the grid space contains a large number of vertices. Electrostatic similarity studies are useful for comparison of global electrostatic properties of families of homologous structures[23,24] or structurally homologous repeat modules in protein structures[25].

C.4.2   Alanine scan mutagenesis

To predict whether electrostatic interactions contribute favorably or unfavorably to association, AESOP implements a computational alanine scan with the Alascan class. This method accepts a single structure file and generates a library of mutants through side-chain truncation. Although the mutation scheme should not generate new clashes in protein structure, the end user has the option to perform minimization of all structures by implementing the conjugate gradient descent algorithm from Modeller. All mutations
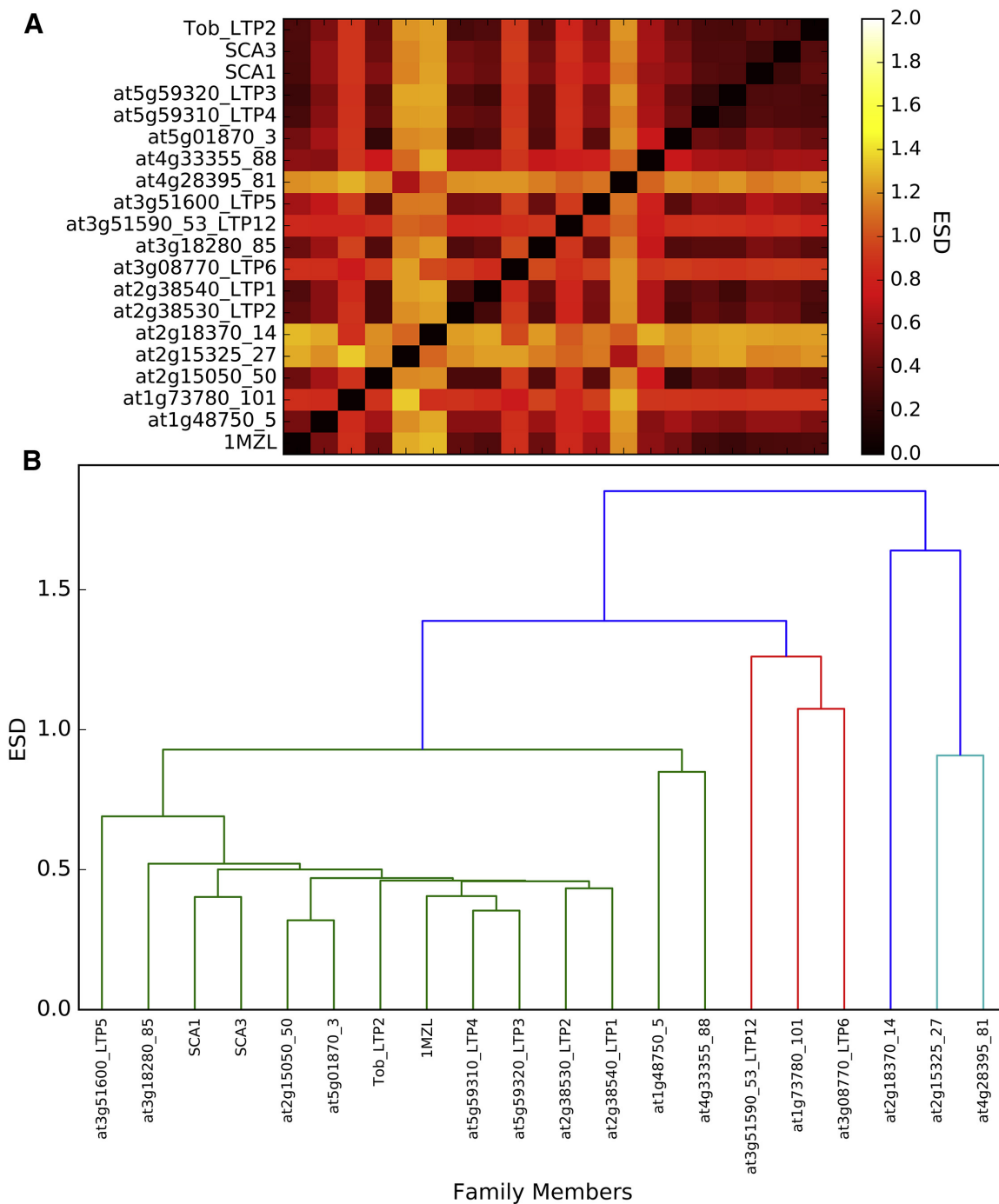
Figure C.2: (A) Example output from the plotESD function. ESD values closer to zero are more similar in terms of electrostatics. (B) Example output from the plotDend function where differences in ESD between family members are shown in a hierarchical dendrogram.

involve only ionizable amino acids, and each mutant structure contains a single mutation. Once generated, solvation and Coulombic free energies are calculated for each structure in the structure library, including the parent structure. Afterward, the user will have the option to view electrostatic free energies of association in a bar graph or a network, if the NetworkX[21] Python library is installed, or to return a matrix of free energies of association or solvation corresponding to the mutant identifiers (Figure C.3). In the default bar graph and network, we typically report relative free energies of association where the parent free energy of association is subtracted from the association free energies of each mutant. Thus, the resulting value for the parent will be zero. Mutations that have more significant effects on complex formation will have larger magnitude deviations from the parent association free energy, observed as outliers outside expected thermal fluctuations ($kT \sim \pm 2.5\,\mathrm{kJ/mol}$). The alanine scan is a perturbation method of identifying individual amino acids that have a significant effect in the formation of the parent protein complex (loss of binding mutations). It can also identify mutations or sites of mutations that will enhance association (gain of binding mutations) in studies of design of protein-protein interfaces[26–28].

C.4.3 Directed mutagenesis

Similar to alanine scan mutagenesis, the directed-mutagenesis class facilitates calculation of free energies of association according to the same thermodynamic cycle. The difference is that the end user must specify a list of exact mutations to be performed, although only mutations involving ionizable amino acids should be considered. If AESOP will be performing a mutation to alanine, it is best to use the Alascan class, as it will be
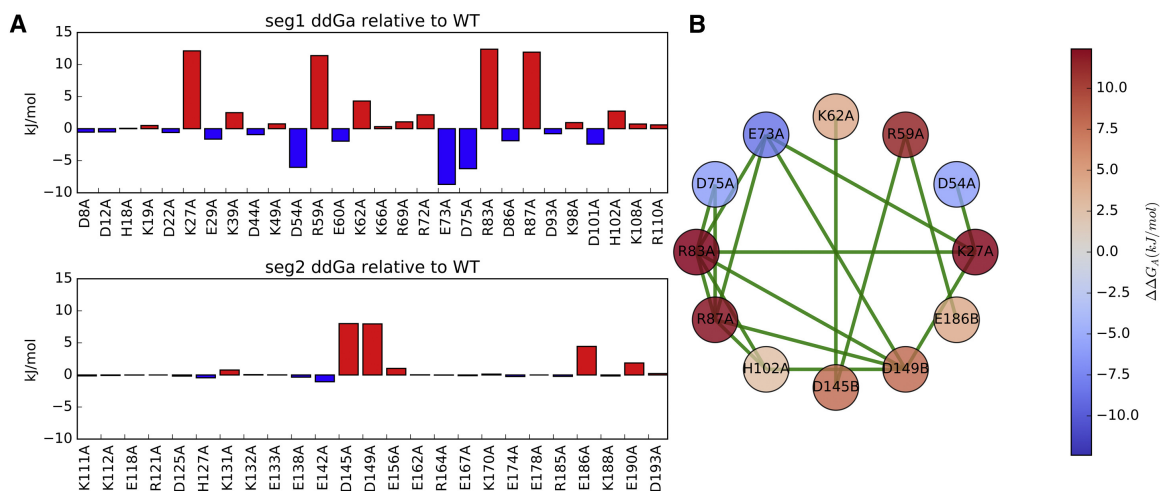
Figure C.3: (A) Example output from plotScan for an alanine scan showing perturbations of free energy of association relative to the wild-type. Perturbations below zero indicate gain-of-binding mutations, and perturbations above zero indicate loss-of-binding mutations. Effects that are significant are typically greater than or less than 2.5 kJ/mol compared to the parent. (B) Example output from plotNetwork for an alanine scan showing electrostatic interactions between residues with changes in free energy of association relative to parent outside thermal effects (± 2.5 kJ/mol). Edges indicate the presence of an interaction; nodes represent individual amino acids; and node colors are scaled according to the free energy of association. Each amino acid label consists of a one-letter amino acid code, the residue number, and the chain identification, in that order.

more efficient and will use a simple side-chain truncation method. To support more complicated mutations, the directed-mutagenesis class depends on Modeller[16,17]. As a result of the mutation process, backbone coordinates may be slightly perturbed around the mutated residue(s). Thus, this method requires extra calculations that the Alascan class does not require. Furthermore, each structure assessed with this method is minimized using the conjugate gradient descent algorithm in Modeller to remove clashes in protein structure. The directed-mutagenesis class supports the same outputs as the Alascan class. The directed mutagenesis studies provide information on the significance of the mutated amino acids in the stability of the structure of the parent complex[28], and they can also be used in the design of tailored protein-protein interfaces[29].

C.4.4   Web server portal

The AESOP alanine scan method is also implemented as a web server (`http://biomodel.engr.ucr.edu/software`) using the Django web framework (`https://djangoproject.com`) to reduce technical barriers and allow for rapid analysis. Results are also displayed with interactive graphs to simplify interpretation without manual scripting. The modular implementation of AESOP and the high scalability of Django facilitates the addition of new features into the framework and web server while servicing a growing number of users.

C.5   Conclusions

The AESOP framework provides a simple interface for quantitative comparisons of electrostatic potentials generated by proteins and predicting electrostatic contributions of ionizable amino acids to protein association. AESOP also contains accessory functions to generate figures for each computational method. As a result, the framework has potential to aid optimization of protein-protein interactions in redesigning protein interfaces as well as to investigate evolution of electrostatics across protein families. In past studies, AESOP has shown agreement with experimental mutagenesis studies involving protein interactions that are known to involve electrostatic interactions[5]. The developers of AESOP welcome and encourage collaboration through the AESOP GitHub repository (`https://github.com/BioMoDeL/aesop`). Installation instructions and example cases are at the AESOP website (`https://aesop.readthedocs.io`), and an archived version of the source code is deposited on Zenodo (`http://doi.org/10.5281/zenodo.269610`).

C.6 References

[1] McCammon, J. A. Darwinian biophysics: Electrostatics and evolution in the kinetics of molecular binding. *PNAS*, 106(19):7683–7684, December 2009. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0902767106. 00015.

[2] Zhou, H. X. Enhancement of protein-protein association rate by interaction potential: Accuracy of prediction based on local Boltzmann factor. *Biophys J*, 73(5):2441–2445, November 1997. ISSN 0006-3495.

[3] Frembgen-Kesner, T. and Elcock, A. H. Striking Effects of Hydrodynamic Interactions on the Simulated Diffusion and Folding of Proteins. *J. Chem. Theory Comput.*, 5(2): 242–256, February 2009. ISSN 1549-9618. doi: 10.1021/ct800499p.

[4] Elcock, A. H., Sept, D., and McCammon, J. A. Computer Simulation of Protein-Protein Interactions. *The Journal of Physical Chemistry B*, 105(8):1504–1518, March 2001. ISSN 1520-6106, 1520-5207. doi: 10.1021/jp003602d.

[5] Gorham, R. D., Kieslich, C. A., Nichols, A., Sausman, N. U., Foronda, M., and Morikis, D. An evaluation of poisson-boltzmann electrostatic free energy calculations through comparison with experimental mutagenesis data. *Biopolymers*, 95(11):746–754, 2011. ISSN 00063525. doi: 10.1002/bip.21644.

[6] Gorham, R. D., Kieslich, C. A., and Morikis, D. Electrostatic Clustering and Free Energy Calculations Provide a Foundation for Protein Design and Optimization. *Ann Biomed Eng*, 39(4):1252–1263, April 2011. ISSN 0090-6964. doi: 10.1007/s10439-010-0226-9.

[7] Kieslich, C. A., Gorham Jr., R. D., and 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*, 357(2):707–716, January 2011. ISSN 0022-3093. doi: 10.1016/j.jnoncrysol.2010.05.087.

[8] Kieslich, C. A., Morikis, D., Yang, J., and Gunopulos, D. Automated computational framework for the analysis of electrostatic similarities of proteins. *Biotechnology Progress*, 27(2):316–325, March 2011. ISSN 87567938. doi: 10.1002/btpr.541.

[9] Długosz, M. and Trylska, J. Electrostatic similarity of proteins: Application of three dimensional spherical harmonic decomposition. *J Chem Phys*, 129(1), July 2008. ISSN 0021-9606. doi: 10.1063/1.2948414.

[10] Blomberg, N., Gabdoulline, R. R., Nilges, M., and Wade, R. C. Classification of protein sequences by homology modeling and quantitative analysis of electrostatic similarity. *Proteins: Structure, Function, and Bioinformatics*, 37(3):379–387, November 1999. ISSN 1097-0134. doi: 10.1002/(SICI)1097-0134(19991115)37:3⟨379::AID-PROT6⟩3.0. CO;2-K.

[11] Richter, S., Wenzel, A., Stein, M., Gabdoulline, R. R., and Wade, R. C. webPIPSA: A web server for the comparison of protein interaction properties. *Nucleic Acids Res*, 36 (Web Server issue):W276–W280, July 2008. ISSN 0305-1048. doi: 10.1093/nar/gkn181.

[12] Krüger, D. M. and Gohlke, H. DrugScorePPI webserver: Fast and accurate in silico alanine scanning for scoring protein–protein interactions. *Nucleic Acids Res*, 38(Web Server issue):W480–W486, July 2010. ISSN 0305-1048. doi: 10.1093/nar/gkq471.

[13] Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res*, 32(Web Server issue):W665–W667, July 2004. ISSN 0305-1048. doi: 10.1093/nar/gkh381.

[14] Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., and Baker, N. A. PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res*, 35(Web Server issue):W522–W525, July 2007. ISSN 0305-1048. doi: 10.1093/nar/gkm276.

[15] Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. Electrostatics of nanosystems: Application to microtubules and the ribosome. *PNAS*, 98(18):10037–10041, August 2001. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.181342398.

[16] Fiser, A., Do, R. K., and Sali, A. Modeling of loops in protein structures. *Protein Sci*, 9(9):1753–1773, September 2000. ISSN 0961-8368.

[17] Martí-Renom, M. A., Stuart, A. C., Fiser, A., Sánchez, R., Melo, F., and Šali, A. Comparative Protein Structure Modeling of Genes and Genomes. *Annual Review of Biophysics and Biomolecular Structure*, 29(1):291–325, 2000. doi: 10.1146/annurev. biophys.29.1.291.

[18] Bakan, A., Meireles, L. M., and Bahar, I. ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics*, 27(11):1575–1577, June 2011. ISSN 1367-4803. doi: 10.1093/bioinformatics/btr168.

[19] van der Walt, S., Colbert, S. C., and Varoquaux, G. The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science Engineering*, 13(2):22–30, March 2011. ISSN 1521-9615. doi: 10.1109/MCSE.2011.37.

[20] Jones, E., Oliphant, T., Peterson, P., et al. SciPy: Open source scientific tools for Python, 2001–. URL `http://www.scipy.org/`.

[21] Hagberg, A. A., Schult, D. A., and Swart, P. J. Exploring Network Structure, Dynamics, and Function using NetworkX. page 5, 2008.

[22] Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science Engineering*, 9(3):90–95, May 2007. ISSN 1521-9615. doi: 10.1109/MCSE.2007.55.

[23] López de Victoria, A., Tamamis, P., Kieslich, C. A., and Morikis, D. Insights into the Structure, Correlated Motions, and Electrostatic Properties of Two HIV-1 gp120 V3 Loops. *PLoS ONE*, 7(11):e49925, November 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0049925.

[24] Kieslich, C. A. and Morikis, D. The Two Sides of Complement C3d: Evolution of Electrostatics in a Link between Innate and Adaptive Immunity. *PLoS Computational Biology*, 8(12):e1002840, December 2012. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1002840.

[25] Kieslich, C. A., Vazquez, H., Goodman, G. N., de Victoria, A. L., and Morikis, D. The effect of electrostatics on factor H function and related pathologies. *Journal of Molecular Graphics and Modelling*, 29(8):1047–1055, August 2011. ISSN 1093-3263. doi: 10.1016/j.jmgm.2011.04.010.

[26] E. S. Harrison, R., Gorham, R. D., and Morikis, D. Energetic evaluation of binding modes in the C3d and Factor H (CCP 19-20) complex. *Protein Science*, 24(5):789–802, May 2015. ISSN 1469-896X. doi: 10.1002/pro.2650. 00000.

[27] Gorham, R. D., Rodriguez, W., and Morikis, D. Molecular Analysis of the Interaction between Staphylococcal Virulence Factor Sbi-IV and Complement C3d. *Biophysical Journal*, 106(5):1164–1173, March 2014. ISSN 00063495. doi: 10.1016/j.bpj.2014.01.033.

[28] Mohan, R. R., Gorham Jr., R. D., and Morikis, D. A theoretical view of the C3d:CR2 binding controversy. *Molecular Immunology*, 64(1):112–122, March 2015. ISSN 0161-5890. doi: 10.1016/j.molimm.2014.11.006. 00002.

[29] Liu, Y., Kieslich, C. A., Morikis, D., and Liao, J. Engineering pre-SUMO4 as efficient substrate of SENP2. *Protein Eng Des Sel*, 27(4):117–126, April 2014. ISSN 1741-0126. doi: 10.1093/protein/gzu004.

Appendix D

Compstatin: β-turn analysis

D.1 Introduction

Compstatin, a cyclic peptide of 13 amino acids in length, is capable of inhibiting complement response through interaction with C3. Compstatin binds to C3 and sterically prevent its cleavage into C3a and C3b by convertases, thus preventing downstream chemotaxis, opsonization and membrane attack complex formation.

The original compstatin has the sequence I[CVVQDWGHHRC]T-NH2, where brackets denote cyclization through a disulfide bridge between C2 and C12, here and hereafter. The original NMR structural study in solution demonstrated the presence of an ensemble of interconverting conformers, with a detectable conformer of estimated 42-63% population comprising a Type I beta-turn[1]. The Type I beta-turn sequence segment is Q5DWG8. Subsequent NMR structural studies of a number of compstatin analogs aimed at detecting structural changes induced by strategically incorporated amino acid replacements with goal to disrupt or enhance the observed beta-turn and surrounding residues[2]. Accord-

ing to nuclear Overhauser effect (NOE) connectivity patterns (Figure 9 of Ref[2], the following structural observations were made: (1) linearization of the peptide through the C2A/C12A replacements (resulting to abolishment of the disulfide bridge) showed that a Type I beta turn structure in the segment D4/Q5DW7/G8 remains. (2) Introduction of the H9A replacement in an N-terminal acetylated original sequence of compstatin showed propensity for Type I beta-turn formation in the segment Q5DWGAHRC12. This is attributed to the helical propensity of the alanine amino acid. Given that a Type I beta-turn is one half of an alpha-helical turn, this data suggests capability for multiple Type I beta-turn formations in the C-terminal segment, Q5DWGAHRC12. (3) Introduction of the V3A replacement in the N-terminal acetylated original sequence of compstatin showed propensity for multiple Type I beta-turn formations in the opposite segment I1AVQDWG8. (4) Introduction of the W7F replacement in the N-terminal acetylated original sequence of compstatin showed that the propensity for Type I beta-turn formation in the segment Q5DWG8 remains, with the possibility of an additional turn in the segment I1CVV4. Another solution NMR structural study of an N-terminal acetylated compstatin analog with the V4W/H9A replacement (sequence Ac-I[CVWQDWGAHRC]T-NH2) suggested the presence of interconverting conformers, with a major conformer comprising a coil with a flexible type I beta-turn in the segment Q5DWG8, based on NOE connectivity and chemical shift analysis[3]. Overall the NMR data suggest mobile Type I beta-turn(s) depending on the sequence of compstatin analogs. A subsequent cocrystal X-ray structure of the V4W/H9A compstatin analog in complex with C3c (a cleavage product of C3)[4], showed a bound conformation with a Type I beta-turn in the segment G8AHR11.

To evaluate if V4W/H9A-compstatin undergoes a conformational transition upon binding or if the difference in the position of its Type I beta-turn compared to that of the original compstatin depends on the V4W/H9A replacement, we performed microsecond molecular dynamics simulations in explicit solvent. The initial structure of the MD simulations was that of the V4W/H9A analog, extracted from the C3d-bound crystal structure.

D.2 Methods

D.2.1 Explicit-solvent molecular dynamics (MD) simulations

Explicit-solvent molecular dynamics (MD) simulations were carried out using the structure of Compstatin from the cocrystal structure of C3c:Compstatin (chain G in PDB: 2QKI)[4]. Initial minimization of the structure in the absence of water was carried out using NAMD[5] and the CHARMM36 forcefield[6]. The structure was then solvated in TIP3P water boxes with dimensions of 97 Å $\times$ 97 Å $\times$ 86 Å and neutralized with sodium and chloride counterions at an ionic strength of 150 mM. Subsequent to addition of solvent, the structure underwent 25,000 steps of conjugate gradient energy minimization followed by heating from 0 to 300 K in 62 ps with all protein atoms constrained to their post-minimization positions. Following heating, the system was equilibrated through five stages for 1 ns/stage with force constants 41.84, 20.92, 8.368 and 4.184 kJ/mol/Å^2 applied during the first four stages respectively to harmonically constrain all protein atoms to their post-minimization positions. During the final stage of equilibration a force constant of 4.184 kJ/mol/Å^2 was applied to constrain only protein backbone atoms to their post-minimization positions. Following equilibration, two 2 µs production runs were carried out using AMBER16[7] with: periodic

boundary conditions, Langevin dynamics, nonbonded interaction cutoff of 12 Å, SHAKE algorithm, and an integration timestep of 2 fs.

D.2.2 Trajectory analysis

Analysis of the trajectory was performed using MDTraj and pandas[8,9]. tICA decomposition was performed on the phi and psi angles observed throughout the trajectory using MSMBuilder[10]. The MiniBatchKMeans method in MSMBuilder was utilized to cluster the components to 50 distinct clusters and cluster centers were extracted as representative structures.

β-turn classification was performed using a custom Python script applying criteria from Hutchinson et al.[11] (as seen in Table D.1) with a tolerance of $\pm 45°$. The classification was applied to each frame in the two simulation trajectories so that an occupancy of each type of turn observed could be calculated. The classifiation was also applied to each representative structure extracted through tICA clustering so that the turn-type occupancy could be calculated.

Table D.1: Compstatin β-turn analysis - Representative Structures

Type	Phi (i+1)	Psi (i+1)	Phi (i+2)	Psi (i+2)
I	-60	-30	-90	0
II	-60	120	80	0
VIII	-60	-30	-120	120
I'	60	30	90	0
II'	60	-120	-80	0
VIa1	-60	120	-90	0
VIa2	-120	120	-60	0
Vib	-135	135	-75	160
IV	Everything else			

Rows represent the type of turn and the column headers represent the ith amino acid (from the beta turn classification i, $i+1$, $i+2$, $i+3$). Classification analysis was performed using 50 representative structures extracted through clustering of

D.3 Results

Tables D.2, D.3, and D.4 show the results of beta-turn classification applied to two 2-µs MD simulation trajectories and clustered representative structures, respectively. Our discussion will focus on well-defined beta-turns of Type I, I', II, and II', and not on generic turns of Type IV, VI, and VIII which may be considered as part of coiled structures. The Type I beta-turn at segment G8AHR11 of the original structure has been lost (occupancies of 0.8% and 1.1%, Tables D.2 and D.3, respectively) and moved to segment Q5DWG8 (occupancies of 20.6% and 5.4%, Tables D.2 and D.3, respectively), the same segment where the Type I beta-turn was observed in the solution NMR structure of the original non-acetylated compstatin. This is also observed in the data of the representative clustered conformations with occupancy of 10% for the segment Q5DWG8 (Table D.4), although these data, and data from Run 2 (Table D.3), also suggest an alternative Type I beta-turn in the segment D6WGA9 with 6% occupancy. Overall, these are low occupancies

Table D.2: Compstatin β-turn analysis - Production Run 1

Type	ILE1	CYS2	VAL3	TRP4	GLN5	ASP6	TRP7	GLY8	ALA9	HIS10
I	0.1	0.5	3.3	1.1	20.6	0	0	0.8	0	0.5
I'	0	0	0	0	0	0	0	0	0.1	0
II	0	0	2.8	0	0.2	0	0	0	0.5	0
II'	0	0	0	0	0	0	0	0	0	0
IV	75.2	76.1	57.2	67	78.8	99.8	97.3	87.4	70	75.5
VIII	0.3	0.4	2.1	1.4	0.1	0.1	0.7	2.3	1.8	1.8
VIa1	0	2.1	0.6	26.2	0.1	0	0.4	5.3	1	1.3
VIa2	0	7.1	1.3	3.6	0.2	0	0.4	1.7	2.2	5
VIb	24.3	13.8	32.7	0.7	0	0	1.2	2.5	24.4	15.9

Rows represent the type of turn and the column headers represent the ith amino acid (from the beta turn classification i, $i + 1$, $i + 2$, $i + 3$).

and suggest flexibility in the compstatin analog, in line with the presence of ensemble of interconverting conformers suggested by the NMR data. Nevertheless, the data suggest that the solution conformation differs from that of the bound conformation, and support the notion that compstatin undergoes a conformational transition upon binding. However, a bound conformation influenced by crystallization conditions and crystal packing effects cannot be excluded. Our conclusions, above, may be taken with caution because our MD simulation data consist only of two MD trajectories. To obtain a more comprehensive picture of compstatin dynamics a larger sample of replicate MD trajectories may be necessary, perhaps spanning longer timescales or performed using advanced sampling methods.

Table D.3: Compstatin β-turn analysis - Production Run 2

Type	ILE1	CYS2	VAL3	TRP4	GLN5	ASP6	TRP7	GLY8	ALA9	HIS10
I	1.5	0	0.3	2.5	5.4	5.7	2.5	1.1	1.9	1.7
I'	0	0	3.7	0	0	0	0	0	0.1	0
II	0	0.2	0	0	0	0	0	0.2	0.8	0
II'	0	0	0	0	0	0	0	0	0	0
IV	69.4	64.8	50.5	76.6	83.9	91.1	91.2	89.3	75.8	66.8
VIII	1.1	7.6	5.7	2.6	0.8	1.8	3	3.6	1.3	1.6
VIa1	4.3	2.5	6.3	9.8	5.1	1.1	0.4	0.5	0.1	2
VIa2	4.1	11.1	20.8	2.7	4.2	0.2	0.6	0.5	1.9	3.9
VIb	19.5	13.7	12.8	5.7	0.5	0.1	2.3	4.7	18.1	24

Rows represent the type of turn and the column headers represent the ith amino acid (from the beta turn classification i, $i+1$, $i+2$, $i+3$).

Table D.4: Compstatin β-turn analysis - Representative Structures

Type	ILE1	CYS2	VAL3	TRP4	GLN5	ASP6	TRP7	GLY8	ALA9	HIS10
I	2	2	2	4	10	6	2	0	2	8
I'	0	0	2	0	0	0	0	0	0	0
II	0	0	2	0	0	0	0	0	4	0
IV	68	62	60	74	78	90	94	90	76	64
VIII	2	4	6	4	4	2	0	4	2	2
VIa1	8	0	4	14	4	0	0	4	0	4
VIa2	2	10	10	2	4	2	2	0	2	4
VIb	18	22	14	2	0	0	2	2	14	18

Rows represent the type of turn and the column headers represent the ith amino acid (from the beta turn classification i, $i+1$, $i+2$, $i+3$). Classification analysis was performed using 50 representative structures extracted through clustering of

D.4 References

[1] Morikis, D., Assa-Munt, N., Sahu, A., and Lambris, J. D. Solution structure of Compstatin, a potent complement inhibitor. *Protein science: a publication of the Protein Society*, 7(3):619, 1998.

[2] Morikis, D., Roy, M., Sahu, A., Troganis, A., Jennings, P. A., Tsokos, G. C., and Lambris, J. D. The Structural Basis of Compstatin Activity Examined by Structure-Function-based Design of Peptide Analogs and NMR. *J. Biol. Chem.*, 277(17):14942–14953, April 2002. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.M200021200. 00060.

[3] Mallik, B., Katragadda, M., Spruce, L. A., Carafides, C., Tsokos, C. G., Morikis, D., and Lambris, J. D. Design and NMR Characterization of Active Analogues of Compstatin Containing Non-Natural Amino Acids. *J. Med. Chem.*, 48(1):274–286, January 2005. ISSN 0022-2623. doi: 10.1021/jm0495531. 00072.

[4] Janssen, B. J. C., Halff, E. F., Lambris, J. D., and Gros, P. Structure of Compstatin in Complex with Complement Component C3c Reveals a New Mechanism of Complement Inhibition. *J. Biol. Chem.*, 282(40):29241–29247, May 2007. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.M704587200.

[5] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16):1781–1802, December 2005. ISSN 0192-8651, 1096-987X. doi: 10.1002/jcc.20289.

[6] MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., and Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B*, 102(18): 3586–3616, April 1998. ISSN 1520-6106. doi: 10.1021/jp973084f.

[7] Case, D. A., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., Iii, V. W. D. C., Darden, T. A., Duke, R. E., Ghoreishi, D., Gilson, M. K., Gohlke, H., Goetz, A. W., Greene, D., Harris, R., Homeyer, N., Izadi, S., Kovalenko, A., Kurtzman, T., Lee, T. S., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Mermelstein, D. J., Merz, K. M., Miao, Y., Monard, G., Nguyen, C., Nguyen, H., Omelyan, I., Onufriev, A., Pan, F., Qi, R., Roe, D. R., Roitberg, A., Sagui, C., Schott-Verdugo, S., Shen, J., Simmerling, C. L., Smith, J., Salomon-Ferrer, R., Swails, J., Walker, R. C., Wang, J., Wei, H., Wolf, R. M., Wu, X., Xiao, L., York, D. M., and Kollman, P. A. *AMBER 2018*. University of California, San Francisco, 2018.

[8] McGibbon, R. T., Beauchamp, K. A., Harrigan, M. P., Klein, C., Swails, J. M., Hernández, C. X., Schwantes, C. R., Wang, L.-P., Lane, T. J., and Pande, V. S.

MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal*, 109(8):1528–1532, October 2015. ISSN 0006-3495. doi: 10.1016/j.bpj.2015.08.015.

[9] McKinney, W. Data Structures for Statistical Computing in Python. page 6, 2010.

[10] Beauchamp, K. A., Bowman, G. R., Lane, T. J., Maibaum, L., Haque, I. S., and Pande, V. S. MSMBuilder2: Modeling Conformational Dynamics on the Picosecond to Millisecond Scale. *J. Chem. Theory Comput.*, 7(10):3412–3419, October 2011. ISSN 1549-9618. doi: 10.1021/ct200463m.

[11] Hutchinson, E. G. and Thornton, J. M. A revised set of potentials for beta-turn formation in proteins. *Protein Sci*, 3(12):2207–2216, December 1994. ISSN 0961-8368.

Appendix E

Crosslinked flagella as a stabilized vaccine adjuvant scaffold

Influenza, one of the most prominent infectious diseases, continues to be a persistent global threat. Control of influenza is reliant on vaccines using conventional technology relying on triggering of immune response in patients. The bacterial protein flagellin is capable of triggering strong inflammatory responses and has, in the form of polymeric flagellin, demonstrated potential as a vaccine adjuvant in the past[1]. In this study, we utilize a computational approach for identifying potential crosslinking disulfide bridges to further stabilize the flagellar filament.

Modeling of the flagellar filament was carried out in UCSF Chimera[2] using an R-type flagellar filament structure (PDB Code 1UCU[3]) with which a multimeric structure, comprised of 26 monomers, was generated. The translation-rotation operation used to generate the multimeric structure was kindly provided to us by Dr. Keiichi Namba (Osaka
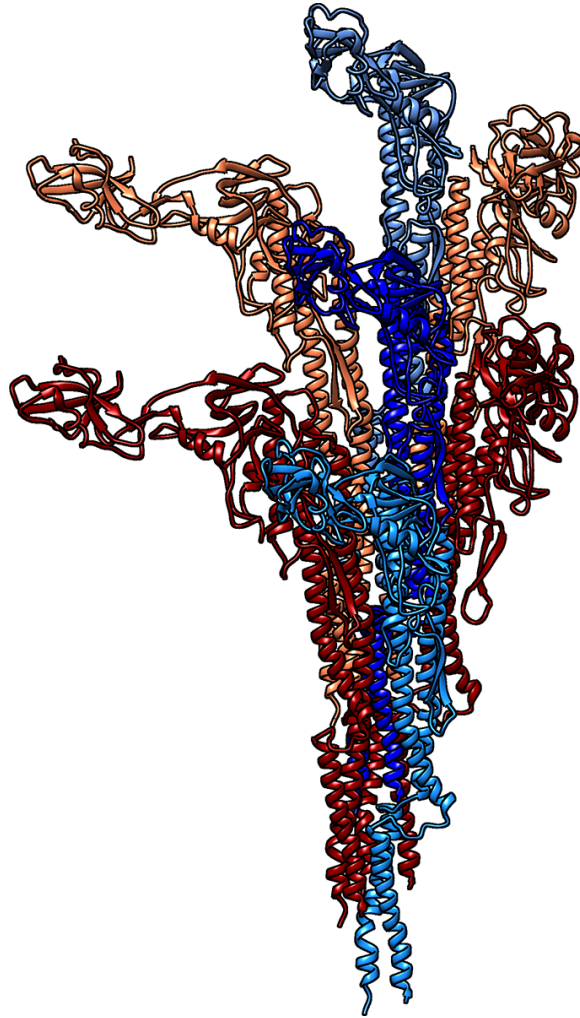
Figure E.1: Molecular graphics of a 11-mer subunit of R-type flagellin.

University). The translation was performed in the z-axis using a value of 4.7058 angstroms and the rotation was performed using an angle of 65.817°. A 11-mer subunit of the filament structure can be visualized in Figure E.1, where we see the upward translation of monomers and the consistency in intermolecular contacts. Through intermolecular interaction analysis as well as structural and physicochemical analysis, potential sites for crosslinking flagellin monomers through disulfide bridges were identified. Crosslinker sites were identified at the D0 and D1 domains for their flexibility while the D2 and antigenic D3 domain were avoided (Figure E.2). Sites were also chosen at locations so as not to disrupt the helical structures of D0 and D1. Additionally, each crosslinker pair contains at least one mutation at a helical terminus or at a loop for the steric accomodation of the disulfide bridge. Asparagine, glutamine, and glutamic acid residues were selected for the introduction of disulfide bridges as their side chain rotameric states are capable of interatomic distances necessary for the formation of a disulfide bridge. The mutation pairs identified are E454(A)-N5(B), N430(A)-N38(B), and Q97(A)-Q62(B).

Through experimental characterization (carried out by our collaborators), we were able to characterize disulfide bridges' impact on structural stability in flagellin. Results demonstrated that the introduction of the disulfidge bridges retained the ability to form flagellae while also having distinct impacts on bacterial motility and thermal stability.
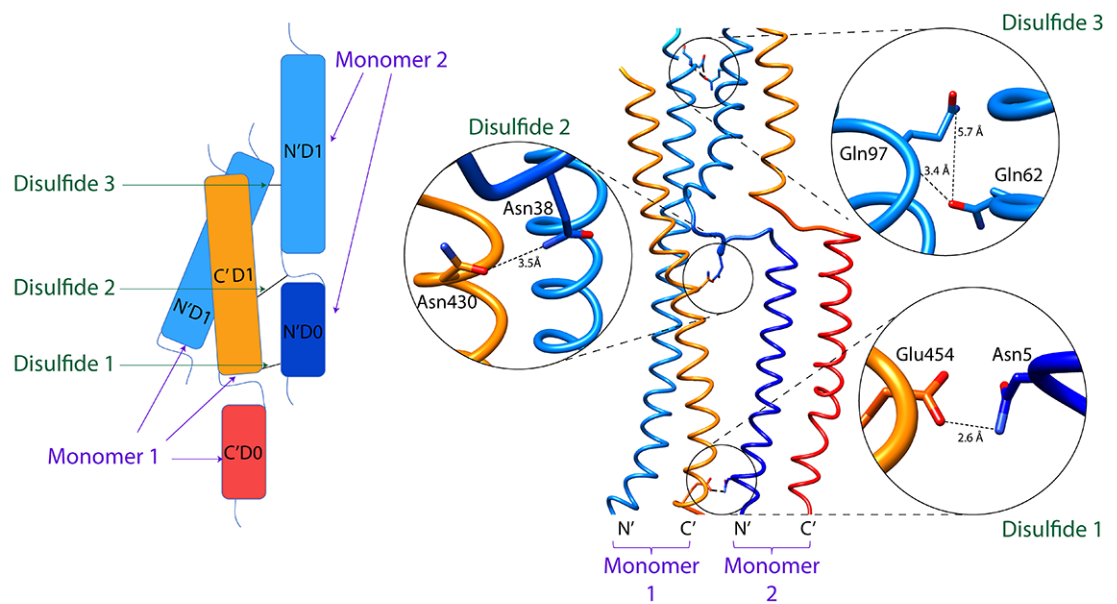
Figure E.2: (Left) Depiction of the helical translations and proposed locations to introduce disulfide bonds between neighboring flagellin protofilaments. Domains N'D0 and N'D1 are colored dark and light blue respectively while C'D0 and C'D1 are orange and red, respectively. (Right) Molecular graphic representation of sites of interest within neighboring protofilament monomers. Insets show a zoomed and rotated view of the amino acid interactions with the specific distances shown: 1) Glu454[Oϵ1] – Asn5[Nδ1], 2) Asn430[Oδ1] – Asn38[Nδ1], and 3) Gln97[N] – Gln62[Oϵ1], Gln97[Nδ2] – Gln62[Oϵ1].

E.1 References

[1] Bennett, K. M., Gorham, R. D., Gusti, V., Trinh, L., Morikis, D., and Lo, D. D. Hybrid flagellin as a T cell independent vaccine scaffold. *BMC Biotechnology*, 15:71, August 2015. ISSN 1472-6750. doi: 10.1186/s12896-015-0194-0.

[2] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. UCSF Chimera?A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, October 2004. ISSN 0192-8651, 1096-987X. doi: 10.1002/jcc.20084.

[3] Maki-Yonekura, S., Yonekura, K., and Namba, K. Conformational change of flagellin for polymorphic supercoiling of the flagellar filament. *Nat Struct Mol Biol*, 17(4):417–422, April 2010. ISSN 1545-9993. doi: 10.1038/nsmb.1774. 00052.

Appendix F

General code utilized for analysis

Code for the following types of analysis and scripts will be available at the BioMoDeL GitHub repository (https://github.com/BioMoDeL):

1. MD setup scripts

2. RMSD timeseries plots for trajectory analysis

3. SASA timeseries plots for trajectory analysis

4. Hydrogen bond analysis and occupancy heatmap plots for trajectory analysis

5. Salt bridge analysis and occupancy heatmap plots for trajectory analysis

6. Hydrophobic contact analysis and occupancy heatmap plots for trajectory analysis

7. Hydrogen bond heatmaps for CPPTRAJ output

8. Salt bridge heatmaps for CPPTRAJ output

9. Interatomic distance timeseries plots with confidence intervals

10. Interatomic distance timeseries heatmaps

11. AESOP codebase

12. Featurize and cluster MD simulation trajectories to extract representative structures

Appendix G

CV

Education	**University of California, Riverside** Ph.D., Bioengineering, 2018. **University of Iowa** B.S.E, Biomedical Engineering, 2013.
Dissertation	"Title of Dissertation" Short description/summary of research and estimation techniques. This can be several lines long because of the paragraph box.
Research	**Graduate Student Researcher** Sept. 2013 - Present BioMoDeL, University of California, Riverside

- Computational analysis of complement response and regulation

- Virtual Drug Screening

- Rational Peptidic Design

- Vaccine Design

- Co-development of electrostatic analysis framework AESOP

Visiting Graduate Student Researcher June 2015 - Sept. 2015
McCammon Lab and SDSC, University of California, San Diego

- Utilized supercomputing infrastructure to conduct accelerated molecular dynamics and Brownian dynamics studies

Programmer May 2013 - Aug. 2013
Public Policy Center, University of Iowa

- Developed a specially instrumented bicycle workflow through a collaborative project to validate the University of Iowa bike simulator.

- Designed a GUI using MATLAB and C# for intake of data (biometric, video, and GPS) and incorporated UX design. Success of project led to a novel pilot study.

Undergraduate Research Assistant Aug. 2011 - Aug. 2013
Dr. Beverly Davidson Lab, University of Iowa

- Investigated a potential transgene solution to SCA6 using molecular techniques, managed a mouse model and performed cell and tissue culture as well as brain tissue extraction.

Industry	**R&D IT Intern**	June 2017 - Sept. 2017
	Biogen, Cambridge, MA	

- Characterized Parkinson's disease through signal processing and machine learning analysis of wearable device data

- Associated patient consent with biospecimens by implementing a Python-based web portal for expediting researcher biospecimen procurement

Teaching and Mentoring	**Consultant**	June 2016 - June 2018
	GradQuant, University of California, Riverside	

Teaching Assistant

BIEN001 - Bioengineering Colloquium	Fall 2016
BIEN165 - Biomolecular Engineering	Spring 2016
BIEN130L - Bioinstrumentation Lab	Spring 2015
BIEN135 - Biophysics and Biothermodynamics	Fall 2014, 2016, 2017

Awards and Fellowships	**Outstanding Teaching Award, Bioengineering**	May 2018
	Graduate Division, University of California, Riverside	

NSF Travel Award	Oct. 2015
Macromolecular Simulation Workshop, CECAM, Jülich, Germany	

SDSC-UC Graduate Student Summer Fellowship	June 2015
San Diego Supercomputing Center, UCSD	

Finalist, NVIDIA GPU Award	March 2015
ACS National Meeting and Exposition, Denver, Colorado	

Best Oral Presentation Award	June 2014
UC Systemwide Bioengineering Symposium, UC Irvine	

Dean's Distinguished Fellowship	2013-2015
University of California, Riverside	

Skills

- **Programming**: Python, R, Perl, C, MATLAB, databases, shell scripting, high performance computing, version control, LaTeX, and UNIX environment

- **Computational Chemistry and Biology**: Molecular Dynamics (MD) (Conventional, Steered, Accelerated), Docking, Virtual Screening, Poisson-Boltzmann Electrostatics, Brownian Dyanmics, Network Analysis, Schrödinger Suite, NAMD, VMD, AMBER, Rosetta

- **Bioinformatics Tools and Databases**: ExPASy, EMBL-EBI, PDB, KEGG, BLAST, and NCBI

- **Statistical Analysis**: Statistical mechanics, validation, dimensionality reduction

- **Machine Learning**: Clustering, classification, and regression

- **Graph Theory and Network Analysis**: Network robustness, identification of communities

- **Data Visualization**: Plotly, ggplot, matplotlib, R-Shiny, Gephi, Cytoscape, Chimera

- **Experimental**: Protein Binding Assays (Microscale Thermophoresis, ELISA), Cell Culture, Tissue Culture, Mouse Perfusions and Surgery, Microarray Hybridization and Analysis

Publications

1. **Rohith R. Mohan**, Ronald D. Gorham Jr., and Dimitrios Morikis. "A Theoretical View of the C3d:CR2 Binding Controversy." *Molecular Immunology* 64, no. 1 (March 2015): 112–22.

2. **Rohith R. Mohan**, Gary A. Huber, and Dimitrios Morikis. "Electrostatic Steering Accelerates C3d: CR2 Association." *The Journal of Physical Chemistry B*, 2016 (April 2016).

3. **Rohith R. Mohan**, Andrea P. Cabrera, Reed ES Harrison, Ronald D. Gorham, Lincoln V. Johnson, Kaustabh Ghosh, and Dimitrios Morikis. "Peptide Redesign for Inhibition of the Complement System: Targeting Age-Related Macular Degeneration." *Molecular Vision* 22 (October 2016): 1280–90.

Publications
(contd.)

4. Reed ES Harrison, **Rohith R. Mohan**, Ronald D. Gorham Jr., Chris A. Kieslich and Dimitrios Morikis. "AESOP: A Python Library for Investigating Electrostatics in Protein Interactions." Biophysical Journal (May 2017).

5. **Rohith R. Mohan**, Mark Wilson, Ronald D. Gorham, Jr., Reed E. S. Harrison, Vasilios A. Morikis, Chris A. Kieslich, Asuka A. Orr, Alexis V. Coley, Phanourios Tamamis, and Dimitrios Morikis. "Virtual Screening of Chemical Compounds for Discovery of Complement C3 Ligands." *ACS Omega* 2018 3 (6), 6427-6438.

6. Nehemiah Zewde, **Rohith R. Mohan**, and Dimitrios Morikis. "Immunophysical Evaluation of the Initiating Step in the Formation of the Membrane Attack Complex". (In revision)

7. Casey M. Gries, **Rohith R. Mohan**, Dimitrios Morikis, and David D. Lo. "Crosslinked flagella as a stabilized vaccine adjuvant scaffold". (Submitted)

Conference
Presentations

1. **Rohith R. Mohan**, Ronald D. Gorham Jr., Dimitrios Morikis. "Investigating the Binding Mode of C3d-CR2 Using Poisson-Boltzmann Electrostatic Calculations and Molecular Dynamics Simulations." Oral presentation, 15th Annual UC Systemwide Bioengineering Symposium, University of California, Irvine, June 2014.

2. **Rohith R. Mohan**, Ronald D. Gorham Jr., Dimitrios Morikis. "Theoretical view of the C3d-CR2 binding controversy." Poster presentation, 249th ACS National Meeting, Division of Computers in Chemistry, Denver, Colorado, March 2015.

3. **Rohith R. Mohan**, Gary A. Huber, J. Andrew McCammon, Dimitrios Morikis. "Interplay between Ionic Strength, Association Rates and Electrostatic Interaction in the C3d:CR2 complex" Poster presentation, Biophysical Society 60th Annual meeting, Los Angeles, California, February 2016.

Conference Presentations (contd.)

4. **Rohith R. Mohan** and Dimitrios Morikis. "Peptide Design for Thernaostic Applications." Poster presentation, 19th Annual UC Systemwide Bioengineering Symposium, University of California, Riverside, June 2018.

5. Reed ES Harrison, **Rohith R. Mohan** , Ronald D. Gorham Jr., Chris A. Kieslich and Dimitrios Morikis. "AESOP: A Python Library for Investigating Electrostatics in Protein Interactions." Gordon Research Conference: Computational Chemistry, West Dover, Vermont, July 2018.

Outreach

Judge 2016, 2017
BioMoDeL community award, Riverside County Science Fair

President June 2015 - June 2016
Bioengineering Interdepartmental Graduate Student Association
University of California, Riverside

Parliamentarian June 2014 - June 2015
Bioengineering Interdepartmental Graduate Student Association
University of California, Riverside

Volunteer Test-writer and Examiner Feb. 2014 - Feb. 2017
Inland Empire Regional Science Olympiad

References

Professor Dimitrios Morikis
Department of Bioengineering
University of California, Riverside
dmorikis@ucr.edu

Dr. Yelda Serin
Graduate Division
University of California, Riverside
yelda.serin@ucr.edu