University of California, Merced

Evaluating Protein-Protein Interactions in Chemokine-Inhibitor Complexes Through MD Simulation

A dissertation submitted in partial satisfaction of the requirements

for the degree Doctor of Philosophy

in

Quantitative and Systems Biology

by

Lauren Stark

2021

Committee

Committee Chair: Professor Patricia LiWang

Committee Member: Professor Christine Isborn

Committee Member: Professor Maria Zoghbi

Advisor: Professor Michael Colvin

The Dissertation of Lauren Elizabeth Stark is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

Chair

University of California, Merced

2021

Dedications

I would like to dedicate my dissertation to my loving husband, who provided me with endless support, advice, and chocolate.

I would like to thank my advisor and our collaborators for their guidance along the way and their infectious enthusiasm on these topics.

Finally, I would also like to thank my friends and family for their constant encouragement that made this journey possible.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Funding received from NSF Graduate Research Fellowship under Grant No. 1744620.

Additional funding was awarded through the Graduate Dean Dissertation Fellowship Spring 2021.

This work used the Extreme Science and Engineering Discovery Environment (XSEDE) Bridges-2 at the Pittsburgh Supercomputing Center with allocation TG-BIO210023, which is supported by National Science Foundation grant number ACI-1548562.

Curriculum Vitae

Education

Ph.D., Quantitative and Systems Biology — Fall 2014 – Summer 2021
University of California, Merced
3.805 cumulative GPA

B.S Biological Sciences, emphasis in Microbiology and Immunology — May 2012
University of California, Merced
3.749 cumulative GPA
Graduated high honors (top 5% in department)

Laboratory Skills Experience

University of California, Merced — July 2017 – , Merced, CA
Conducted research for Dr. Michael Colvin
Research: Model chemokine and chemokine inhibitor binding interactions using Molecular Dynamics simulations to identify key residues in binding strength and specificity.
Techniques: Python, R, GROMACS, bash scripting

University of California, Merced — August 2014 – July 2017, Merced, CA
Conducted research for Dr. Fabian Filipp
Research: Model metabolic flux for cancer cells using patient data from The Cancer Genome Atlas (TCGA). From this model, identify key metabolic switches affected in cancer that explain metabolic shifts observed. Compare changes in metabolism through tumor progression.
Techniques: Python, R, and bash scripting for metabolic modeling and data analysis

University of California, Merced — August 2011 – July 2014, Merced, CA
Conducted research for Dr. Linda Hirst
Research: Project 1: Use experimental and computational techniques to examine the effects crosslinker morphology, filament length, and persistence length have on semi-flexible polymer network formation and structure. Use actin as model.
Project 2: Analyze liquid crystal (LC)/ Quantum dot (QD) dispersions. Test the effect of LC ligands attached to the QD on the overall dispersion/clustering. Identify packing order and structure of QD clusters using X-ray diffraction data.

Techniques: Coarse-grained molecular dynamics simulations using NAMD; statistical analysis using Python and Matlab; confocal, phase contrast, and fluorescence microscopy; X-ray diffraction analysis

Additional Work Experience

Teaching Assistant – UC Merced

Mathematical Modeling for Biology Spring 2018 – Fall 2020, Merced, CA

Ran two to three discussion sections each week, graded homework and tests, proctored exams

General Virology Summer 2014, Fall 2018, Merced, CA

Ran two discussion sections, graded homework, presentations, and tests, proctored exams

General Microbiology Lab Fall 2014, Merced, CA

Ran a lab section twice a week, graded reports and exams, generated quizzes, proctored exams

Research Assistant, Lab Technician – Hirst Lab January 2014 – July 2014, Merced, CA

Research assistant – assigned to the LC/QD project. Prepare samples for a variety of experiments; assist in x-ray diffraction experiments at SLAC's SSRL facility; assisted in analysis of results and image collection.

Lab technician – roles included general lab maintenance through cleaning, stocking, and organization, hazardous waste management, safety management, and purchasing of lab supplies.

Research Interests

Use Molecular Dynamics (MD) simulations to identify key residues involved in protein-protein interactions and predict effects of mutations at these sites.

Publications

"The Binding and Specificity of Chemokine Binding Proteins Through the Lens of Experiment and Computation" L. E. Stark, W. Guan, M. E. Colvin, and P. J. LiWang, Biomedical Journal (submitted)

"Hypermutation of DPYD Deregulates Pyrimidine Metabolism and Promotes Malignant Progression" L. Edwards, R. Gupta, and F. V. Filipp, Mol Cancer Res, 2016 Feb. doi: 10.1158/1541-7786.MCR-15-0403

"Self-assembled nanoparticle micro-shells templated by liquid crystal sorting" A. R. Rodarte, B.H. Cao, H. Panesar, R.J. Pandolfi, M. Quint, L. Edwards, S. Ghosh, J.E. Hein and L.S. Hirst, Soft Matter, 2015, doi: 10.1039/C4SM02326A

"An Analytic Toolbox for Simulated Filament Networks" R. J. Pandolfi, L. Edwards, and L.S. Hirst, Mrss14-1688-y05-18 Spring 2014 MRS proceedings, 2014

"Designing Highly Tunable Semiflexible Filament Networks" R. J. Pandolfi, L. Edwards, D. Jonston, P. Becich, and L. S. Hirst. Phys. Rev. E., 2014, doi: 10.1103/PhysRevE.89.062602

Awards

Graduate Dean Dissertation Fellowship Spring 2021
Fellowship for final semester of graduate school to assist in completion of degree

NSF-GRFP – Graduate Research Fellowship Program September 2015
Three year fellowship supporting graduate students in the STEM fields funded by NSF

Presentations

Biophysical Society Meeting – Poster Presentation February 2020, San Diego, CA
Presentation "*Determining Factors that Influence vCCI Loop Interactions in vCCI-Chemokine Binding Through MD Simulation"*

Biophysical Society Meeting – Poster Presentation February 2018, San Francisco, CA
Presentation "*Evaluating Protein-Protein Interactions in Chemokine-Inhibitor Complexes Using MD Simulation"*

Conferences

Biophysical Society Meeting February 2020, San Diego, CA
Annual meeting for Biophysical Society

Biophysical Society Meeting February 2018, San Francisco, CA
Annual meeting for Biophysical Society

The Cancer Genome Atlas (TCGA) Conference May 2015, Bethesda, MA
Annual meeting at NIH on progress and findings associated with TCGA project

Service

Graduate Student Association (GSA) August 2018 – May 2019, Merced, CA
Position: Treasurer – Manage GSA budget and funding opportunities for graduate students. Serve as GSA representative on Student Fees Advisory Committee

Student Fees Advisory Committee August 2017 – May 2019, Merced, CA
Position: Graduate Student Representative – Meet with campus departments that receive funds from student fees to review how funds are spent and provide suggestions on allocating funds in the future.

NSF-CREST CCBM Student Leadership October 2017 – May 2018, Merced, CA
Plan events to increase member participation and interaction, meet with Center directors and staff to address student issues and ideas to improve the Center.

Graduate Student Association (GSA) August 2015 – May 2016, Merced, CA
Position: Academic Affairs Officer – Coordinate inclusion of and updates from graduate students on committees throughout campus; share updates from committees with GSA. Serve as Graduate Student Representative on the Committee on Academic Planning and Resource Allocation (CAPRA)

QSB Graduate Student President June 2015 – May 2016, Merced, CA
Serve on the QSB Executive Committee as the graduate student member (since March 2015); chair QSB Social Wellness Committee to plan community building events (since April 2015); serve on GSA Delegate Assembly; assign and coordinate with graduate members on QSB committees

Abstract

Viruses have evolved multiple strategies to suppress host immune responses to improve their survival, including expressing proteins that act to mimic or inhibit components of the immune machinery. One protein produced by poxviruses, viral CC chemokine inhibitor (vCCI), is capable of binding nearly all human CC chemokines at low nanomolar to picomolar concentrations. vCCI sequesters the chemokines, blocking interaction sites necessary for binding to their cognate receptors, and thus disrupting immune signaling. Understanding vCCI's remarkable ability to bind specifically to such a diverse set of chemokines would increase our knowledge of both the immune system and viral strategies to evade it. Additionally, the ability to engineer vCCI analogs could open the door to a new class of anti-inflammatory drugs.

We used MD simulations of vCCI bound to several CC chemokines and the herpesvirus HHV8 decoy chemokine vMIP-II to reveal how vCCI manages both specificity and breadth of its interactions with members of the CC chemokine family. Along with key hydrophobic interactions and salt bridges between vCCI and the chemokines, we identified an additional beta strand formed along the CC chemokine N loop, which has previously been found for other chemokine binding proteins but not reported for vCCI. Further building on these models, we simulated published vCCI mutations, such as the Y80A loss of function, to identify a plausible cause of the loss of function through acidic loop collapse due to the loss of the bulky residue. Finally, we used these results to generate new hypotheses of other mutations to "tune" the binding of vCCI, to be tested by further experiments.

1 Background

1.1 Chemokines

Chemokines are chemotactic cytokines, a family of small secreted proteins (~70 amino acids) that recruit cells of the immune system to sites of injury or infection[1]. Damaged or infected cells secrete chemokines which form a concentration gradient by binding glycosaminoglycans (GAGs) on the cell surface. The chemokines are then recognized by their cognate receptor present on immune cells, guiding the cell to the site of injury or disease. The timing and type of cell needed are crucial to the healing process, while overactivity or improper timing has been linked to autoimmune diseases.

The family of chemokines, which includes around 50 different proteins in humans, is further divided into four subfamilies (C, CC, CXC, and CX3C) based on the spacing of two conserved cysteines near the N-terminal (see Figure 1.1a for examples of CC and CXC chemokine sequences). Chemokines are named based on their subfamily, followed by L to indicate they are a ligand, and numbered to differentiate them (e.g. CCL4). Despite the high variability in their amino acid sequences, chemokines share a conserved tertiary structure of a three-strand antiparallel beta sheet and a C-terminal alpha helix (Figure 1.1b). Figure 1.1c shows the key binding features of chemokines, which include the N-terminal tail, the N loop (between the conserved cysteines and beta strand 1), the 30s loop (between beta strands 1 and 2) and the 40s loop (between beta strands 2 and 3). The N loop is key for receptor binding affinity, while the N-terminal tail is required for receptor activation[2–5].

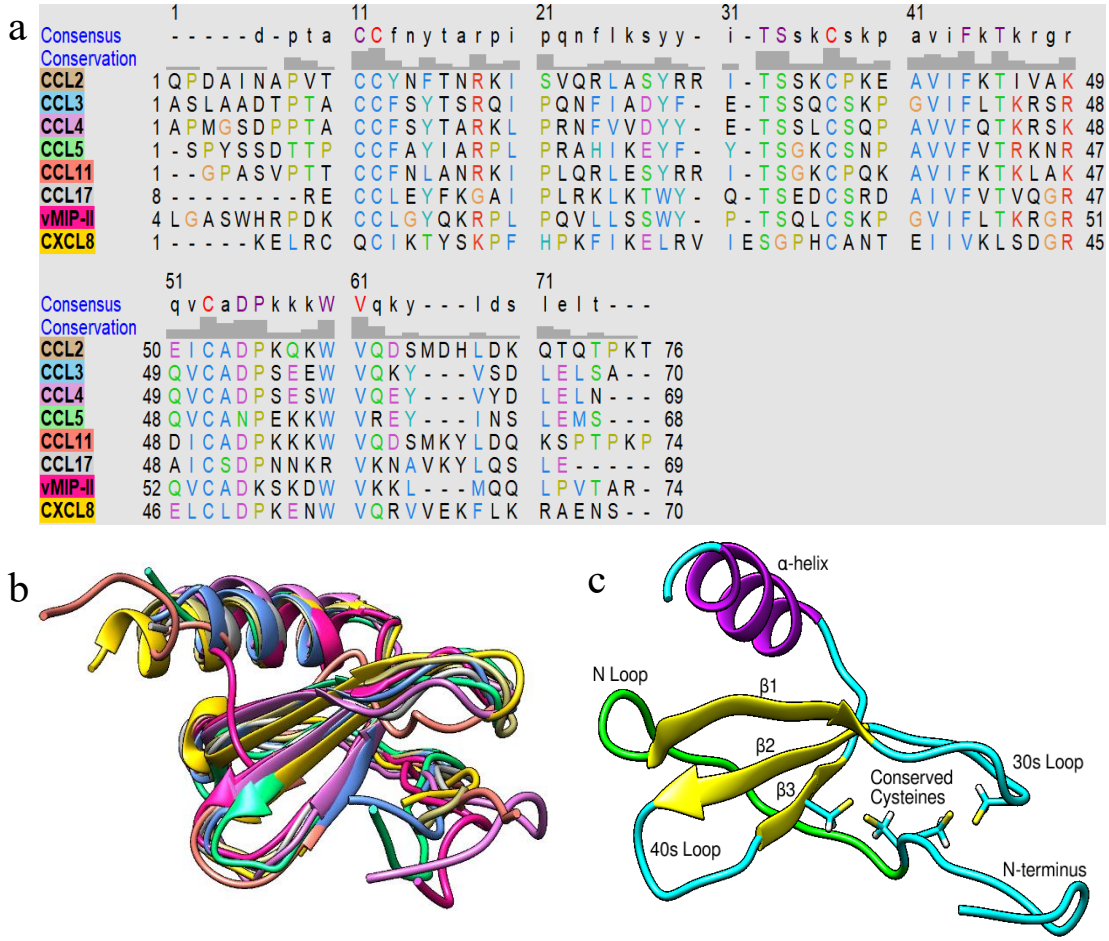


Figure 1.1: Chemokine sequence alignment and structure comparison. a) Sequence alignment for six CC chemokines, a CXC chemokine, and a viral decoy chemokine. The consensus sequence at the top shows the amino acid most present at that position, or a dash for gaps. Capital letters represent all (red) or nearly all (purple) residues are conserved at that position. The height of the gray bars in the conservation header indicate what fraction of the sequences the residue is conserved. b) Chemokines from (a) with backbone atoms superimposed to show similarities in secondary structure. Chemokines are colored matching labels in (a) The PDB IDs for the structures are as follows: CCL2: 3IFD, CCL3: 2X69, CCL4: 2FFK (vCCI with CCL4 bound), CCL5: 1RTO, CCL11: 1EOT, CCL17: 1NR2, vMIP-II: 1VMP, and CXCL8: 5D14. c) Key structural regions of a CC chemokine are labeled on the structure of CCL4. The N loop is colored in green to better show the region it covers between the conserved cysteines and the first beta strand.

1.2 Viral CC Chemokine Inhibitor (vCCI)

Many viruses produce inhibitors and decoys to evade detection by the host immune system. Some viruses create decoy chemokines (e.g. herpesvirus HHV8 decoy chemokine vMIP-II) which bind chemokine receptors to block the binding site for chemokines[6,7]. Others secrete chemokine binding proteins, such as poxvirus viral CC chemokine inhibitor (vCCI), to disrupt the chemokine gradient and prevent chemokines from binding to their receptors.

vCCI (also referred to as 35kDa, p35, or T1) is a soluble chemokine inhibitor that has been shown to be able to bind up to 80 different CC chemokines across multiple species, but not to bind other classes of chemokines[8]. vCCI binds the chemokines at a 1:1 ratio at picomolar to low nanomolar concentrations[9–11]. Mutational studies and NMR experiments have shown that vCCI inhibits immune response by sequestering chemokines and obscuring key residues involved in receptor-binding and GAG-binding[12,13]. vCCI has even been shown to be effective at reducing inflammation in the lungs due to allergen-induced asthma in rodents[14]. It is due to these properties that vCCI has been suggested as a possible treatment for autoimmune diseases and other conditions related to inflammation[15].

vCCI forms a beta sandwich fold with a unique topology so far only seen in some poxvirus immune evasion proteins, such as A41 and the SECRET domain of CrmB[16]. Beta sheet 2 and the long acidic loop between beta strands 2 and 3 are both key in forming contacts between vCCI and the chemokines (Figure 1.2c). While the length of the loop varies between poxvirus species, with cowpox vCCI having 14 residues to

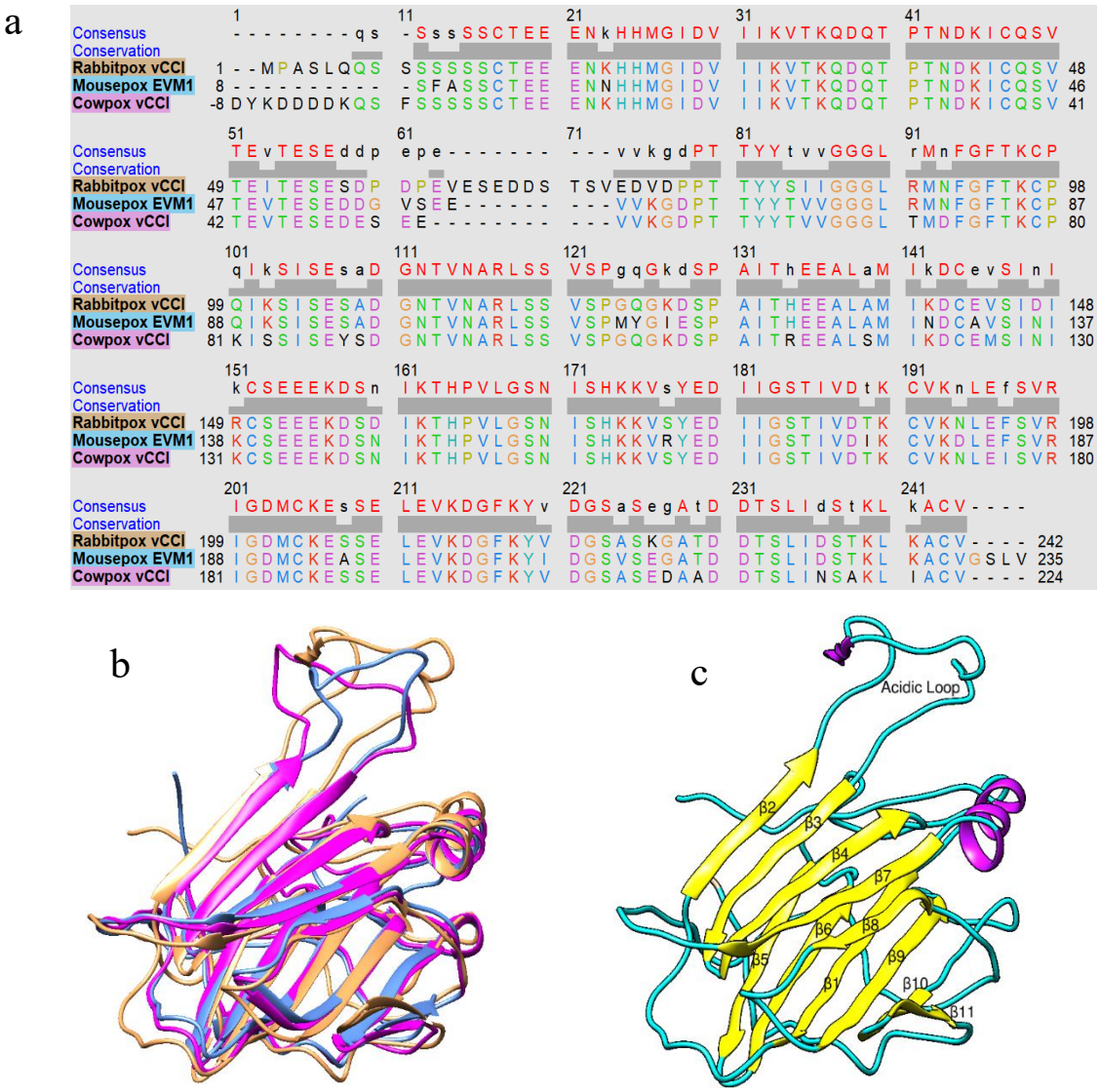


Figure 1.2: Sequence and structure of vCCI. a) Sequence alignment of rabbitpox vCCI, mousepox EVM1 (vCCI homolog), and cowpox vCCI. The sequence is highly conserved, as indicated by the red capital letters in the Consensus header, and the height of the gray bars in the conservation header. b) Structures of the viral inhibitors with backbone atoms superimposed to show structural similarities. Coloring of the inhibitors matches the labels in (a), with rabbitpox vCCI in tan, EVM1 in blue, and cowpox vCCI in purple. PDB IDs as follows: rabbitpox vCCI: 2FFK, mousepox EVM1: 2GRK, cowpox vCCI: 1CQ3. c) Rabbitpox vCCI with acidic loop and beta strands labeled, showing topology of the beta sheets, which is unique to poxvirus proteins. Beta sheet 2 of the beta sandwich fold is shown on top, with beta sheet 1 on the bottom of the vCCI structure.

rabbitpox (vCCI) having 25 residues, the loops are all composed of ~50% acidic residues. This high concentration of acidic residues allows the loop to find complementary charges on the chemokine, such as the conserved basic residues in the N loop and 40s loop (See Figure 1.1).

1.3 Molecular Dynamics

Molecular Dynamics (MD) simulations provide a unique perspective into the structure and function of proteins and other biomolecules. Utilizing structures obtained through X-ray crystallography, NMR, and cryo-EM, MD simulations have been used to predict possible conformations and motions, compare potential costs and benefits of a mutation, and guide the design of small molecules to inhibit or enhance binding[17,18]. As computing power becomes more affordable and available, MD simulations become a more accessible tool for studying proteins. Additionally, improvements in the field are resulting in more accurate models and simulations able to reach longer time scales than previously possible[19–21].

The purpose of the MD simulation is to calculate the motions of the atoms within the system over a set period of time, typically nanoseconds to microseconds. The type of MD simulation to use depends on what types of questions are being addressed based on these motions. Classical, atomistic MD simulation uses an all-atom approach, with each atom within the system being represented as a sphere and all bonds as springs. With this simplification, classical Newtonian equations can be used to calculate the forces and positions of the atoms at each timestep. Two other commonly used types of simulations exist to probe other scales for proteins: quantum and coarse-grained. Quantum MD

simulations zoom in on the electron densities and are capable of modeling bond formation and breaking, but require lots of computational time and power, limiting the size and simulated time of simulations. Quantum MD simulations have been used for many systems, such as understanding proteolytic enzyme activity and inhibition[22] or effects of DNA methylation[23]. Conversely, coarse-grained simulations are used to calculate much larger complexes by combining multiple atoms into a single group, from a few atoms, such as the side chain of an amino acid, to hundreds or more, such as domains or whole proteins. This method is ideal for modeling large protein complexes that would be unfeasible through other methods, but loses some accuracy by approximating aggregate behaviors of atoms and molecules. These larger systems, such as the nuclear pore[24], actin polymer chains[25], or the cell membrane[26], have been successfully modeled to provide a better picture of how they function.

Another factor to consider for MD simulation is the timescale. Protein motions of interest can range from nanoseconds to minutes, so choosing the right model is critical. These motions can include resolving poor contacts and loop fluctuations (ns to μs) to protein folding (μs to seconds)[27,28]. Classical MD simulations typically use a timestep of 2fs to resolve the fastest motions (hydrogen bond stretching). Quantum MD requires a much smaller timestep and more complex equations to resolve subatomic motions, while coarse-grained simulations can resolve longer timesteps depending on how the model is built. The larger the difference between the timestep and the desired timescale, the longer a simulation takes to run and the larger the datafile will be. As computers continue to improve in both computational power and storage, longer timescales can be reached at each level of simulation[29].

For this work, we used Classical MD simulation (from here forward referred to as MD simulation) to observe feasible conformations and interactions of the modeled proteins. Based on the small size of the system and the types of interactions we are observing, this scale is ideal for modeling both the size (~4600 atoms in the complex) and time scale (1μs) of the protein-protein interactions of vCCI and chemokines.

1.4 Simulation Set-up

MD simulation begins with a starting structure. The structure for rabbitpox vCCI by itself and with chemokines bound was based on the NMR structure of rabbitpox vCCI bound to a mutant of CCL4 K45A/R46A/K48A (referred to as CCL4 mutant from here forward) (PDB ID: 2FFK)[13]. Chemokine structures were obtained from the Protein Data Bank[30] with the following PDB IDs: 1VMP (vMIP-II), 2FFK (CCL4, wildtype and mutant), 1RTO (CCL5), and 1NR2 (CCL17). Chemokines were docked onto vCCI by aligning the desired chemokine Cα backbone to the CCL4 mutant Cα atoms on the 2FFK structure. The structure for CCL11 bound to vCCI was determined by a combination of NMR and the docking program HADDOCK[11].

Mousepox EVM1 structure is from PDB ID 2GRK[31]. Since no structures exist with EVM1 with a bound chemokine, EVM1 was first aligned to the Cα atoms of vCCI in 2FFK in conserved regions. After EVM1 was positioned with the CCL4 mutant, other chemokines were aligned similarly to how they were with vCCI.

All mutations to vCCI, EVM1, and chemokines were added using the pyMOL[32] mutagenesis tool to create the new structures. Simulations of mutated inhibitors with a chemokine docked used the same starting structure as the wildtype simulation.

All simulations were run using GROMACS 5.0.7[33] with the AMBER99SB-ILDN forcefield[34]. All proteins were solvated with the TIP3P water model. Na+ and Cl- ions were added to the solvent to set the net charge of the system to zero and the ionic concentration to 70mM to mimic cellular conditions. All simulations go through several steps of energy minimization to relax unfavorable contacts before the start of the simulation. All simulations were run at 300K and 1 bar, with a timestep of 2fs. Unless otherwise stated, simulations of vCCI and EMV1 with a chemokine ran for 1μs, while simulations with vCCI and EVM1 alone, both wildtype and mutated, ran for 2μs.

1.5 Analysis Tools

Several tools were used to analyze the trajectories for the simulation. Table 1.1 includes a list of the types of calculations used, a brief description of the calculation, and the purpose of the calculation.

1.5.1 GROMACS tools

RMSD, RMSF, Mindist, and DSSP[35] were calculated using the tools provided in GROMACS 5.0.7. RMSD calculations were based on the Cα backbone atoms, while RMSF and DSSP were based on all protein atoms. Mindist atom selections were designated by an index file based on what was being tracked (acidic loop, R80, R149).

1.5.2 Hydrogen Bonding

Hydrogen bonding and salt bridge counts were obtained using the web application PDBePISA[36] at https://www.ebi.ac.uk/pdbe/pisa/. PDB files were generated every 2ns from each simulation trajectory and submitted to PDBePISA using the Selenium package for Python3[37] to automate the process.

Table 1.1: Analysis tools

Tools	Description
GROMACS tools	
RMSD Root Mean Squared Deviation	For each frame, the designated atoms of the reference and current structure are first superimposed to minimize distances, then the sum of the root of the average distance between the atoms is calculated. This value is used to compare how much the structure changes over time, as well as see when the system reaches an equilibrium - the RMSD is consistent for some time.
RMSF Root Mean Square Fluctuation	RMSF is calculated for each residue as the RMSD of the residue, averaged over the course of the simulation. Residues in regions of high flexibility, such as the N or C terminal or loop regions, will have higher RMSF compared to residues in stable secondary structures like beta strands. This analysis is useful for comparing the stability of the protein, seeing if secondary structures are denaturing or stable.
DSSP	DSSP assigns the most likely secondary structure based on geometrical parameters drawn from previously determined protein structures in the Protein Data Bank. This analysis is useful for determining the stability of the protein, whether the secondary structure is stable or fraying.
Mindist	Mindist tracks the shortest distance between two selected groups of atoms. This was used in the context of this project to track loop positioning compared to a reference amino acid on the opposite end of beta sheet 2 to observe loop collapse in mutants.
PDBePISA analysis	
Hydrogen bonds	The heavy atoms of the acceptor – donor pair must be less than 3.89Å from each other, with a A-H-D angle between 90-270° to be considered a hydrogen bond.
Salt bridges	Salt bridge designation uses the same parameters to those used for hydrogen bonds, with the exception of the distance between the heavy atoms of the acceptor – donor pair set to 4Å
BSA, ASA Buried Surface Area, Accessible Surface Area	A probe is rolled over the surface of the structure to estimate the surface area. The probe radius for BSA and ASA is 1.4Å. ASA is determined by calculating the surface of each protein chain individually. BSA is determined by calculating the surface of the whole complex and taking the difference of it to the ASA.

1.5.3 Interface Contacts

Interface contacts were calculated in Python to select a residue and find all residues on the other chain within 2.8 Å of it, and repeat for each residue. This distance was chosen to be close enough to exclude water molecules, and thus identify hydrophobic contacts that would be missed through other analyses.

1.5.4 Mapping Persistent Interactions

Persistent interactions were plotted in R to show connections and stability of the interactions. For an interaction to be considered persistent, the interaction between the two atoms had to be present for at least 33% of the evaluated time (the last 500ns of the simulation unless stated otherwise). Lines and residue numbers were colored based on the type of interaction: black for salt bridges, red for hydrogen bonds, and blue for interface contacts. The sum of interactions for each type is shown in the lower right of each graph to give a comparable value to evaluate how connected a complex is. This value is the sum of the fraction of the evaluated time the interaction is present for each interaction of that type.

1.5.5 Buried Surface Area (BSA)

Buried surface area (BSA) and accessible surface area (ASA) were calculated using PDBePISA. BSA and ASA for each residue were averaged over the last 500ns to ensure the system has reached an energy-minimized state. Percent BSA was calculated as BSA divided by ASA to normalize peak heights for residue size.

2 MD Simulation of Chemokine-Inhibitor Complexes

2.1 vCCI-chemokine Complexes

Both the breadth and specificity of targets of vCCI make it ideal as a possible anti-inflammatory treatment[8]. Understanding what drives the interactions between vCCI and CC chemokines will allow for better protein engineering of vCCI towards a specific chemokine or even one that it does not bind normally. Experiments have provided great insight into key residues involved in these complexes from both the chemokine[12] and vCCI itself[38]. MD simulation offers a closer view of these structures and how they interact.

Four CC chemokines (CCL4 as both wildtype and K45A/R46A/K48A mutant, CCL5, CCL11, CCL17) and a viral decoy chemokine (vMIP-II) were selected to compare and contrast how vCCI binds each of them. MD simulations showed the same key regions used by vCCI to selectively bind CC chemokines as previously seen in experiments[9,12] (see Figure 2.1), such as the acidic loop of vCCI interacting with the N loop and 40s loop of the chemokines. Additionally, simulation revealed a previously unseen structural change: a beta strand on the chemokine forming alongside beta strand 8 of vCCI.

2.2 Complex Stability

The stability of the secondary structure for each complex can be seen in Figure 2.2. After a few hundred ns to allow the simulated structures to relax into a more favorable conformation, the system shows only minor fluctuations in the RMSD. The RMSF reveals that only the less structured regions of the proteins (the N-terminal, C-terminal,

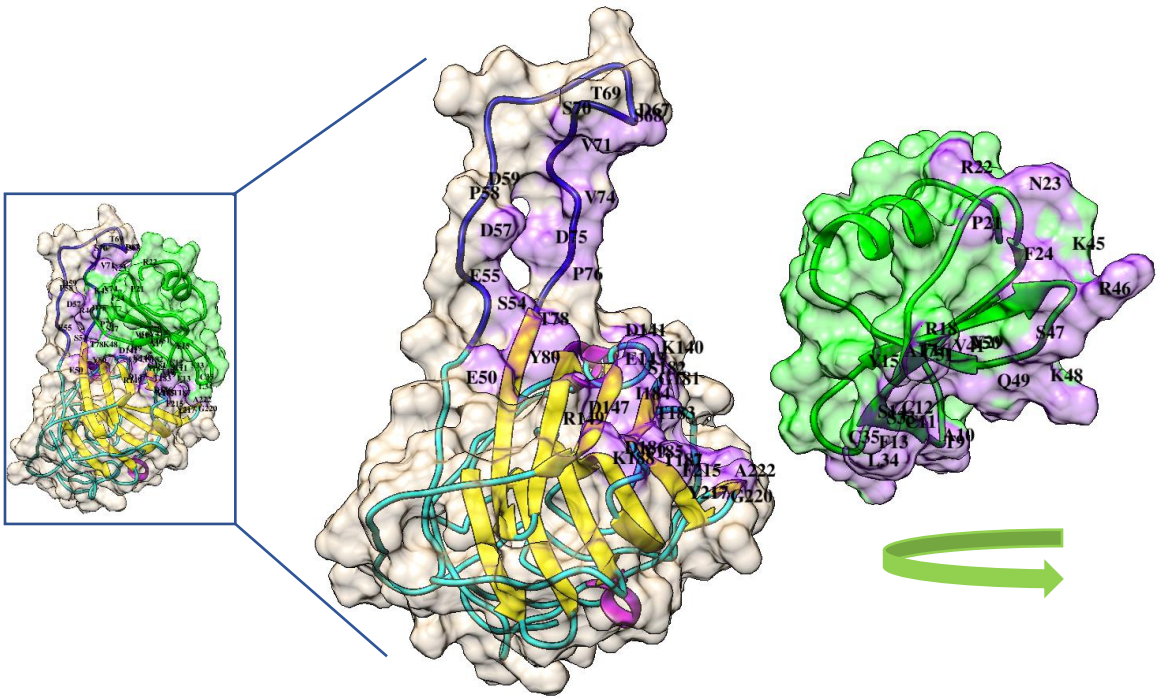


Figure 2.1: Structure of vCCI and chemokine in complex. The structure of CCL4 (green) bound to vCCI (tan) after 1us of simulation, with both ribbon and surface area shown. The interface surface is colored purple on both proteins, and interfacing residues labeled. The smaller image shows the two proteins in complex. The larger image has the two proteins pulled away from each other and CCL4 rotated 180° to reveal the interface on each protein.

and the acidic loop of vCCI) show high fluctuations, meaning the core of the proteins are not denaturing or unfolding (Figure 2.2b). Of particular note, the new beta strand observed in the chemokine N loop in some of the 3D structures after simulation can be seen in the DSSP (Define Secondary Structure of Proteins) analysis[35] (Figure 2.3). Additionally, the beta strand is very stable in most of the simulations, persisting for a majority of the simulated time.

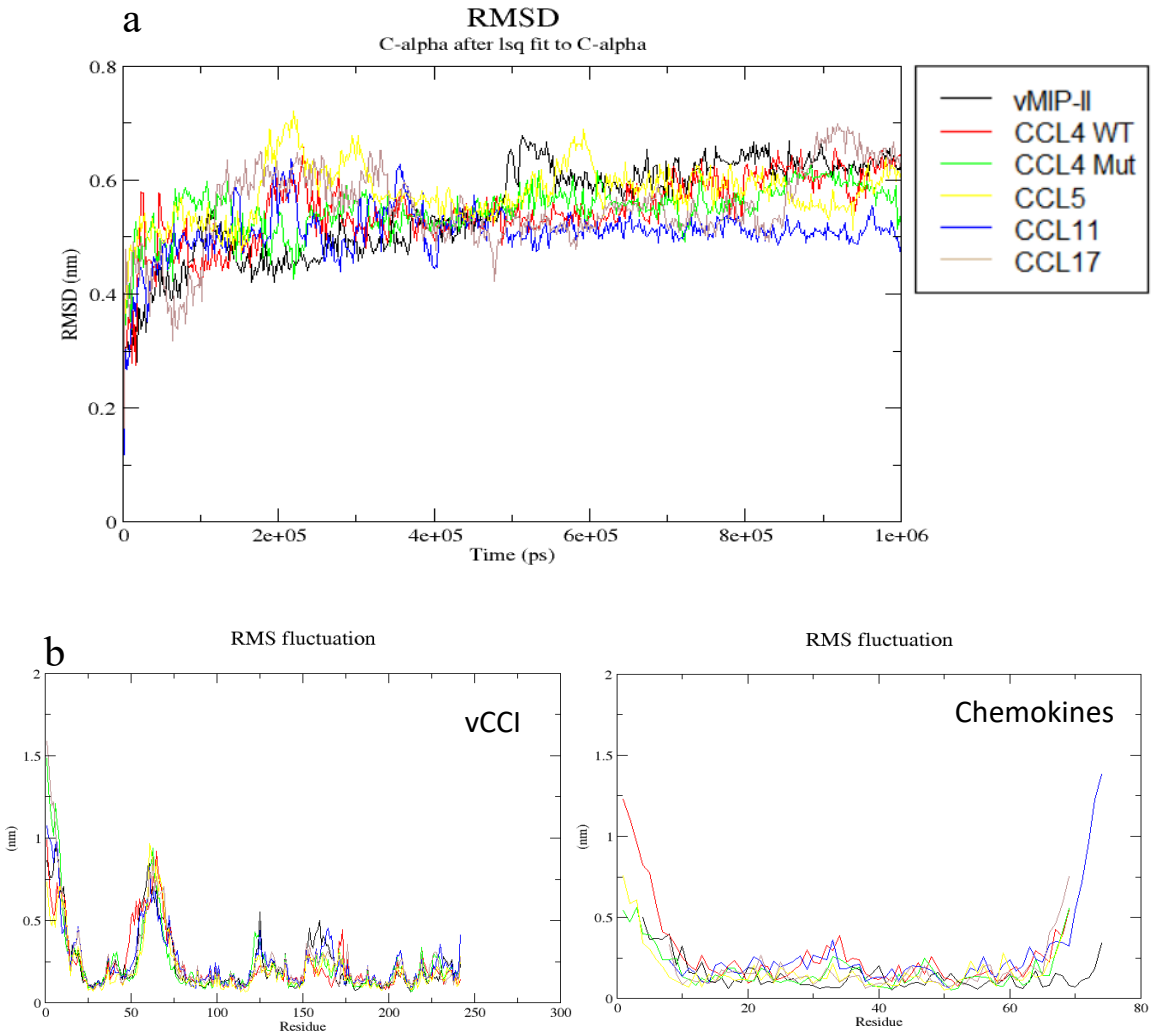


Figure 2.2: Stability of vCCI-chemokine simulations. RMSD (a) and RMSF (b) for each vCCI-chemokine simulation. The lines in all three graphs are colored based on the chemokine used in that simulation, and listed in the legend in (a). (a) The x-axis represents the simulation time and the y-axis represents the calculated RMSD between the first timestep and the timestep at x. The simulations all reach a steady state around 400ns. b) RMSF for vCCI and the chemokines (labeled), with x-axis representing the residue and y-axis the average fluctuation, with peaks revealing the more flexible regions of the protein. The lines follow the same coloring designation as (a).The N- and C- terminal tails, as well as the vCCI acidic loop, have high RMSF due to their flexibility.

The formation of an additional beta strand has been seen in crystal structures of chemokines bound to other virally produced inhibitors, such as ORF virus CKBP with CCL2 (PDB ID: 4ZK9)[39] and Evasin-1 with CCL3 (PDB ID: 3FPU)[40], as well as the oligomer structure of CCL5 (PDB ID: 5COY)[41]. While the NMR structure used as the reference for these simulations (PDB ID: 2FFK)[13] does not identify a beta strand at the chemokine N loop, the simulation of this structure (vCCI-CCL4 mutant)[10] does. Seeing the beta strand form in simulation, when not found in the initial structure, suggests the beta strand formation observed in the other inhibitor-chemokine complexes may not be an artifact of crystallization.

2.3 Hydrogen Bonds and Salt Bridges

The level of hydrogen bonds and salt bridges between the chemokine and the inhibitor can serve as a potential metric of the strength of their interaction. These two types of interactions contribute greatly to the strength and stability of the protein complex[42,43]. Both the number and persistence of these interactions reveal how vCCI is optimized for interacting with the variety of chemokines it binds. Table 2.1 compares the different counts of these interactions across the simulated complexes.

Simulations of vCCI with CCL11, CCL17, and CCL5 show nearly identical levels of average hydrogen bonding at each frame, followed closely by vMIP-II. The simulations with CCL4, both wildtype and mutant, have the lowest average number of hydrogen bonds. The order of simulations based on average hydrogen bonding somewhat reflects what was predicted based on previous experimental results on relative binding affinities[8] ,

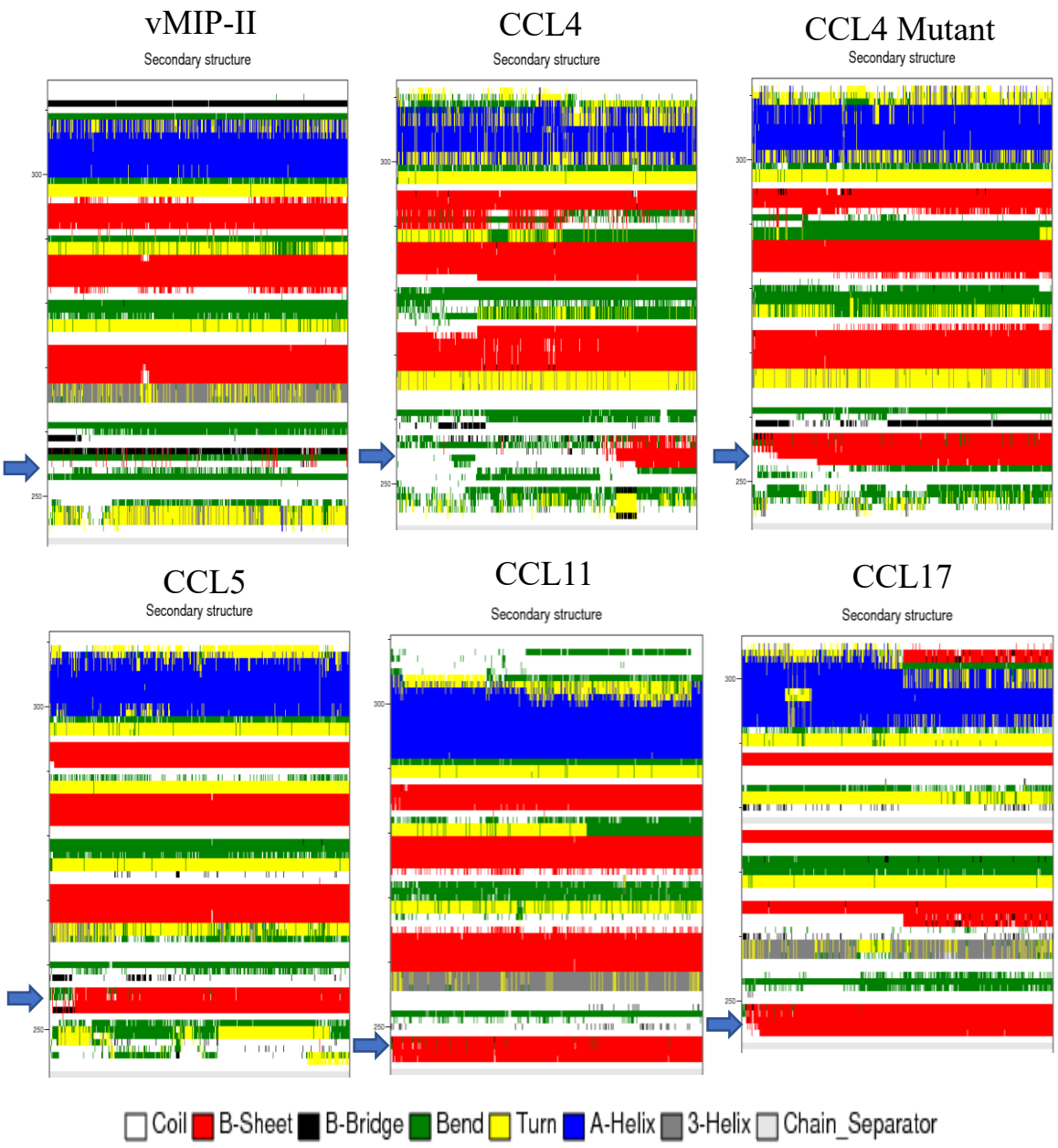


Figure 2.3: DSSP of chemokines throughout simulation. Secondary structure plots for vMIP-II, CCL4, CCL4 mutant (K45A/R46A/K48A), CCL5, CCL11, and CCL17 when bound to vCCI through the course of the simulation. The x-axis represents simulation time and the y-axis represents chemokine residue. The color legend for each secondary structure is shown at the bottom. Arrows on the left of each image indicate where the additional beta strand around residues 8-14 forms upon complex with vCCI. The secondary structure for each 2 ns frame in the MD trajectories was computed using the algorithm Define Secondary Structure of Proteins (DSSP).

Table 2.1: vCCI-chemokine interactions

	CCL4	CCL4 Mutant	CCL5	CCL11	CCL17	vMIP-II
Binding Constants*	1.16 ± 0.17nM	Data below	0.22 ± 0.087nM	0.65 ± 0.17nM	No data	0.06 ± 0.006 nM
Hydrogen Bonds						
Average	20.2302	14.8214	24.7738	25.9365	25.0913	24.3730
Unique bonds	263	81	180	231	278	208
Persistent bonds	18	16	25	24	17	23
Unique residues	28	16	24	25	27	28
Persistent residues	12	9	12	11	9	12
Salt Bridges						
Average	12.8095	7.4722	17.5198	21.4206	18.1667	18.4643
Unique bonds	78	21	79	83	89	73
Persistent bonds	14	9	21	25	23	29
Unique residues	8	3	10	13	12	10
Persistent residues	4	1	4	8	8	7

*Binding constants for CCL4, CCL5, and CCL11 determined by fluorescence anisotropy[11]. The EC_{50} for CCL4 mutant was determined from a quantitative ELISA to compare wildtype and mutant chemokine binding to vCCI (CCL4 wildtype EC_{50} 0.66 ± 0.4nM)[13]. The K_D for vMIP-II was obtained through competition fluorescence anisotropy[10].

The average number of bonds was calculated over the last 500ns of the trajectory. "Unique bonds" is the count of unique interactions of that type identified in the trajectory. Persistent bonds are interactions that are present in at least 33% of the evaluated trajectory (last 500ns). Unique residues and persistent residues are the number of chemokine residues that are involved in the unique bonds and persistent bonds, respectively. These two counts are to see how much of the chemokine forms interactions.

with the exception of CCL17. CCL17 was expected to have much lower hydrogen bonding due to its limited ability to compete off CCL4 in experiments. Additionally, CCL17 has been categorized as a poor target for vCCI to bind due to its sequence missing key conserved basic residues at residue 18, 45, and 46 (CCL4 numbering, see Figure 1.1 for sequence alignment). These interactions will be discussed in more detail below (see Section 2.5).

Looking beyond the average number of hydrogen bonds, we can also look at how long a hydrogen bond is present, or how persistent it is throughout the simulation. An interaction that is maintained over time is better for evaluating binding than more transient interactions that constantly shift, which would indicate a lot of movement or less stability in the complex. Persistent interactions were determined by tracking the number of frames an interaction is present in the trajectory. The cutoff selected for a persistent interaction is it must be present for at least 1/3 of the trajectory. With this criteria, there is a shift in the order of interaction counts, with CCL17 falling between CCL4 wildtype and mutant. This is more in line with what has been shown experimentally. While CCL17 may form lots of hydrogen bonds at each time point, and in fact forms the largest number of unique hydrogen bonds throughout the simulation, the interactions themselves are less stable.

While hydrogen bonds are more common, a salt bridge represents a much stronger electrostatic interaction between two complementary charged residues compared to the dipole-dipole interactions involved in hydrogen bonding. The average number of salt bridges show greater deviation between the simulations, with CCL11 having the most, then vMIP-II, CCL17, and CCL5 around the same range, and finally both CCL4 simulations form the lowest number of salt bridges.

2.4 vCCI-Chemokine Buried Surface Area

The buried surface area of vCCI reveals the regions of the inhibitor that make contact with the bound chemokine. By comparing the different simulations, commonalities appear in the chemokine-inhibitor interface (see Figure 2.4). The most consistent

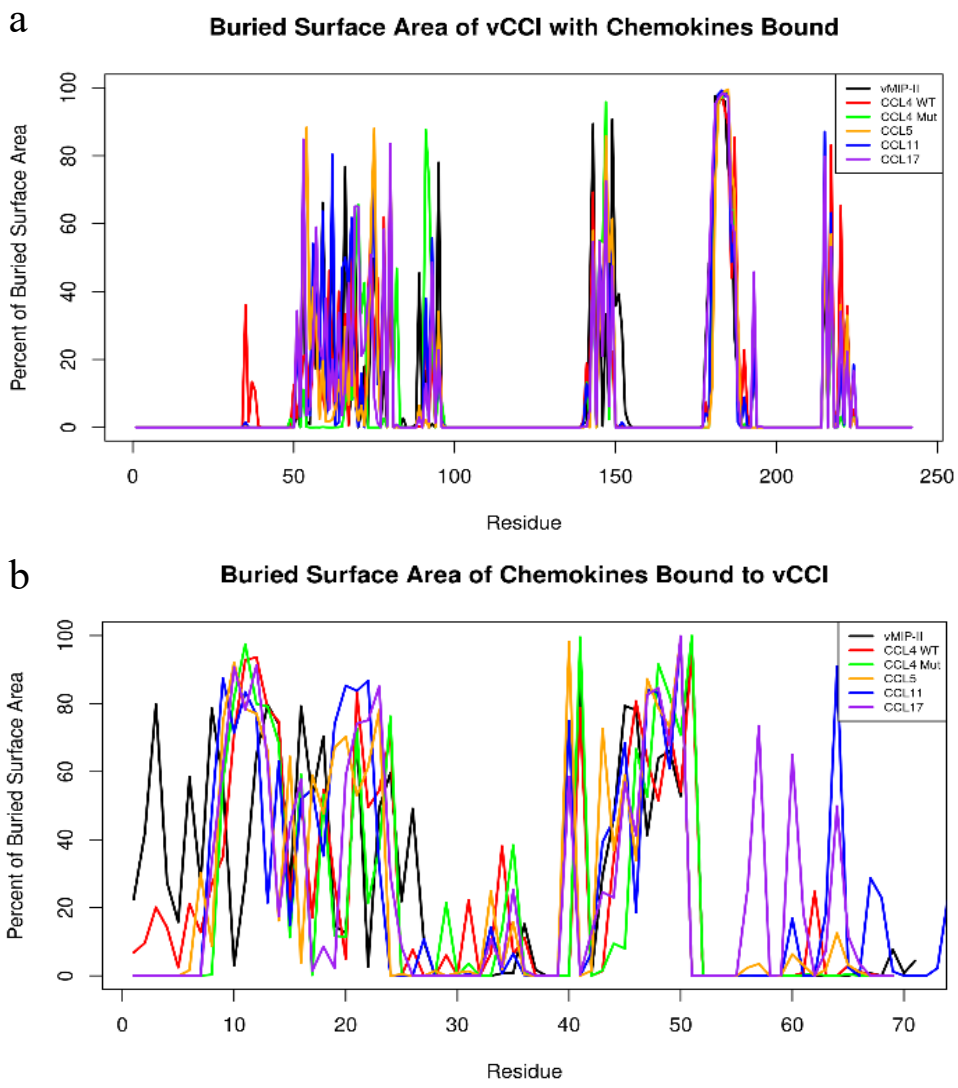


Figure 2.4: Buried surface area of vCCI and chemokine in complex. Graphs show the percent of buried surface area (BSA) for each residue for vCCI (a) and the chemokines (b) when in complex. The x-axis represents the residue for each protein and the y-axis represents the percent BSA, which is the BSA divided by the accessible surface area (ASA) for the residue to normalize for size of the residue. Regions like the peak at 180-186 of vCCI (a) indicate shared contacts across chemokines and likely indicates a region key for binding chemokines.

interaction between vCCI and the chemokines can be seen at residues 180-186 of vCCI (beta strand 8). This region is buried to an almost identical degree for all chemokine complexes simulated, highlighting the importance of this region in the binding of CC chemokines.

Other regions of note for vCCI are residues 88-96 (beta strand 4), 143-150 (beta strand 7), and 214-220 (beta strand 10). Beta strands 4, 7, and 8 are a part of beta sheet 2, which is the binding face the chemokine docks on. Beta strand 10 is a part of beta sheet 1 on the underside of vCCI, which forms a hydrophobic pocket with beta strand 8 that interacts with a large hydrophobic residue conserved on CC chemokines (for example, F13 in CCL4). While these three regions show a lot more variability in the amount of BSA amongst the different chemokines, each region is buried for each chemokine.

The final shared region, with the largest variability amongst chemokines, is residues 50-80, which contains the acidic loop of vCCI (residues 53-77). The flexibility of the loop accounts for some of the variation, as well as what regions of the chemokine the loop can interact with. For example, the CCL4 mutant results in much lower BSA around the loop region at residues 50-65 compared to the other chemokines. This could be due to the mutations in the 40s loop of the CCL4 mutant, which are typically conserved basic residues which were mutated to alanine (K45A,R46A,K48A).

The BSA for the chemokines themselves, when bound, show greater variation due to their sequences, but key conserved regions and interactions can still be identified. Most notably from Figure 2.4b, the 40s loop is buried in all the simulations, from residues 44-52. While the peak positions based on sequence are different (residue 50 peak in CCL5, CCL11, and CCL17, vs residue 52 in CCL4 and vMIP-II), the trend of the BSA is identical, indicating a shared interaction region with vCCI.

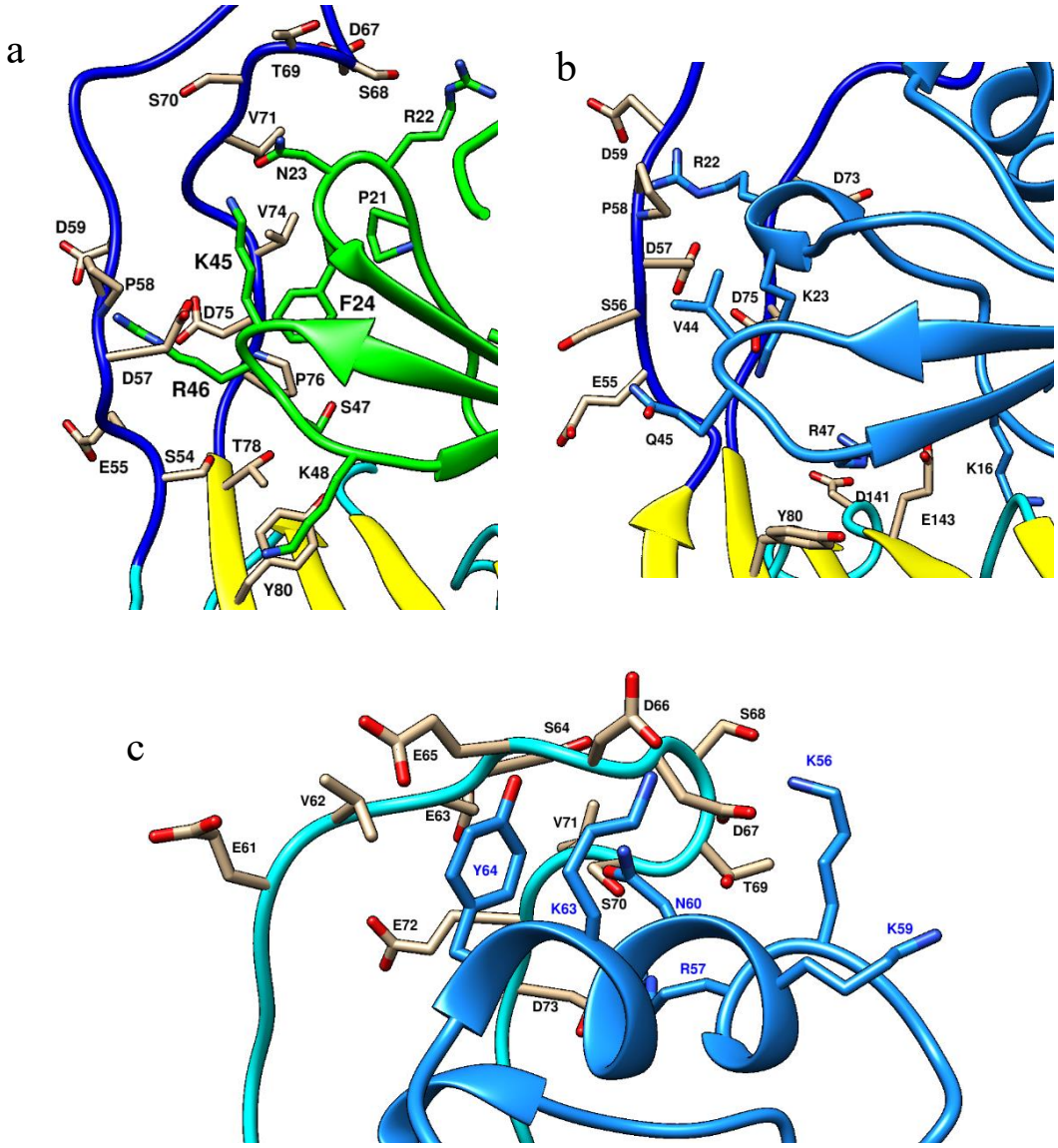


Figure 2.5: Differences in vCCI binding of CCL4 and CCL17. a) The N loop and 40s loop residues of CCL4 (green) interacting with the vCCI acidic loop (blue). The N loop interacts with the upper half of the acidic loop, while the 40s loop interacts closer to the beta sheet 2 face. b and c) CCL17 (light blue) interacts with the acidic loop of vCCI (blue in b, cyan in c). Due to the lack of basic residues in the CCL17 40s loop, the N loop interacts closer to the beta sheet 2 face (b) while the top of the vCCI acidic loop interacts with the CCL17 alpha helix (c).

The residues 8-14 of the human CC chemokines, which contain the conserved cysteines and part of the N loop, also share an increased BSA, while vMIP-II does not follow the pattern of buried residues at this site. This region contains the conserved cysteines that define the CC chemokine family and a large hydrophobic residue (F13 in CCL4). This region forms an additional beta strand with vCCI along beta strand 8, which we see in all simulations with human CC chemokines (See Figure 2.3), but only appears briefly in the vMIP-II simulation.

An interesting difference in BSA appears around residues 55-65, which contains the alpha helix in the chemokine. Here, CCL17 has much higher BSA for some residues (K56, R57, N60, and Y64) compared to most of the other chemokines which show essentially no BSA in this C-terminal helix region. If we look at the last frame for the simulations of CCL17 (Fig 2.5c), we see that the acidic loop of vCCI stretches up to interact with the alpha helix. This extended loop conformation has not been observed experimentally. This could explain why vCCI may not bind CCL17 as well as the other chemokines as this could be a less stable conformation overall to compensate for the missing charges vCCI typically interacts with on a chemokine (residues 18, 24, and 45/46 based on CCL4 numbering). The peaks in the average BSA supports that this interaction of the vCCI loop with the alpha helix of CCL17 is present for an extended time in the simulation, and is not just a transient contact. While the simulation time may not be long enough to conclusively prove the stability of the extended loop interaction, it does suggest that this interaction is a possible conformation that vCCI might use to compensate when a less ideal chemokine is bound.

CCL11 also shows a large BSA in this same region, at residue 64, as well as smaller peaks around 67, 68, and 74. The first three residues are found in the alpha helix of CCL11, buried by the acidic loop of vCCI for part of the simulation, similar to CCL17. Even with the 40s loop interactions at the base of the acidic loop, which positions the alpha helix a little further back than with CCL17, vCCI is still able to make contact for part of the simulation. CCL11 also has a longer C-terminal tail after the alpha helix than the other chemokines, which was able to reach back down to vCCI and interact during the simulation (residue 74).

2.5 Persistent Interactions and Proximity Contacts

Figure 2.6 shows a mapping of the persistent interactions between vCCI and the specified chemokine to better evaluate the residues involved in the interface and the different types of interactions. Persistent interactions are defined as interactions that are present for at least 1/3 of the evaluated simulation time, in this case, the last 500ns of the simulation.

Three types of interactions were evaluated: salt bridges, hydrogen bonds, and interface contacts. The first two were discussed above; the third, interface contacts, identifies all residues within 2.8 Å of another residue on the other protein. This third interaction is used to identify residues in the interface space that might not be involved in other non-bonding interactions, such as those involved in hydrophobic interactions. The distance cutoff of 2.8 Å was chosen based on the average size of a water molecule, meaning residues closer than that distance would prevent water from entering. While these interactions are the weakest in maintaining the complex, identifying these residues helps in mapping the interface space.

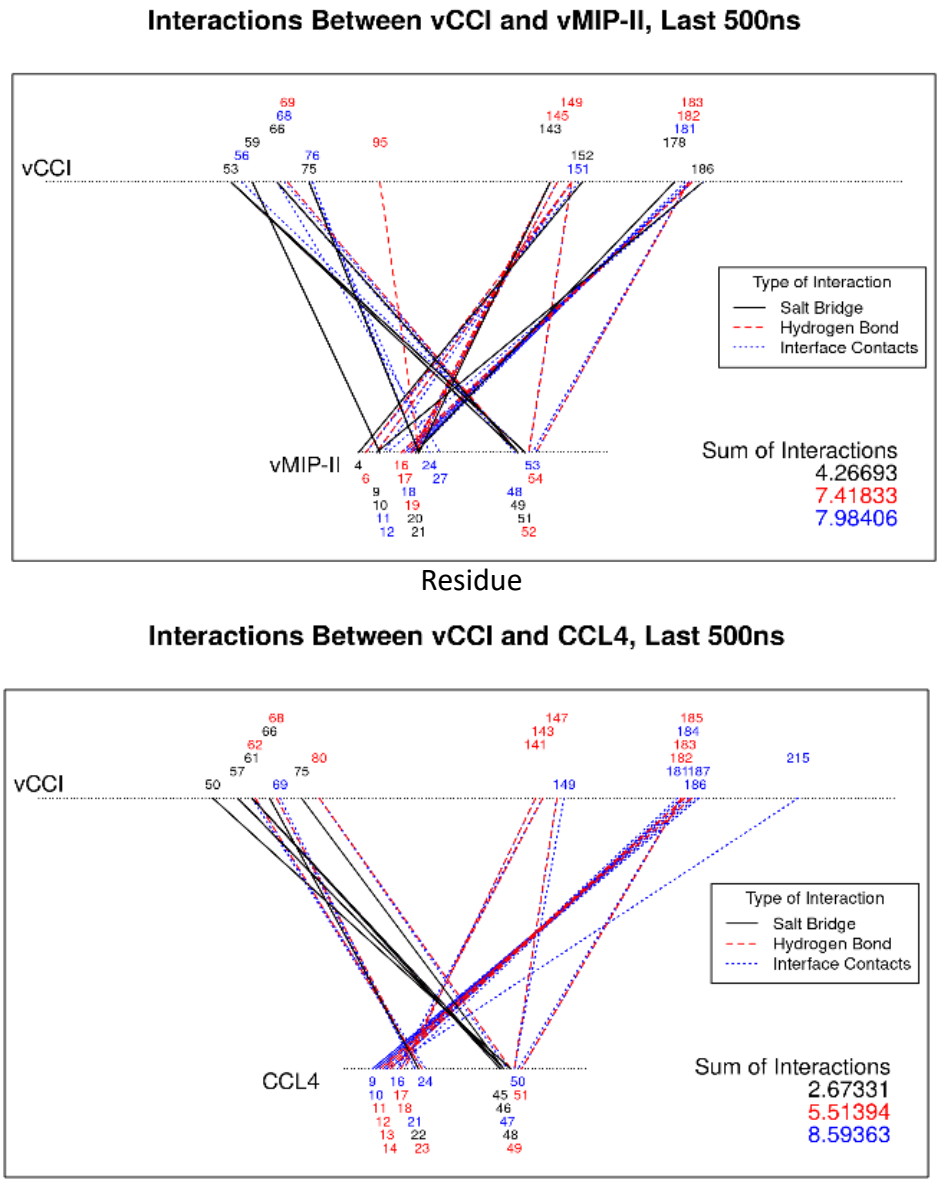
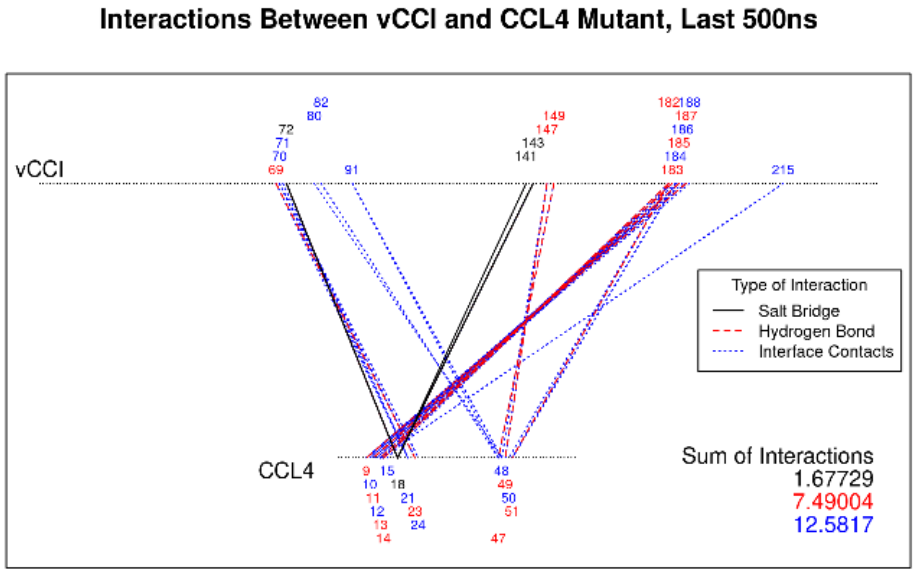


Figure 2.6: Mapping of persistent interactions between vCCI and chemokines. The points comprising the top line represent the residues of vCCI while the bottom points represent the residues of each chemokine. The transecting lines indicate the persistent interactions (present at least 33% of the observed time) between vCCI and the chemokine. The color and style of the line indicates the type of interaction: salt bridge (black solid), hydrogen bond (red dashed), and interface contact (blue dotted).The numbers above and below the horizontal lines list the residue numbers of vCCI and the chemokine, respectively, that are involved in an interaction. The color of the number shows the strongest interaction that residue is a part of, matching the color used for the transecting line. The sums in the lower right of each graph are the sum of the fraction of time during the simulation that the indicated interactions are observed, to provide a quantitative value to compare simulations.



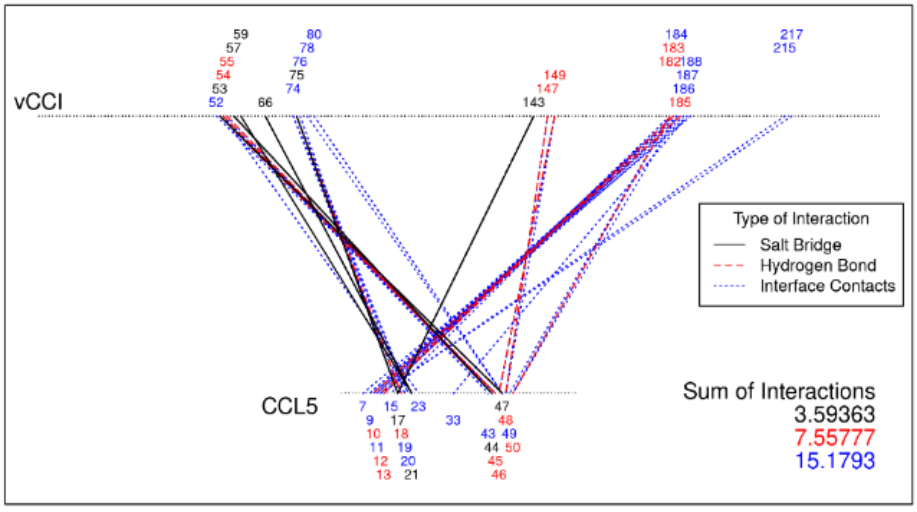
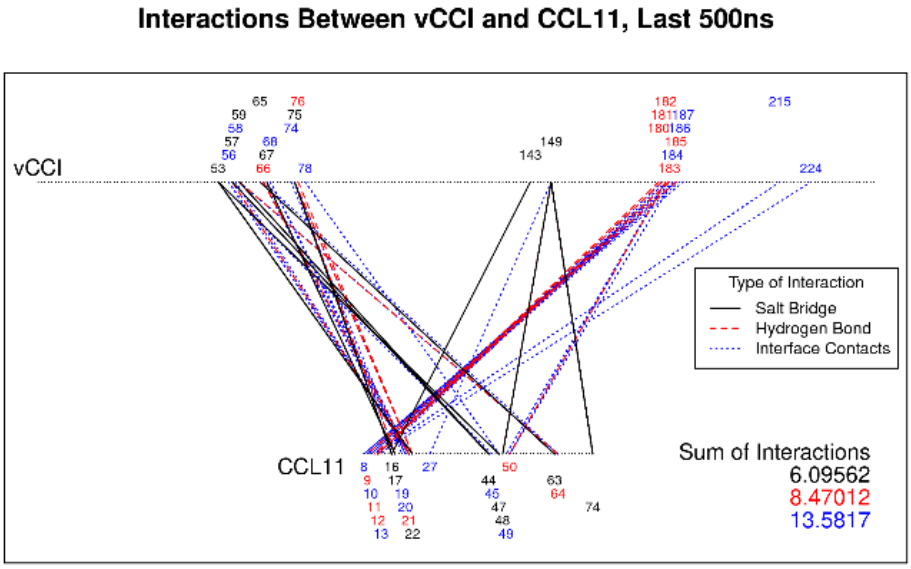


Figure 2.6 continued



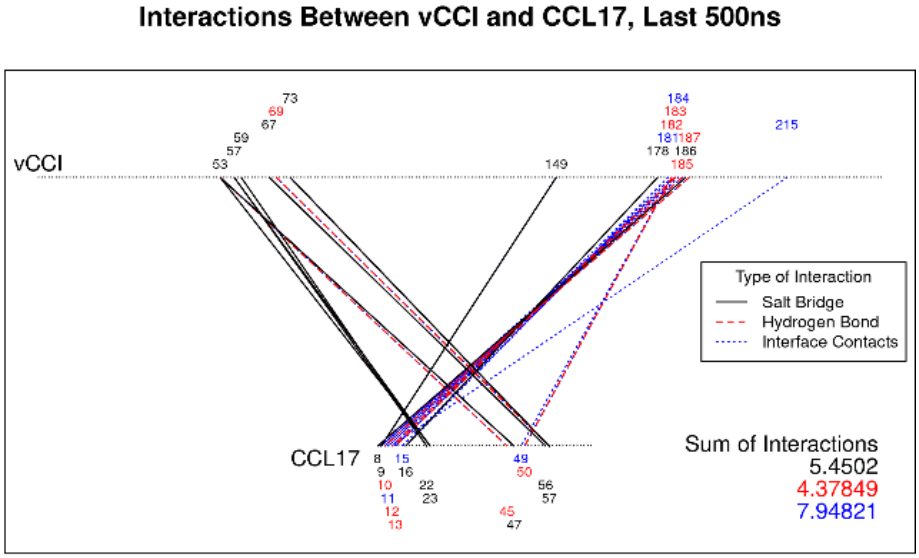


Figure 2.6 continued

Two key groupings of interactions are observed in each of the simulations: beta strand 8 of vCCI (residues 180-186) with the first part of the N loop of the chemokine (residues 8-14) and the acidic loop of vCCI (residues 53-77) interacting with the 20s part of the N loop and the 40s loop of the chemokines (residues 18-24 and 45-51 respectively)

The interaction of vCCI beta strand 8 with the N loop of each CC chemokine is clearly visible in all simulations. As seen in the DSSP and the structures after simulation (Figure 2.3), the N-loop of each chemokine forms an additional beta strand along vCCI beta strand 8. While these connections are less persistent in the vMIP-II and CCL4 wildtype simulations, they are still present and match the general pattern of interactions as seen in the other simulations.

Within the N loop of the chemokine, there is also a conserved large, hydrophobic residue (residue 13/14 depending on sequence) that sits in a hydrophobic pocket created between beta strand 8 and 10 of vCCI (residues I180, T183, V185, F215, and Y217). This pocket has been identified as significant in chemokine binding[9,12]. Among the human CC chemokine simulations, this interaction can be seen as well. While not as persistent as the beta strand interactions, the interactions of residue 13/14 of the chemokine with residues F215 and Y217 of vCCI are most frequently present.

Due to the high flexibility of the acidic loop of vCCI, persistent interactions during the course of a simulation are less expected, yet each chemokine reveals at least some connections to the loop to last over several hundred nanoseconds. These interactions occur with the 20s and 40s loop of the chemokines, where despite low conservation of sequence between different chemokines, key basic and polar residues are found. The

distribution of acidic residues throughout the loop allows it to move and position to best find the complementary charges on the CC chemokine.

In contrast to this, CCL17, which lacks key basic residues in its 40s loop and has minimal persistent contacts in this region, instead shows an interaction with the vCCI loop at residues 57, which is part of the alpha helix. As shown in Figure 2.5c, the vCCI loop is seen as extending to interact with the alpha helix of CCL17. This is in agreement with the buried residues seen with BSA along the alpha helix (Figure 2.4b). This extended loop is not seen in any other simulations, suggesting it may be less energetically favorable, and could explain why CCL17 binds poorly to vCCI.

2.6 EMV1-Chemokine Simulations Compared to vCCI

EVM1, the ectromelia (mousepox) virus homolog of vCCI, shares the same ability to bind many CC chemokines, but does so with an acidic loop almost half the size. Analysis of vCCI suggests the length of its acidic loop allows it to interact with the key residues on chemokines despite their high sequence variability. EVM1 was selected to elucidate the effects loop length has on chemokine binding. Additionally, we want to determine how EVM1 might compensate when binding chemokines with its shorter loop. EVM1, as seen in Figure 1.2b, has the same PIE domain fold as vCCI and shares nearly 80% sequence similarity. The main difference between the two inhibitors lies in the acidic loop, which contains 25 residues in vCCI and 15 residues in EVM1. EVM1 maintains the 50% acidic residue composition as vCCI and other members of the protein family.

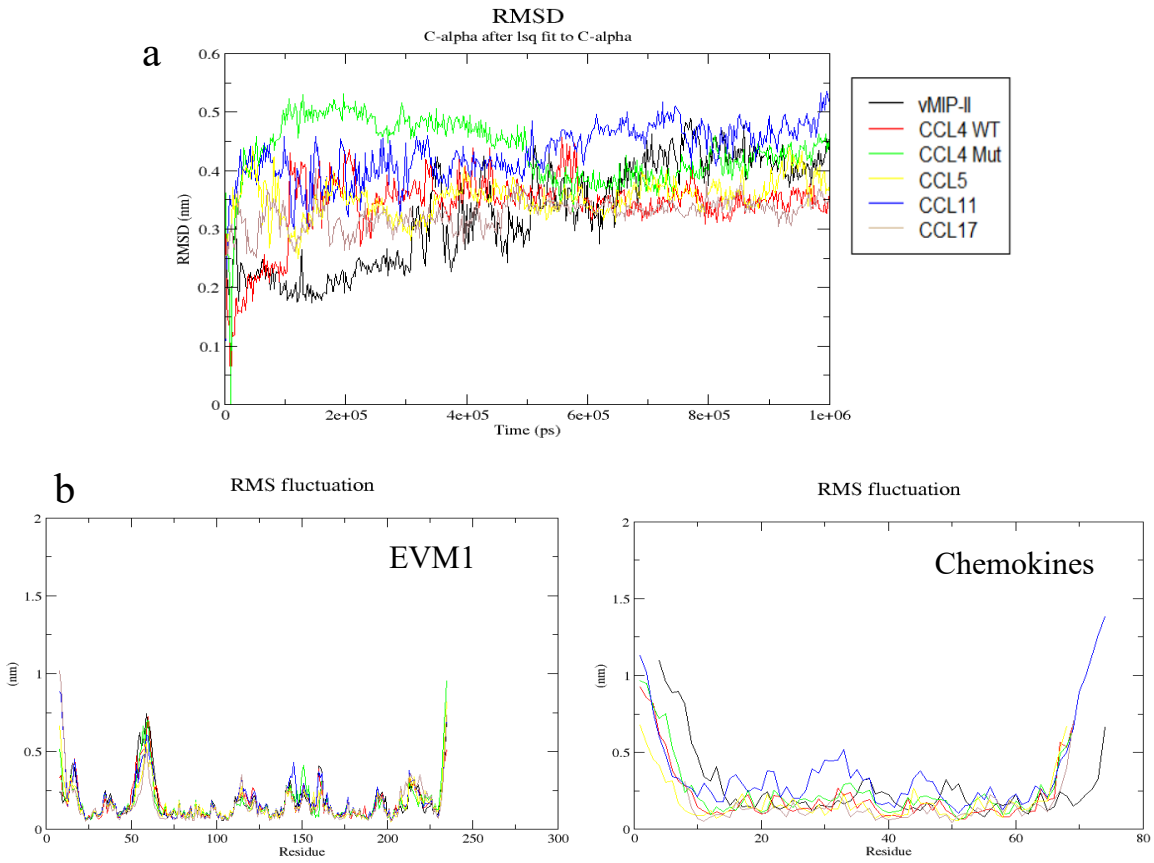


Figure 2.7: EVM1 simulation stability. a) RMSD of EVM1-chemokine simulations. Simulations reach a relative equilibrium around 500ns. The deviation for the simulations is small, around 0.5nm, showing the structures are holding together. b) RMSF for EVM1 and the chemokines. EVM1 has less fluctuation in its N-terminal and loop, which is expected from the smaller size of both compared to vCCI. The first seven residue of the N-terminal are absent in the EVM1 structure. The chemokines retain a similar level of fluctuation, though CCL11 has higher fluctuation in the middle residues.

2.6.1 Complex Stability

The simulations of EVM1 with chemokines were able to settle into a stable conformation quickly, with only minor changes in the RMSD after a few hundred nanoseconds as well (Figure 2.7a). Plotting the RMSF of EVM1 residues (Figure 2.7b) shows that the protein is folded as expected, with structured regions having low fluctuations averaged over the simulation, while less structured regions like the acidic loop are higher. The RMSF for the individual chemokines (Figure 2.7c) show very similar levels of fluctuations,

comparable to their values while bound to vCCI, though CCL11 has slightly higher peaks in its beta strand 1 when bound to EVM1. The N-terminal tail of vCCI has higher RMSF, while the RMSF is higher for EVM1 in its C-terminal tail. CCL11 shows the highest RMSF peaks amongst the chemokines, has even greater mobility around residues 24-36 and 44-48 when bound to EVM1. This greater RMSF for CCL11 could represent weaker binding of the inhibitor, allowing more freedom of motion for the chemokine.

Looking at the DSSP of the chemokines with EVM1 (Figure 2.8), the additional beta strand in the N loop can be seen in each simulation. This interaction is more stable in the EVM1 complex for some simulations, such as vMIP-II and CCL4 wildtype. This could be a result of more favorable starting positions than the prior simulations. The identification of this additional beta strand in all simulations, even with a different inhibitor, offers further support for this structural change forming upon binding.

2.6.2 Buried Surface Area

The peaks of buried surface area appear in the same corresponding regions for EVM1 as they did in vCCI, showing the chemokines did not need to reposition in the simulation to find more favorable contacts (Figure 2.9). CCL5 results in higher BSA compared to other chemokines in several regions on EVM1 (residues 130-136, 170-174, and 204-207), whereas it resulted in identical peaks to the other chemokines for vCCI. CCL5 appears to cover a larger interface space on EVM1 than the other chemokines, covering a few additional residues when bound. The lone peak around residues 32-35 are a result of CCL5 N-terminal tail falling across these residues and remaining, similar to CCL4 with vCCI. For the chemokines, one of the most notable changes in buried surface area is the reduced peaks in the alpha helix for CCL11 and CCL17. This change corresponds to the

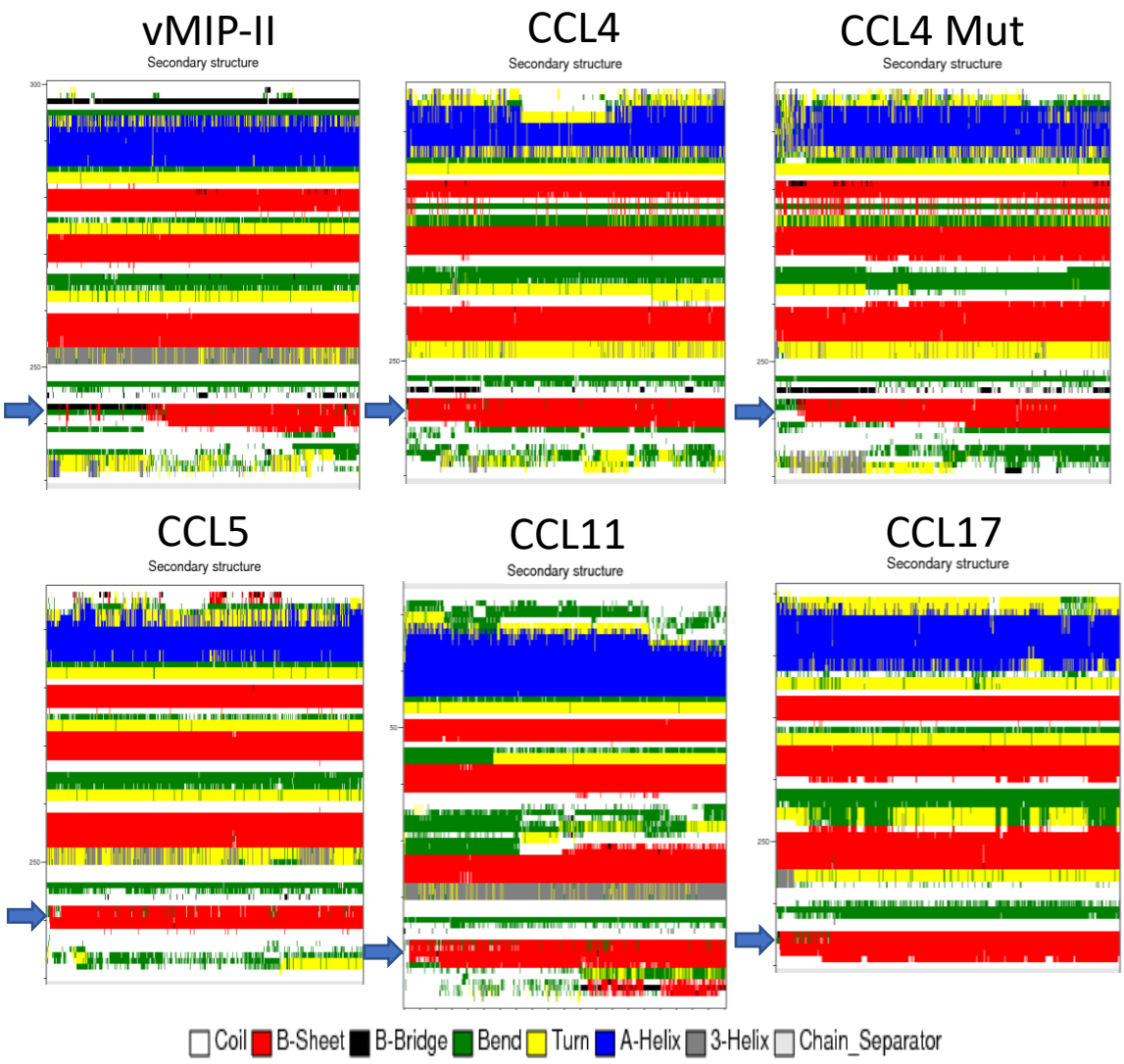


Figure 2.8: DSSP of chemokines bound to EVM1. Secondary structure depiction of chemokines while bound to EVM1.The x-axis represents the simulation time and the y-axis represents the chemokine residue. The blue arrows to the left of each graph point to the additional beta strand formed along the first half of the N loop of the chemokine. The additional beta strand in vMIP-II and CCL4 wildtype are more stable for the EVM1 simulation compared to the vCCI simulations (see Figure 2.3).

Table 2.2: EVM1-chemokine interactions

	CCL4	CCL4 Mutant	CCL5	CCL11	CCL17	vMIP-II
Binding Constants	No data	No data	0.230 ± 0.17nM	No data	No data	No data
Hydrogen Bonds						
Average	17.1753	13.4382	19.6175	18.6335	19.7291	16.5976
Unique bonds	138	137	184	181	102	201
Persistent bonds	17	8	14	12	19	11
Unique residues	21	15	25	25	18	21
Persistent residues	10	7	9	7	9	8
Salt Bridges						
Average	6.9681	5.1873	16.6494	10.2590	13.5299	11.4263
Unique bonds	53	47	80	54	35	85
Persistent bonds	8	4	24	12	18	13
Unique residues	7	3	9	9	6	10
Persistent residues	3	1	5	2	5	4

Binding constant for EVM1 – hCCL5 (human CCL5) determined by SPR (surface plasmon resonance)[31]. This closely matches the K_D found for rabbitpox vCCI – CCL5 in Table 2.1. Other binding constants found by Arnold and Fremont include EVM1-hCCL2: 20 ± 4nM and EVM1-hCCL3: 0.029 ± 0.011nM.

shorter acidic loop of EVM1 making less contact with the alpha helix of the chemokines. The acidic loop either covers a smaller amount of the residue surface or the residue is buried for less time during the simulation.

Another major difference can be seen around the 40s loop of the chemokines, where the percent of BSA is less overall around residues 42-46. The spread of values in this region is greater than with vCCI. Finally, the N loop of vMIP-II, which forms a more stable beta strand during the simulation with EVM1, more closely matches the other chemokine peaks in this region. This is in contrast to the BSA when bound to vCCI, where the N terminal region of vMIP-II did not match that of the other chemokines, reflecting the less stabilized region that did not form the additional beta strand.

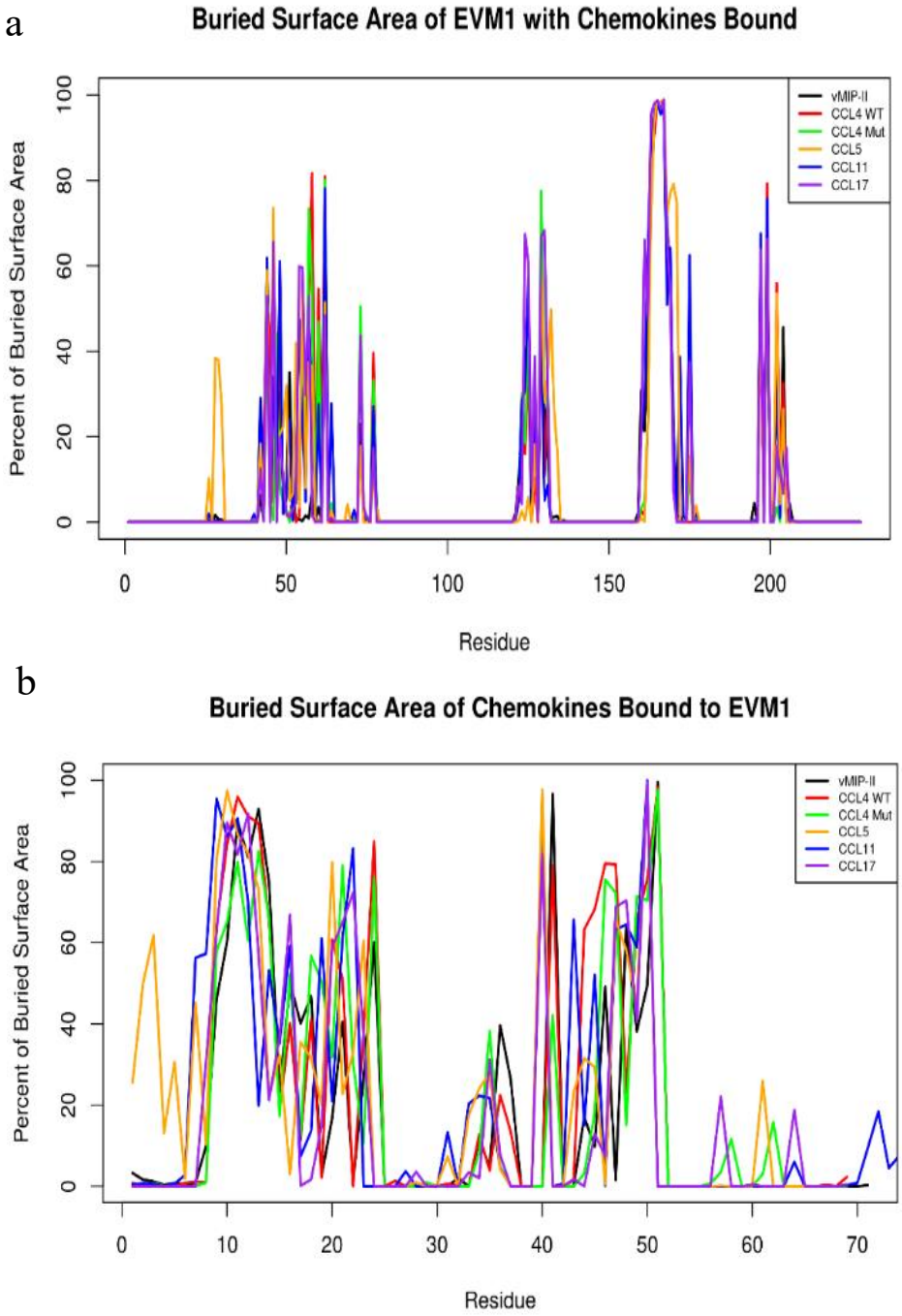


Figure 2.9: Buried surface area of EVM1 and chemokines in complex. The graphs represent the percent BSA for each residue for EVM1 (a) and the chemokines (b), averaged over the last 500ns of the simulations. EVM1 also shows the same shared peak in BSA in the last strand of beta sheet 2 (residues 168-174), similar to vCCI.

2.6.3 Interactions

Binding constants for EVM1 and several chemokines have matched close to some published values for vCCI[31]. However, our simulations of chemokines with EVM1 consistently have fewer hydrogen bonds and salt bridges compared to those for chemokines bound to vCCI. Additionally, the order of simulations based on average hydrogen bonds and salt bridges differs for some chemokines. CCL17 maintains the most hydrogen bonds in each frame, followed closely by CCL5 and CCL11, and finally CCL4 wildtype, vMIP-II, and CCL4 mutant. CCL17 also maintains the most persistent hydrogen bonds, despite the expectation that the shorter loop of EVM1 should make it harder to reach the interactions that were key in CCL17 binding to vCCI.

The shorter loop does greatly reduce the hydrogen bonds that form with the alpha helix, and instead emphasizes interactions with R22 in the 20s region of the N loop. Despite the decrease in average hydrogen bonds, CCL17 actually gains more persistent hydrogen bonds compared to the simulation of it bound to vCCI. Additionally, CCL17 with EVM1 has far fewer unique hydrogen bond interactions, the least among the other EVM1 simulations. This all suggests that CCL17 moves around much less during the simulation when bound to EVM1 despite its lack of key residues utilized by other chemokines to maintain their interactions.

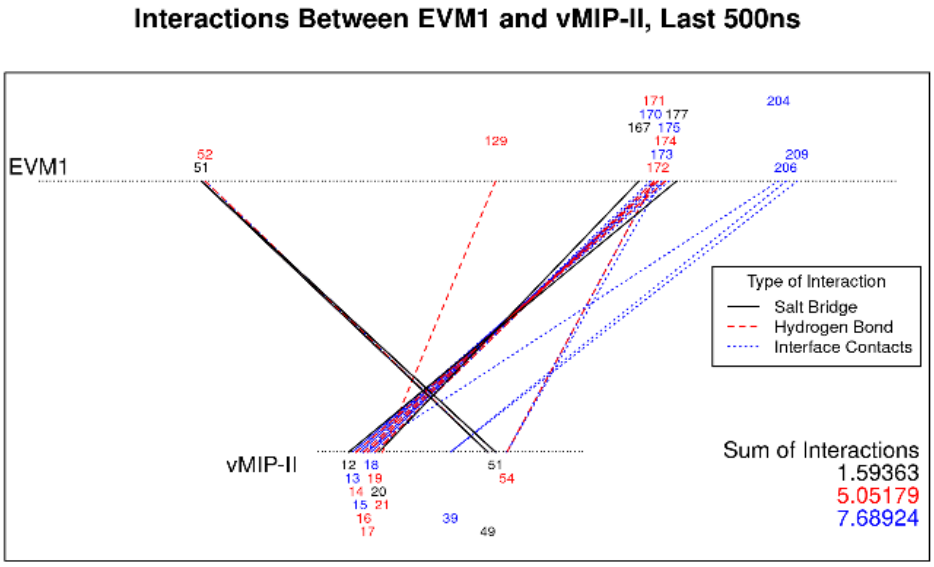
Salt bridges in EVM1 simulations show a similar drop compared to their vCCI counterparts. CCL5 maintains the highest average number of salt bridges, to almost the same level as with vCCI, followed then by CCL17, vMIP-II, CCL11, CCL4 wildtype, and lastly CCL4 mutant. The residues of CCL5 are able to maintain the same salt bridges

formed with the vCCI acidic loop when interacting with EVM1 (residues R17,R21, R44, K45, and R47). In contrast, CCL11, which formed the most salt bridges with vCCI, formed half as many with EVM1. There is a noticeable absence of persistent salt bridges formed along the CCL11 40s loop, which make up half of the persistent salt bridges in the vCCI simulation.

Figure 2.10 shows the persistent interactions between EVM1 and the chemokines. The same two major interactions seen with vCCI are present in the EVM1 simulations as well; the acidic loop (residues 53-65) interacts with the 20s part of the N loop and the 40s loop of the chemokines while beta strand 11 (residues 168-174) and beta strand 14 (residues 204-209), analogous to beta strands 8 and 10 in vCCI, interact with the first part of the N loop of the chemokines. The shorter acidic loop does reduce the overall number of interactions, occasionally only binding one of the two basic loops of the chemokines (40s loop of vMIP-II, 20s region of the N loop of CCL11).

The interactions with vMIP-II and EVM1 seem much weaker than with vCCI. While vMIP-II was able to form the additional beta strand when interacting with EVM1, the contacts with the acidic loop are much fewer, forming only a few persistent interactions along the 40s loop with the acidic loop, and none with the 20s region of the N loop.

Once again, the hypothesis that CCL17 should be a poor target for inhibitor binding does not match with the results of the simulation. Despite the shorter loop, which is unable to reach the alpha helix of CCL17, the chemokine still forms a comparable number of interactions as the other chemokines.



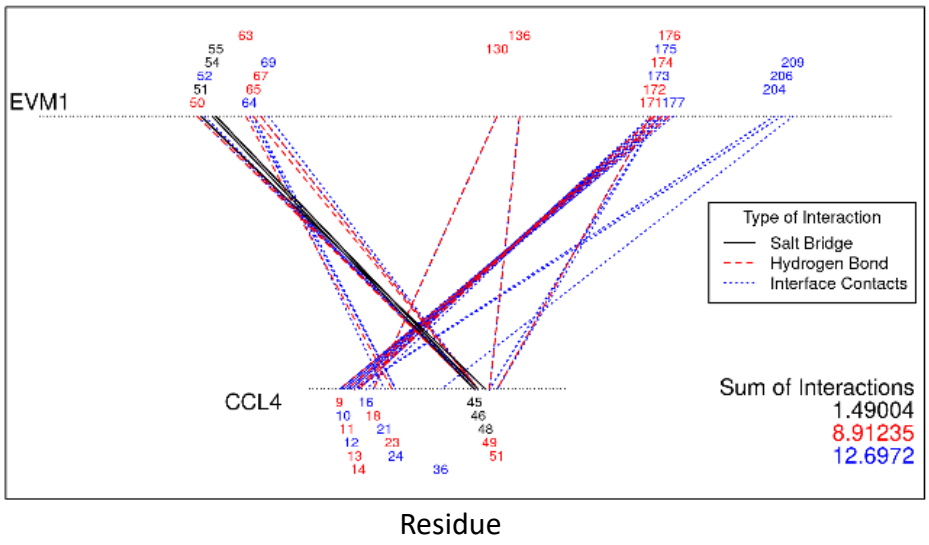
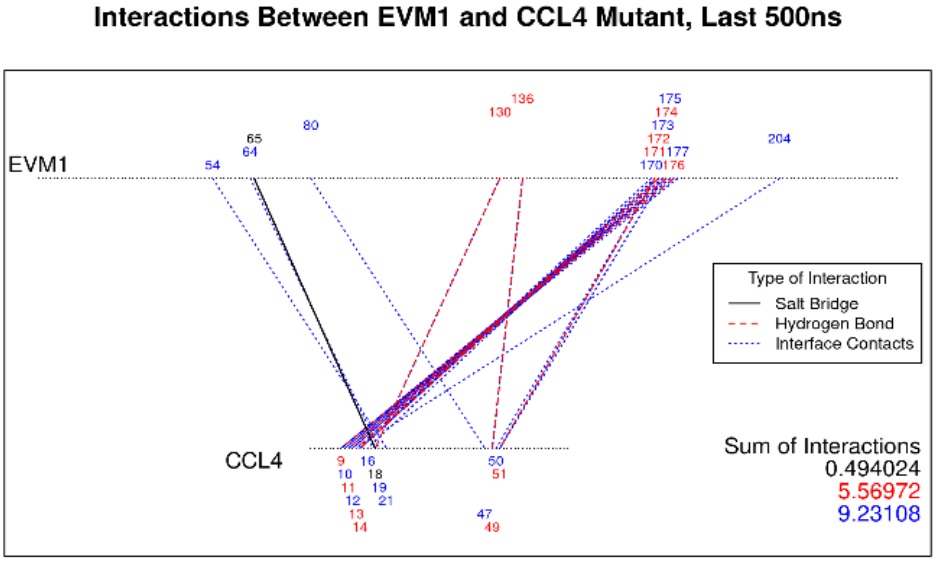


Figure 2.10: Persistent interactions of EVM1-chemokine complex. Mapping of persistent interactions colored by type of interaction: salt bridge (black), hydrogen bond (red), and interface contact (blue). Persistent interactions are interactions that are present for at least 33% of the evaluated trajectory (last 500ns). The dotted horizontal lines along the top and bottom of the graph represent residues of EVM1 and the specified chemokine respectively. The colored numbers indicate the residues involved in persistent interactions, with the strongest interaction determining the color. Fewer persistent interactions are formed between EVM1 and the bound chemokines compared to vCCI, though the pattern of interactions is consistent (compare to Figure 2.6).



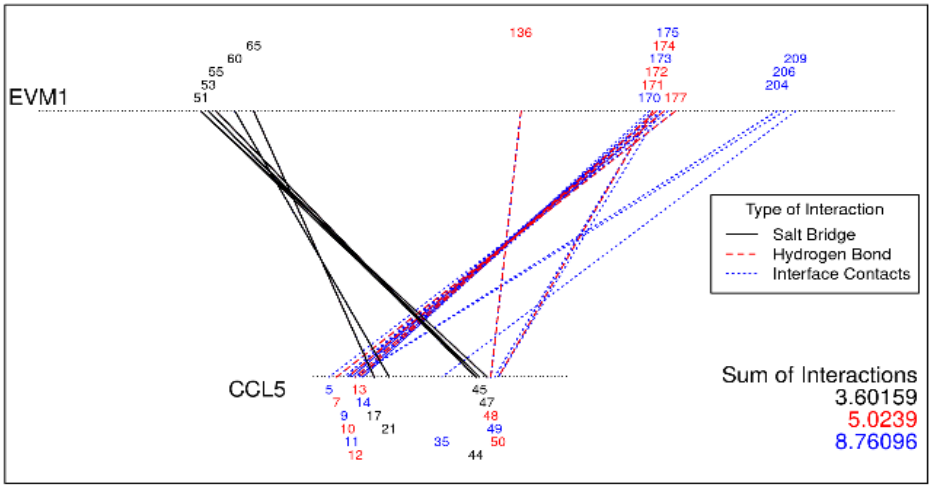
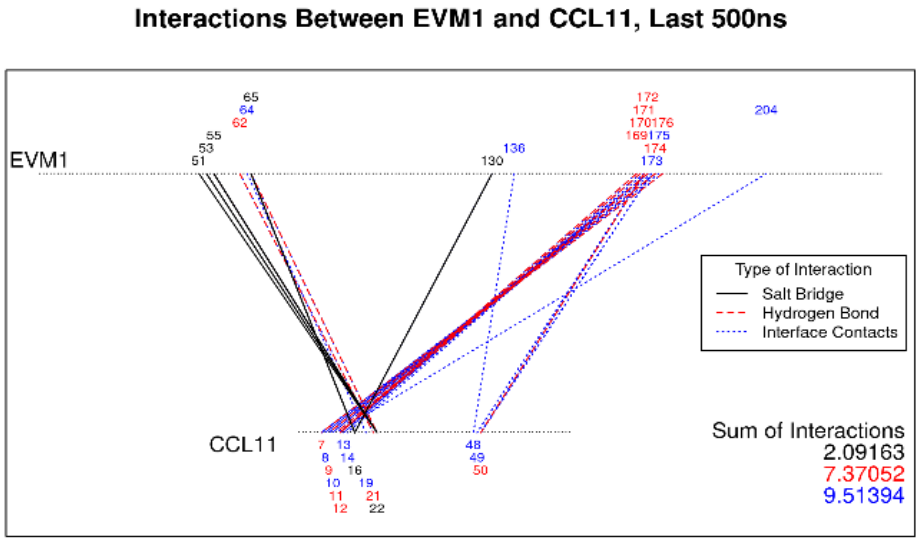


Figure 2.10 continued



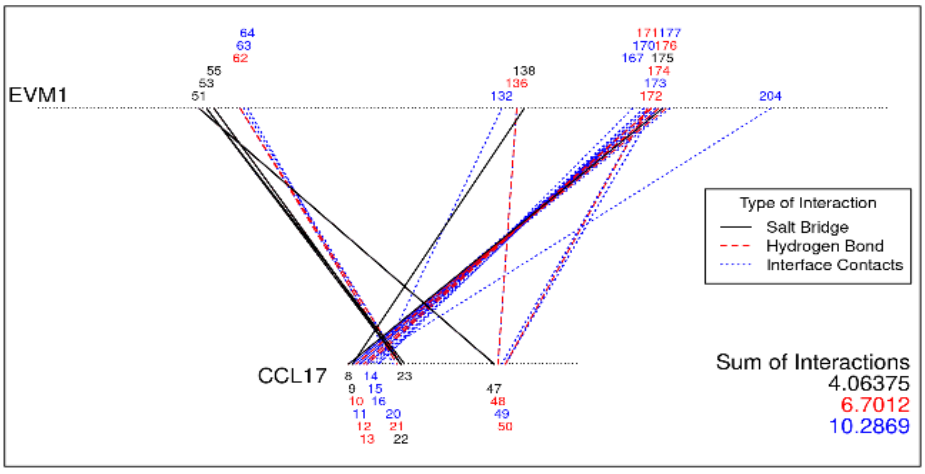


Figure 2.10 continued

2.7 Conclusions

The MD simulations of the vCCI-chemokine complexes repeatedly showed strong interactions supported by experimental findings, such as the interactions between R18 (CCL4 numbering) of the chemokine with D141 and E143 of vCCI or the hydrophobic interactions between the F13 of the chemokine and F215 and Y217 of vCCI. Additionally, a stable beta strand was identified between the chemokine N loop and the vCCI beta strand 8, which has not been previously reported for vCCI but has been seen with several other chemokine binding proteins.

While simulation can provide a good approximation of binding and conformations with current forcefields, there may be other factors that influence chemokine binding to vCCI, especially CCL17 binding, that are outside of the current scope of MD simulation. For example, the simulations run for a total of 1μs, a small snapshot of time for these complexes. This means the complex is unlikely to disassociate unless the proteins are in undesirable positions, such as clashing atoms. For this reason, we rely on the number and persistence of the interactions to suggest how well a complex holds together for the simulated time. Other factors, such as changes in free energy between being isolated in solvent versus in a complex, are not included in simulations of the bound complexes. Simulation techniques such as free energy perturbation have been used to calculate accurate binding free energies[44]. Such methods, previously limited to small molecules for drug design, are starting to be applied to larger systems as the availability and capabilities of computational power improve.

3 Simulating Mutation Effects on vCCI-chemokine Interactions

Experimental mutational studies have been the primary tool in determining the amino acids that mediate protein-protein interactions. Several studies have been conducted on vCCI and various chemokines to determine which residues are essential to the complex. To engineer vCCI to bind more specifically to desired targets as an auto-immune treatment, it is key to understand what effect mutations have on their interactions. The mutations tested below were based on previous experimental results. We studied the effects of these mutations to determine if simulation can match the observed effect or provide evidence as to why a mutation caused the observed change.

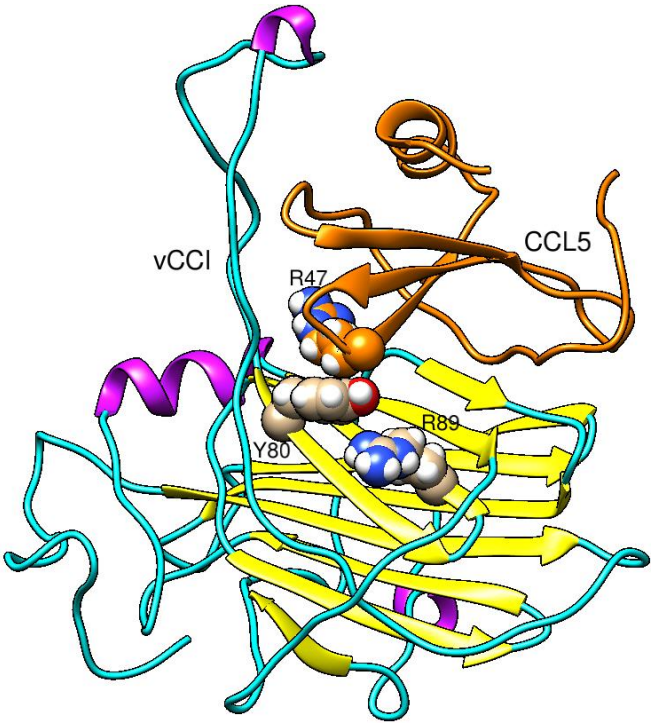


Figure 3.1: Residues of vCCI targeted for mutation. Ribbon structure of vCCI (cyan, with secondary structures colored) with CCL5 (orange) bound. Y80 and R89 of vCCI, two residues selected for mutation, are depicted as spheres to show steric and electrostatic clash with chemokine residue R47 (also represented as spheres), a conserved basic residue in CC chemokines.

3.1 vCCI Y80A

Experimental chemokine mutation studies revealed that mutating K49 (CCL2 numbering; K48 in CCL4) to an alanine improved binding to vCCI[9,12]. The structure of vCCI bound to a chemokine revealed that the residues on vCCI in proximity to K49 – Y80 and R89 – pack closely to this chemokine residue, likely leading to steric and/or electrostatic clash (see Figure 3.1)[45]. This led White et al[38] to select these residues for mutation in vCCI. In their study, using vaccinia virus 35K (another name for vCCI), they found R89A improved chemokine binding as expected, while Y80A resulted in a loss of function. They did not look into this mutation further, making it an ideal candidate for further study.

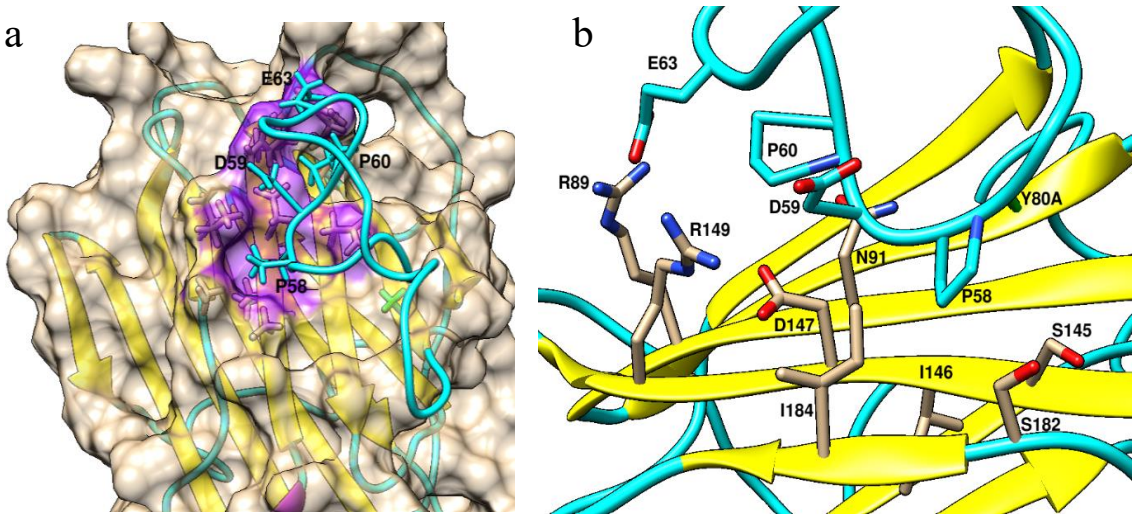


Figure 3.2: Loop collapse of Y80A. a) Top view of loop collapse, with loop (cyan) beta sheet 2 surface shown (tan), and contact between loop and beta sheet 2 (purple). P58 of the loop falls into a pocket on the beta sheet to stabilize the collapsed form. b) Side view of collapsed loop to show side chains of the loop residues (cyan) and beta sheet 2 residues (tan) interacting during loop collapse.

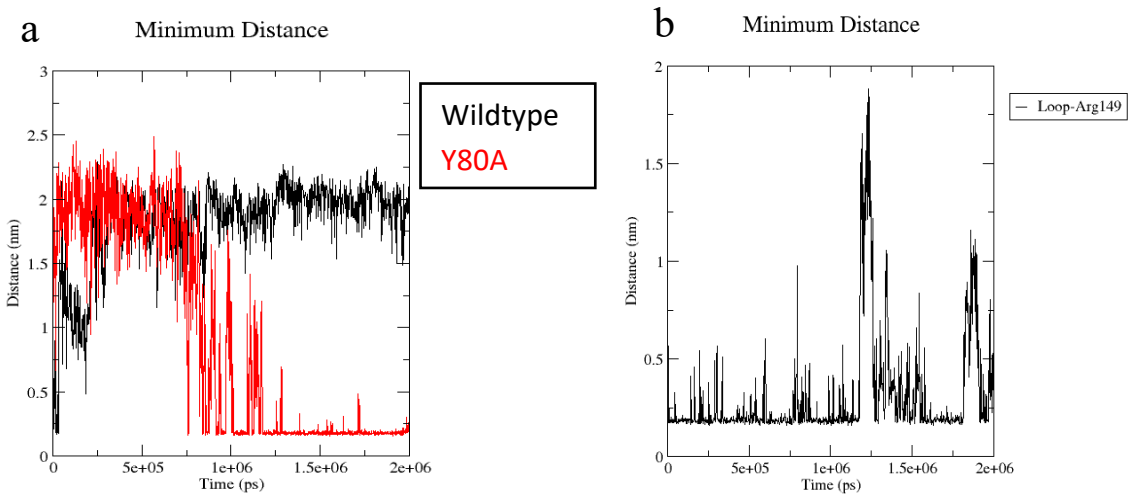


Figure 3.3: Tracking loop collapse through R149. a) Plot of the minimum distance between the residues of the loop and residue R149, which is on the opposite end of beta sheet 2, when simulated with no chemokines. For vCCI wildtype (black), the loop initially starts close to R149, but quickly moves away. In contrast, the vCCI Y80A loop (red) starts far away, initially collapses around 700ns, and fully collapses around 1.2μs. b) A second simulation of vCCI Y80A was run to see if loop collapse would be repeated. The loop collapsed very quickly towards R149, though did occasionally move away and back over the course of the simulation.

Y80 is a highly conserved residue in vCCI across poxviruses (see Figure 1.2), located on beta strand 3 near the base of the acidic loop (Figure 3.1). When a chemokine is bound, the aromatic ring sits parallel between beta sheet 2 of vCCI and the 40s loop of the chemokine. When vCCI is by itself, Y80 is seen upright, parallel to the loop. From this, we hypothesized that Y80 may act as a wedge keeping the loop up to keep the binding face clear. If the loop collapsed onto beta sheet 2, this may prevent the chemokine from binding, resulting in the loss of function as they observed.

Simulations of vCCI support this hypothesis, with loop collapse observed in two separate simulations, while simulations of vCCI wildtype never show loop collapse. Figure 3.2 shows the structure of the loop collapsed onto beta sheet 2. The loop, searching for basic

residues, interacts with R149 on the opposite end of beta sheet 2 (See Figure 3.3), resulting in the loop covering most of beta sheet 2 and burying many residues key to chemokine binding (See Figure 3.4). Two proline residues in the loop, P58 and P60, fit into pockets in the beta sheet 2 face. The peaks of buried surface area on the beta sheet face caused by loop collapse coincide with those observed with a chemokine bound.

Initial loop collapse, involving only interactions with R149, is relatively unstable due to the high flexibility of the acidic loop and few interactions to pin the loop in place, as observed in the second simulation (Figure 3.3b). However, the loop appears to become more stabilized when P58 falls into a hydrophobic pocket in beta sheet 2, seen in Figure 3.2. During the initial collapse, the loop is still very flexible, with the distance between the loop and R149 fluctuating (see 0.8-1.3μs in Figure 3.3a). Once P58 falls into the hydrophobic pocket, around 1.3μs, the fluctuations in the loop drop dramatically, stabilizing the loop's position in relation to R149.

In contrast, the vCCI wildtype loop started close to R149, but quickly moved away and remained far away for the majority of the simulation. In this simulation, the acidic loop instead moved in the opposite direction, interacting with the vCCI N-terminal tail, leaving the beta sheet 2 face exposed to solvent. The differences in the buried surface area caused by the acidic loop can be seen in Figure 3.4. For vCCI wildtype, there is a peak in BSA around the first ten residues due to this interaction. This leaves the key beta strands of beta sheet 2 (residues 143-149, 180-186) exposed, while for vCCI Y80A, these two beta strands are somewhat buried by the collapsed loop.

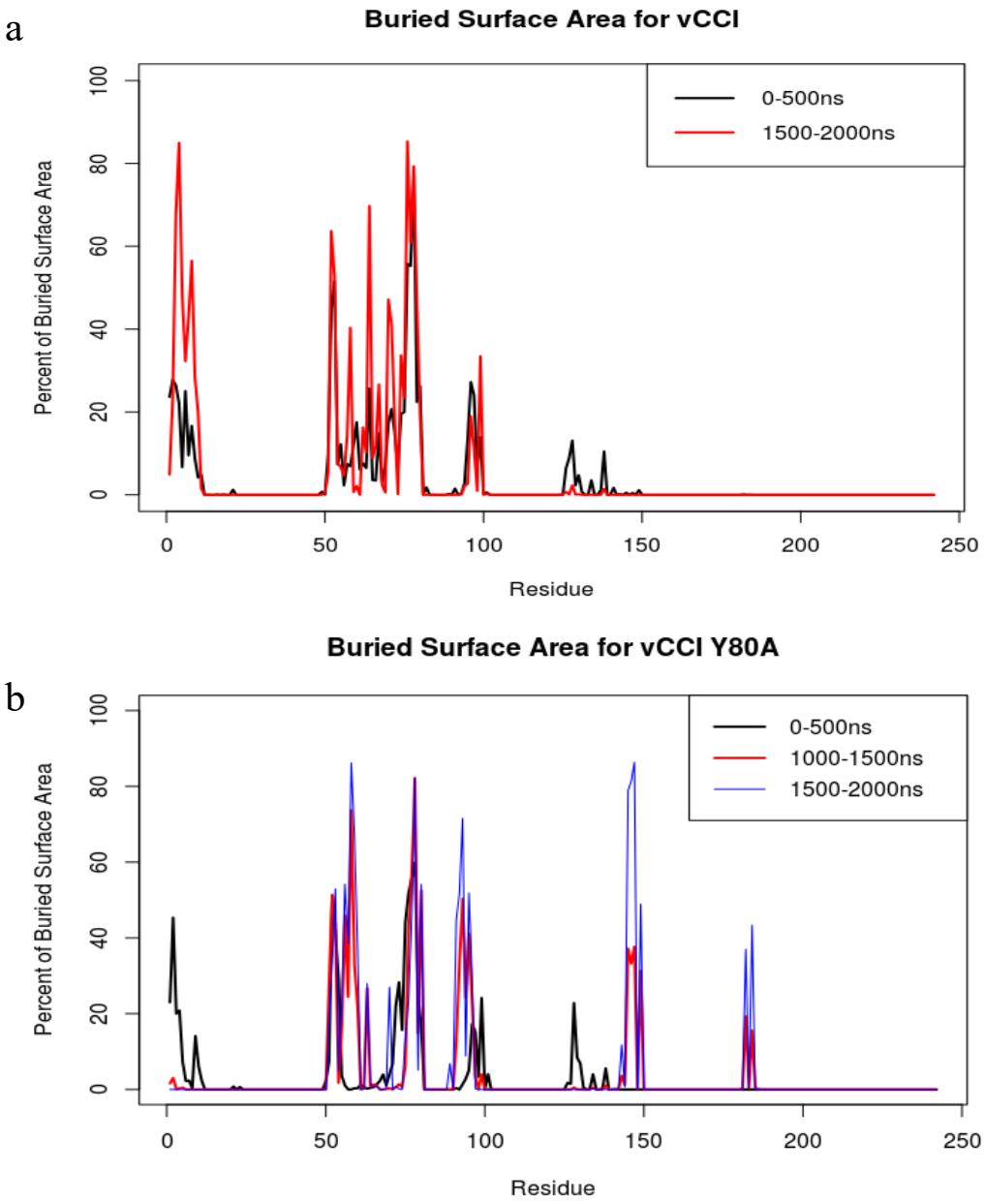


Figure 3.4: Comparison of BSA for vCCI wildtype and Y80A. Percent BSA for vCCI wildtype (a) and vCCI Y80A (b) is shown. The loop for both simulations is labeled as a separate chain to create the buried surface area. a) The time ranges, 0-500ns (black) and 1500-2000ns (red), were selected to show the differences between the start and end of the simulation. The acidic loop interacts with the N-terminal tail during most of the simulation, keeping the beta sheet 2 face clear. b) The time ranges for vCCI Y80A, 0-500ns (black), 1000-1500ns (red), and 1500-2000ns(blue), were selected to show the start of the simulation, the beginning of loop collapse, and full loop collapse respectively. Loop collapse results in peaks of BSA around regions key to chemokine binding.

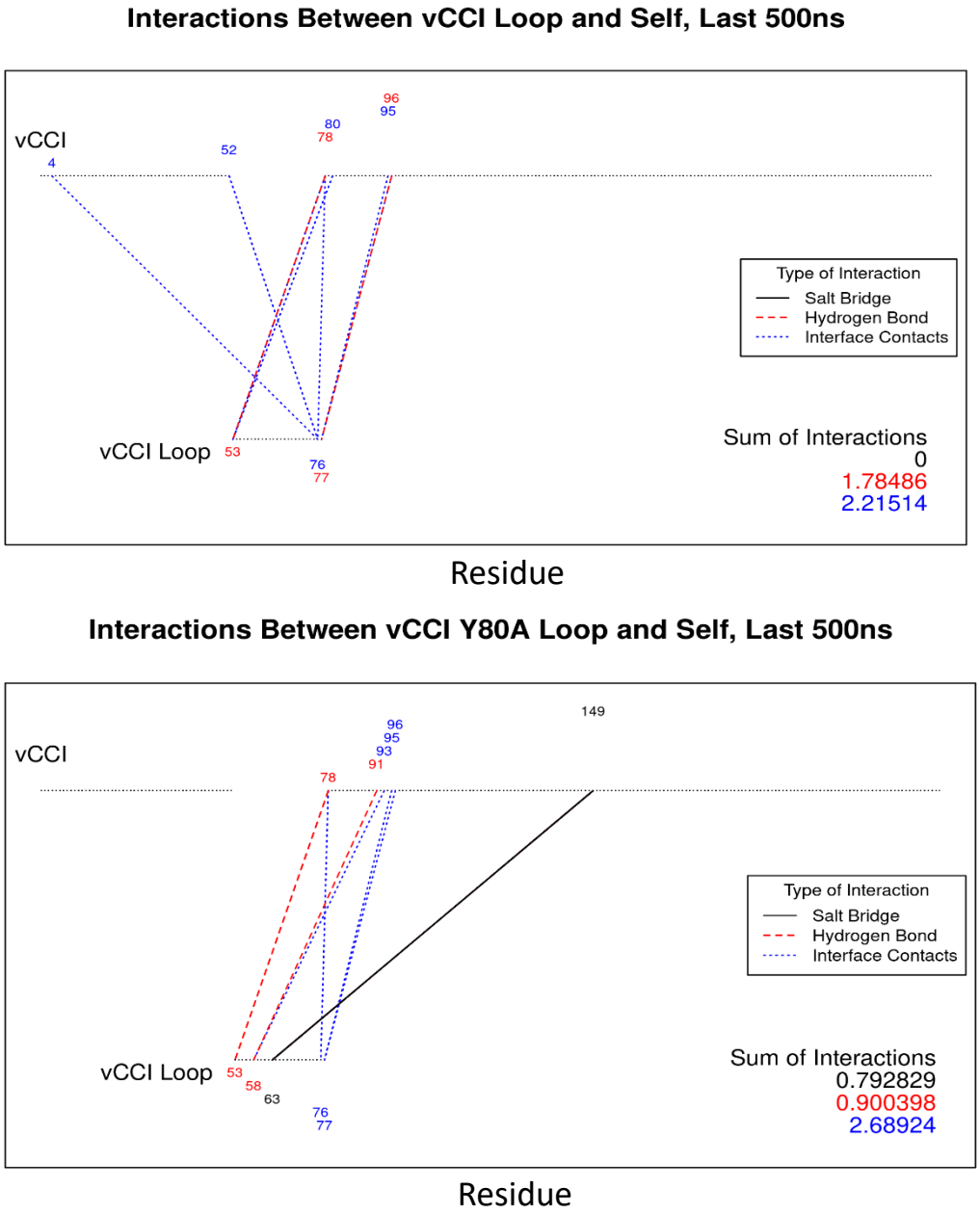


Figure 3.5: Persistent Interactions of vCCI loop in wildtype and Y80A. Mapping of persistent interactions between the vCCI loop and the rest of vCCI. The gap in points in the top horizontal line are for the acidic loop, which is represented by the bottom points. In the vCCI wildtype simulations, the vCCI loop tended to interact with the N-terminal tail, though these interactions are transient. The other interacting residues (52, 78, 96) are where the loop connects or in close proximity. For the vCCI Y80A simulation, interactions between 58 of the loop and 91/93 on beta sheet 2 represents the pocket the P58 collapses into. The loop collapse is also stabilized by the salt bridge between E63 of the loop and R149 of beta sheet 2.

Looking at the hydrophobic pocket on beta sheet 2 in wildtype and chemokine-bound structures, Y80 can be seen covering this region. This could be further support of Y80 preventing loop collapse by preventing stabilization of this collapsed state. Due to the limitations of simulation time, we are unable to precisely estimate how stable either state of loop collapse is. However, with the loop remaining in the down position for up to 1μs, it suggests it is a feasible conformation the loop could adopt. With this conformation, we would predict a slower observed on-rate in competition assays.

3.2 vCCI Y80A bound to vMIP-II

The goal of the study that looked into the Y80A mutation was to reduce steric clash between the chemokine and vCCI to improve binding. Simulations suggest that Y80 may serve to keep the long acidic loop open and away from the binding face, and that mutating to Y80A can result in loop collapse, which obscures the residues involved in chemokine binding, which could explain the results previously seen. To test the initial goal to evaluate the effects of steric clash, vMIP-II was simulated bound to vCCI Y80A to observe if the reduced crowding improved binding compared to wildtype.

The interactions seen in the vCCI Y80A – vMIP-II simulation are nearly identical to those with wildtype, showing that the mutation does not cause unfavorable interactions with the viral chemokine, nor is there an increase in binding due to the reduction in steric clash from Y80 (Fig 3.6a). The main difference between the two appears in the reduction of persistent salt bridges along the N loop of vMIP-II and beta strand 8 of vCCI Y80A (residues 180-186). The same residues between the two proteins still interact, suggesting

Table 3.1: vCCI-chemokine interactions between wildtype and mutants of vCCI

Complex	**vCCI – vMIP-II**		**vCCI – CCL5**	
vCCI	wildtype	Y80A	wildtype	R89A
Hydrogen Bonds				
Average	24.3730	24.1633	24.7738	17.6773
Unique bonds	208	217	180	192
Persistent bonds	23	21	25	11
Unique residues	28	27	24	20
Persistent residues	12	12	12	8
Salt Bridges				
Average	18.4643	17.1155	17.5198	13.8566
Unique bonds	73	89	79	98
Persistent bonds	29	18	21	22
Unique residues	10	10	10	8
Persistent residues	7	4	4	6

no major shift in how the viral chemokine binds. Looking at the summary of interactions (Table 3.1, first two columns), the two simulations show very similar levels of hydrogen bonding, but the mutant shows a drop in the average number and persistence of salt bridges.

The buried surface area (BSA) for both vCCI and vMIP-II show a lot of similarity between the wildtype and Y80A simulations, suggesting no major changes in the interface between the two proteins (Fig 3.6b). The greatest variation for vCCI Y80A is in the acidic loop region (residues 53-77). The acidic loop is highly flexible, making its buried surface area highly variable across simulations. Additionally, the arrangement of negatively charged residues allows the loop to find complementary charges more easily despite the variability of their spread among different chemokines. This means the residues in the loop interacting with the chemokine at a given time can change, resulting in different BSA, even across similar simulations. vMIP-II shows only minor changes as

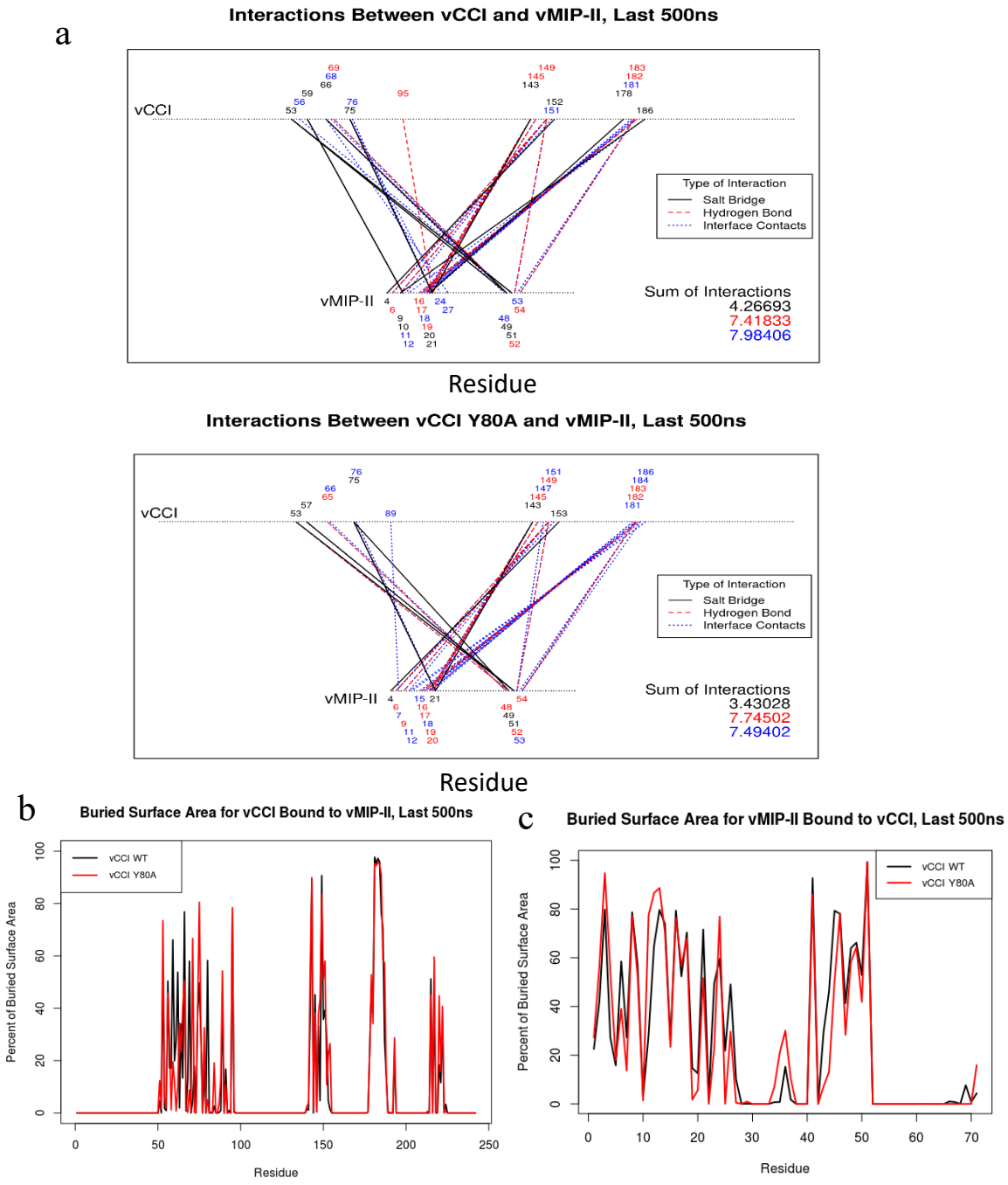


Figure 3.6: Comparison of vMIP-II binding to vCCI wildtype and Y80A. a) Mapping of interactions between vCCI and vMIP-II. The interactions are very similar between vMIP-II bound the vCCI wildtype and vCCI Y80A, with the most notable change being a reduction in salt bridges formed with vCCI Y80A. b) BSA for both vCCI and vMIP-II are nearly identical between the vCCI wildtype (black) and vCCI Y80A (red) simulations, showing the Y80A does not affect the interface with the chemokine.

well, mostly around the N loop and the 30s loop, which are both near beta strands 8 and 10 of vCCI. The mutation to Y80A may cause slightly more flexibility of the chemokine when bound due to the reduced strain, causing it to tilt towards those residues, where small increases in BSA for vCCI can be seen along beta strand 10.

Preliminary experimental results from our collaborators have shown the vCCI Y80A mutant does successfully fold and is capable of binding chemokines at higher concentrations than those used by the original experiment. Ongoing experiments are looking into how Y80A changes the binding constant. Based on the simulations performed, we predict the collapse of the loop to decrease the on-rate of the chemokine by covering key residues involved in the inhibitor – chemokine interface. Once bound, however, there should be little difference between how the wildtype and the mutant inhibitor interact with the chemokine, and so we expect a similar off-rate.

3.3 EVM1 Y69R / vCCI Y80R

Studies of EVM1 have also looked into mutations that affect chemokine binding[31]. The mutations that most reduced chemokine binding were I173R/Y, S171Y/W, andY69R. Y69R (analogous to Y80 in rabbitpox vCCI) reduced binding by 52-fold. The mutation to a large, basic residue near the acidic loop in a position shown to pack tightly near basic residues on chemokines strongly supports the reduced binding observed. Due to the similarity in the position of the mutation Y80A, it posed an ideal target to test through simulation. Two simulations were tested to observe how this mutation might affect chemokine binding and interactions with the loop: Y69R for EVM1 and Y80R for vCCI.

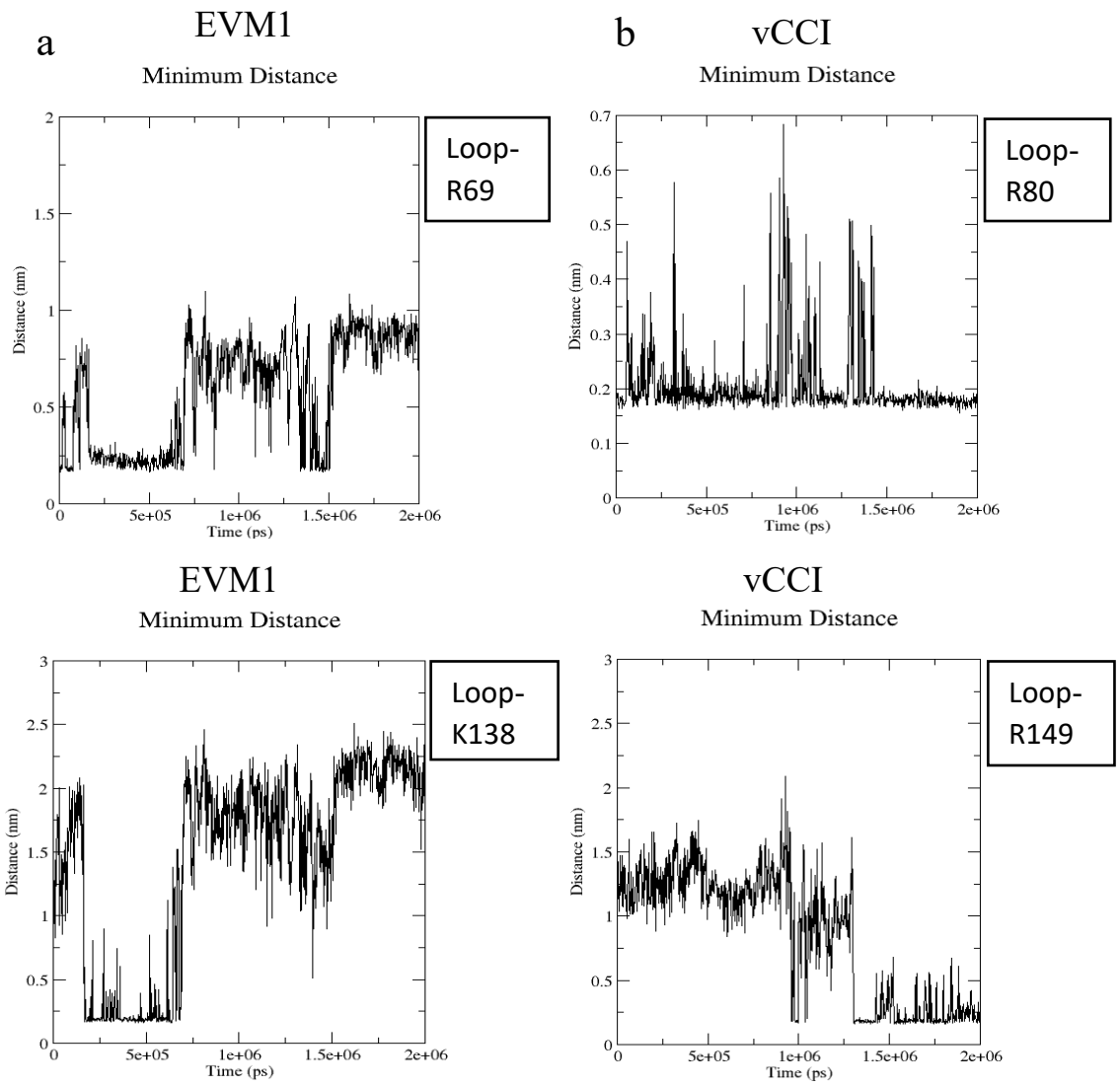


Figure 3.7: Tracking loop position in EVM1 Y69R / vCCI Y80R simulations. Minimum distance between the loop and selected residue for EVM1 (a) or vCCI (b). The top graphs track the distance between the loop and the mutated residue (R69 or R80) to show when they interact. The bottom graphs track the distance between the loop and a residue on the opposite end of beta sheet 2 (K138 or R149) to determine if the loop collapses. The vCCI loop consistently interacts with the mutated residue, while the EVM1 loop only maintains the contact for a few hundred nanoseconds.

Simulation of the EVM1 Y69R mutation shows the acidic loop interacting with the new basic residue. This conformation, however, is less stable than the interaction in vCCI Y80R. As seen in Figure 3.7a, the EVM1 loop maintains contact with Y69R for about 40% of the simulation. In contrast, the Y80R mutation in vCCI maintains contact with the loop for nearly the whole simulation (Figure 3.7b). While the loop of vCCI fluctuates, the number of nearby acidic residues ensures that a complementary charge can be found to maintain the interaction with Y80R. The shorter length of the loop in EVM1 likely makes it more strained to maintain the interaction with Y69R due to backbone constraints, which results in the loop breaking free in the last 400ns. This collapse of the acidic loop for vCCI Y80R somewhat differs from that seen in the vCCI Y80A mutation, with the majority of the loop interactions for vCCI Y80R pulling the loop to the side, not covering the beta sheet 2 binding face, as seen by the lack of loop interaction with R149 until the end of the simulation. It is only when the loop interacts with the R149 residue that the loop obscures the binding face. In contrast, for EVM1 Y69R, the loop interacts with R69 and K138 at the same time (see Figure 3.7a top and bottom), meaning the beta sheet 2 binding face is obscured while the loop interacts with the Y69R mutation. The contact of Y69R/Y80R with the acidic loop could explain part of the change in the binding affinity of a chemokine to the inhibitor.

Another factor of this mutation that would explain the 52-fold reduction in binding comes from the change to a basic residue in close proximity to a similar charge on the chemokine. As discussed above with vCCI Y80A, the tyrosine at this position packs tightly with a conserved basic residue on the chemokine. By mutating to Y69R/Y80R in

this region, the similar charges between the chemokine and the inhibitor should repel each other, making it very unfavorable for a chemokine to dock.

3.4 vCCI R89A

The other mutation that White et al. suggested to fix steric clash for the chemokine, R89A, was able to increase chemokine binding as they predicted. In their study, they tested the ability of vCCI to inhibit CC chemokines through indirect assays of chemokine ligand / receptor activity, and found that the vCCI R89A mutation resulted in a reduction in CCL5/CCR5 activity to 18% of that when no inhibitor is present, while wildtype vCCI reduced CCL5/CCR5 activity to 34%, suggesting tighter binding of the chemokine to vCCI with the R89A mutation compared to wildtype[38]. R89 of vCCI packs close to a conserved basic residue in most CC chemokines (See Figure 3.1). The proximity of these similar charges seen in the NMR structure led to the suggested mutation, which in turn showed that the removal of the basic residue improved chemokine binding.

The buried surface area of vCCI itself, when bound to CCL5, is mostly consistent between wildtype and the R89A mutant (Figure 3.9a), with a few peaks in the acidic loop and beta strand 4 having lower BSA for the mutant, while beta strand 10 and 11 have higher BSA for the mutant. The mutation, R89A, is present on beta strand 4, which would explain a small shift in BSA due to the smaller residue, however, most of the residues of beta strand 4 are buried less. This suggests the mutation caused the chemokine to shift slightly away from the beta strand, which is unexpected with the removal of what should be a repulsive residue. The R89A mutant has higher BSA along

beta strand 10 and 11 (residues 215-222), which helps make up part of the hydrophobic pocket utilized by the conserved hydrophobic residue in the N loop (F12 for CCL5).

Looking at the buried surface area of CCL5 (Figure 3.8c), there is a sharp drop in the BSA of the 20s region of the N loop, while there is an increase in BSA in the 30s loop. The 20s region of the N loop typically interacts with the middle half of the acidic loop (residues 63-67) of vCCI, which also shows lower BSA. The 30s loop of the chemokine typically lays near the end of beta strand 8 (residues 186-187) and beta strand 10 and 11 (residues 215-222). The increased BSA in this region is likely responsible for the increased BSA of beta strand 10 and 11. These results suggest that CCL5 is tilted slightly away from the beta sheet 2 face, moving the 20s region of the N loop away from the acidic loop and moving the 30s loop closer to the beta strand 10 and 11 residues. The absence of the R89 could cause the chemokine to lean more into the vacant region and tilt the chemokine relative to where it would bind normally.

The average hydrogen bonds and salt bridges formed between vCCI and CCL5 are reduced in the R89A simulation compared to wildtype vCCI (see Table 3.1). The repositioning of the chemokine reduced the expected interactions, which matches with the reduced buried surface area for the 20s region of the N loop. Interestingly, vCCI R89A with CCL5 forms more persistent salt bridges, as well as more unique salt bridges overall, than with wildtype vCCI. The interaction plots for vCCI wildtype and R89A mutant (Figure 3.8a) echo the stronger interactions along beta strand 10 and 11, as well as the reduced 20s region of the N loop interactions, as observed in the changes in BSA.

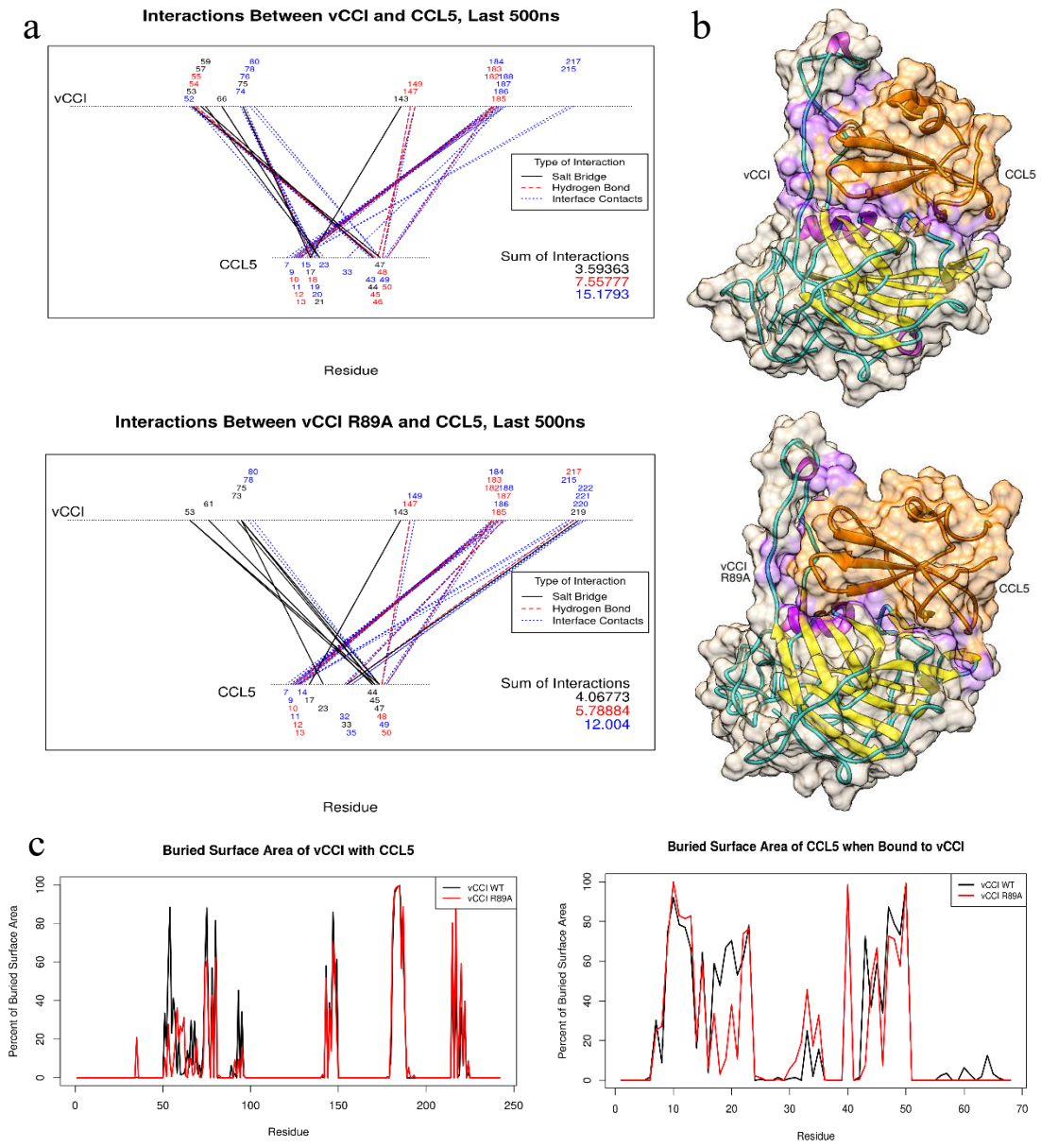


Figure 3.8: Comparison of CCL5 binding to vCCI wildtype and R89A mutant. a) Mapping of the persistent interactions shows that the vCCI R89A forms more salt bridges with CCL5, most notably between the 30s loop of CCL5 and beta strand 10 of vCCI. b) Structure of vCCI-CCL5 complex (top) and vCCI R89A-CCL5 complex (bottom), with vCCI colored tan and CCL5 in orange. The vCCI R89A mutant causes CCL5 to shift towards the site of the mutation, resulting in the N loop and C-terminal alpha helix of CCL5 to move away from the vCCI acidic loop, as well as the 30s loop to make more contact with beta strand 11. c) Percent BSA for vCCI and CCL5 with vCCI wildtype (black) and vCCI R89A (red). For vCCI, the acidic loop and beta strand 4 have lower BSA while beta strand 10 and 11 are higher for the R89A mutant. For CCL5, the N loop is less buried, while the 30s loop is more in the R89A simulation.

Although CCL5 tilts slightly away from the acidic loop of vCCI, as seen in the reduced interface contacts and hydrogen bonds that have a shorter range of interaction, it maintains all of the salt bridges seen when bound to vCCI wildtype.

This mutation is an ideal model for testing the capabilities of simulation to predict the effects a mutation may have on the chemokine-inhibitor complex. Based on the experimental results from White et al, we would expect to see more interactions to indicate the tighter binding and thus greater inhibition of the bound chemokine. However, from the results of the simulation, we instead see a reduction in interactions due to the repositioning of the chemokine caused by the mutation on vCCI. This repositioning does cause slightly more salt bridges to form, but fewer hydrogen bonds and less contact along the CCL5 20s region of the N loop with the vCCI acidic loop. This reduction in interactions may come from other forces, such as the reduction in steric clash and the clustering of similar charges. Extending the simulation may also allow for the structure to be more settled in the new orientation and improve the sampling of interaction conformations, much like what was observed in the vCCI Y80A loop collapse. Additional replicates will be needed to evaluate the likelihood of these observed shifts in the chemokine-inhibitor complex.

4 Future Prospects for the vCCI Project and other MD Simulations

The results shown thus far have revealed how MD can compare interactions of different complexes as well as provide insight into the effect mutations may have on the complex. Further simulations to continue building on this project are listed below. Additionally, other simulation techniques, such as free energy perturbation, could provide further insight into the difference between the chemokine-inhibitor complexes. Finally, these techniques can be applied to other protein-protein complexes, and are currently being used by our group to help design an inhibitor for ASC, a major component of the inflammasome.

4.1 Additional vCCI MD Simulations

4.1.1 Mutations to vCCI to improve binding of CCL17

From the results of the simulations, CCL17, which was previously shown to be a poor competitor to bind vCCI, forms a comparable number of interactions with vCCI as other tight binding chemokines like vMIP-II and CCL5. The main difference in the simulations arises in how CCL17 is positioned when bound to vCCI (see Figure 2.5). Due to the lack of charges in the 40s loop, CCL17 tilts forward for the 20s region of the N loop to interact where the 40s loop typically sits, causing the C-terminal alpha helix to be within reach of the vCCI acidic loop. While this new conformation allows CCL17 to form a comparable level of interactions as the other chemokines, this could cause strain to the complex and be less stable.

One way we could modify vCCI to address this issue could be mutating the acidic residues at the base of the vCCI loop (such as residues 53,55, and 75) to alanine. This should allow the 20s region of the N loop of CCL17 to position closer to the expected residues on vCCI (residues 62, 66, 68, and 69). This change should still allow the other chemokines to bind, but may weaken their binding overall.

Another option is to help improve binding in the first half of the N loop. Looking at the sequence alignment for the chemokines in Figure 1.1, CCL17 is missing a basic residue around residue 17/18. Instead, this residue is earlier, at residue 16 (K16). In the simulation, K16 can be seen close to D141 and E143 of vCCI (see Figure2.5c), which have been shown to be key in interacting with this basic residue on other chemokines[13]. However, there are no persistent interactions between these residues, and instead, K16 forms a salt bridge with D178 of vCCI during the simulation. Shifting the position of the acidic residue two residues later (residue 145) may help the positioning of CCL17.

While MD simulation does allow for a detailed look at how mutations might affect stability and binding affinity, there is the caveat that most simulations start with a pre-folded structure. MD simulation quality continues to improve, but it is important to work with experimental labs to test any results.

4.1.2 vCCI with CXC chemokine CXCL8

It has been shown previously that vCCI selectively binds only CC chemokines, with no detected evidence of binding from the other three subfamilies of chemokines[8]. Despite the similarities in tertiary structure, there is only around 20-40% sequence similarity between different subfamilies, as well as differences in receptor specificity and their

quaternary structure[41,46]. Simulations of CXC chemokines with vCCI can reveal what differences in interactions occur that may lead to the reduced affinity for other chemokines, and potentially find mutations that could improve binding to other chemokine targets.

The initial set up of the simulation, docking the CXC chemokine CXCL8 to vCCI, has proven more challenging than the others. The structure of CXCL8 (PDB ID: 5D14) has several side chains in unfavorable positions when aligning the backbone to the reference structure (PDB ID: 2FFK). Several of these side chains, such as K13 and R45, were repositioned to remove steric clash using the pyMOL mutagenesis tool, keeping the same residue but changing the atom starting positions. Some clash still remained, so the system underwent a series of energy minimization and position restraint steps to allow the system to relax enough to run. At this time, the simulation has run for 50 nanoseconds and should provide some interesting results to compare to CC chemokines. While it is too early to draw any conclusions from this simulation, the side chain clash observed when setting up the simulation may suggest that the positioning of the side chains for CXCL8 make it a less ideal target for vCCI to bind.

4.2 Free Energy Perturbation Simulations

Free energy perturbation (FEP) simulations have emerged as a tool for evaluating protein-ligand interactions and are constantly expanding to larger systems with the same approach[44,47]. This technique has been largely used in drug design, testing the relative binding free energy of small molecules to the target protein[48–50]. FEP has also been used in comparing how the effects of a mutation in a binding pocket can affect stability[51] and

the binding affinity of a ligand[52]. These techniques have proven to be robust for addressing different aspects of free energy[53], and are fairly accurate, with calculated free energy of binding within a few kcal/mol of established values[54].

These calculations typically involve "phasing out" of a subset of atoms, such as the ligand or mutated residue, to calculate the intermediate changes in free energy between the two states (e.g. bound vs. unbound for relative free binding energy). The free energy change between bound and unbound is too large to converge to a reliable value, so intermediate steps, representing "partially bound" states, are required. These intermediate steps are not physically possible, but through MD simulation, this technique is effective, where you can represent an atom as "30% present" by changing the forces it exerts.

Current FEP models do have some limitations that make modeling vCCI-chemokine complexes a challenge. First, the chemokine in the complex is much larger than a typical ligand used for drug design, making it much more computationally demanding. The large changes in energy would require many more transition states, and likely longer simulation times than current studies use, to provide reliable results. As computers continue to improve, this will become less of a concern.

A larger issue is chemokines have lots of basic residues. The "phasing out" of the chemokine to calculate the relative free energy of binding or solvation will result in a non-neutral charge in the system, which will cause errors in the simulation. This same issue arises with mutating acidic or basic residues using FEP. Some solutions could be to simultaneously phase in or out ions, or mutate a distal residue to counteract the change in charge, to set the system to neutral at each tested state[47]. This could change the ionic

concentrations of the solution, or affect the stability of the protein, which may affect how reliable the calculations are. As this field continues growing, a reliable solution for addressing these issues may not be too far off. Other techniques to determine protein binding affinities are constantly growing[55,56] to address these issues and may be better suited for the vCCI-chemokine complex.

4.3 ASC Simulations

Finally, the same analysis we have used for evaluating the vCCI-chemokine complexes is equally applicable to other protein complexes. In collaboration with Professor Eva De Alba's group who study the inflammasome complex, we have simulated several of their proposed inhibitors to narrow down which designs for her group to test in the lab.

Apoptosis-associated speck-like protein containing a CARD (ASC) is a major component of the inflammasome, which drives inflammation induced by the innate immune system[57]. Overactivity of the inflammasome has been linked to several auto-immune diseases and some cancers[57]. ASC is composed of two domains, a Pyrin domain (PYD) and a caspase activation and recruitment domain (CARD) connected by a linker, as shown in Figure 4.1b (in red). Recent work in inhibiting ASC speck formation has looked into small peptides that can bind and inhibit these domains[58–60].

We have simulated three peptides (referred to as 3b, 3e, and 3f) designed by our collaborators with the goal to bind the PYD domain tightly to prevent speck formation. A fourth design (referred to as ASCc) is based on the PYD domain of a splice variant which has been shown to disrupt inflammasome formation[59]. These peptides were docked onto

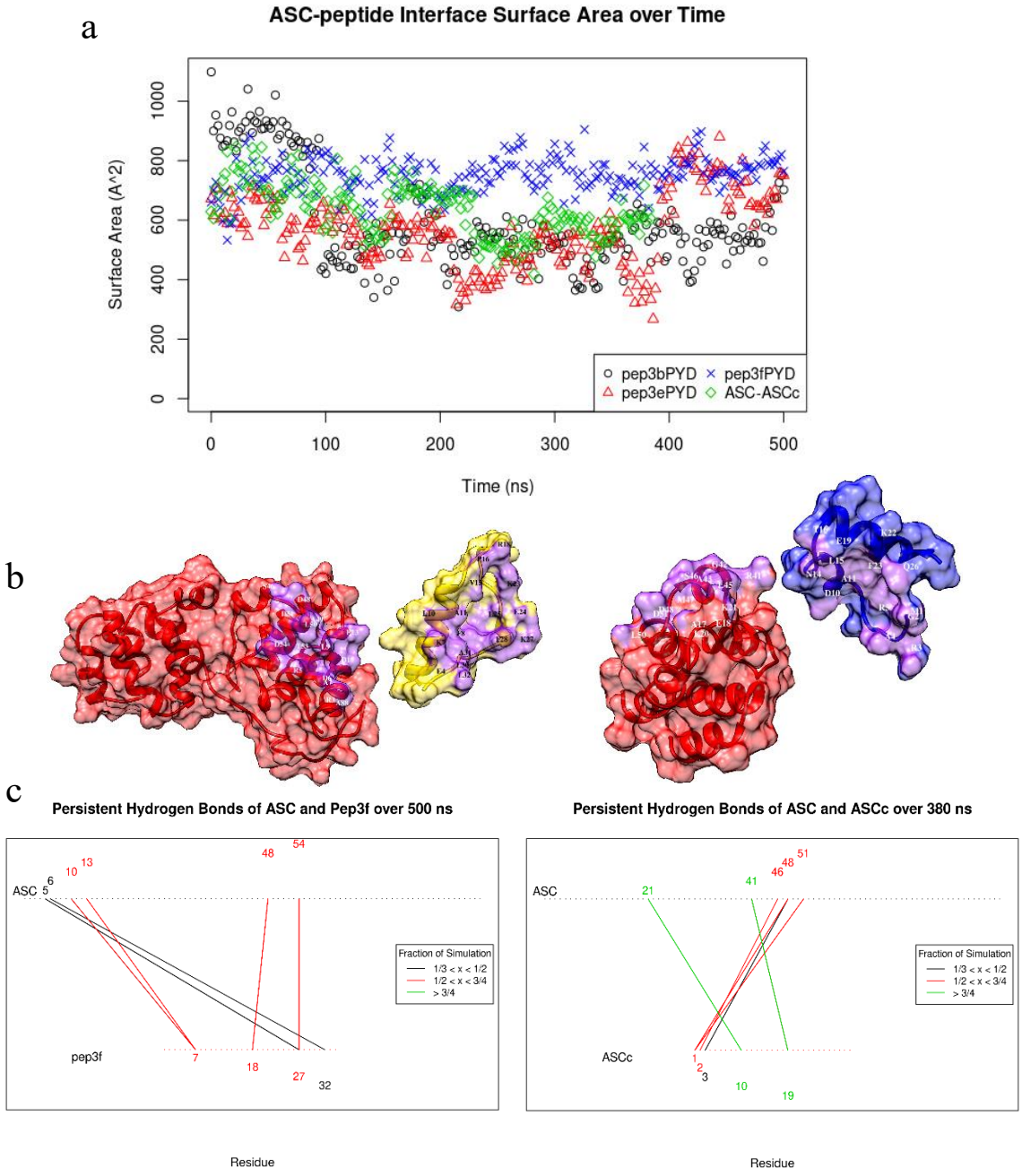


Figure 4.1: ASC – inhibitor simulations. a) Surface area of the interface between the ASC PYD domain and the indicated inhibitor over the course of the trajectory. b) Structures of ASC with inhibitor. The inhibitor is pulled away to show the interface between the two proteins (colored purple). Left: ASC (red) with peptide 3f (yellow). Right: ASC PYD domain (red) with ASCc (blue). c) Mapping of hydrogen bonds between ASC PYD domain and specified inhibitor. Lines are colored based on the fraction of the trajectory the interaction is present: black for 1/3 to 1/2 , red for 1/2 to 3/4, and green for > 3/4. Peptide 3f interacts mostly with helix 1 and 4, while ASCc interacts with helix 2 and 4.

the structure of ASC using HADDOCK by our collaborators based on NMR results and we have used these structures as the starting points for all-atom MD simulations. Each of the three peptides has been simulated for at least 500ns. We also ran simulations with the ASCc peptide docked on the ASC PYD domain, with the linker and CARD domain removed from the structure to reduce the computational load. This simulation has run out to 380ns.

Figure 4.1 shows some of the results of these simulations. Comparing the four different peptide inhibitors bound to the PYD domain of ASC, peptide 3f forms the largest interface and maintains these interactions the best. In contrast, peptide 3b starts with the largest interface space, but decreases to half the size at around 100 ns as the peptide shifts out of the pocket, suggesting a less favorable starting position. Interfaces for peptide 3f and ASCc are shown in Figure 4.1b. Peptide 3f interacts with alpha helix 1 and 4 of the PYD domain, while ASCc docks mostly on helix 3, with interactions also at helix 2 and 4. This is further supported by the persistent hydrogen bonds (Figure 4.1c). These two peptides represent two different interaction types seen for binding the PYD domain[61].

Based on these results, the collaborators will move forward with these two peptides for further study. In this way, MD simulation shows one of its most useful functions, serving as a tool to optimize time and resources in the lab by narrowing down better targets for further study.

References

1. Griffith, J. W., Sokol, C. L. & Luster, A. D. Chemokines and Chemokine Receptors: Positioning Cells for Host Defense and Immunity. *Annual Review of Immunology* **32**, 659–702 (2014).

2. Clark-Lewis, I. *et al.* Structure-activity relationships of chemokines. in *Journal of Leukocyte Biology* vol. 57 (1995).

3. Hemmerich, S. *et al.* Identification of Residues in the Monocyte Chemotactic Protein-1 That Contct the MCP-1 Receptor, CCR2 †. *Biochemistry* **38**, 13013–13025 (1999).

4. Stone, M., Hayward, J., Huang, C., E. Huma, Z. & Sanchez, J. Mechanisms of Regulation of the Chemokine-Receptor Network. *International Journal of Molecular Sciences* **18**, 342 (2017).

5. Bhusal, R. P., Foster, S. R. & Stone, M. J. Structural basis of chemokine and receptor interactions: Key regulators of leukocyte recruitment in inflammatory responses. *Protein Science* **29**, 420–432 (2020).

6. Liwang, A. C., Wang, Z. X., Sun, Y., Peiper, S. C. & Liwang, P. J. The solution structure of the anti-HIV chemokine vMIP-II. *Protein science : a publication of the Protein Society* **8**, 2270–80 (1999).

7. Fernandez, E. J., Wilken, J., Thompson, D. A., Peiper, S. C. & Lolis, E. Comparison of the structure of vMIP-II with eotaxin-1, RANTES, and MCP-3 suggests a unique mechanism for CCR3 activation. *Biochemistry* **39**, 12837–12844 (2000).

8. Burns, J. M., Dairaghi, D. J., Deitz, M., Tsang, M. & Schall, T. J. Comprehensive mapping of poxvirus vCCI chemokine-binding protein. Expanded range of ligand interactions and unusual dissociation kinetics. *Journal of Biological Chemistry* **277**, 2785–2789 (2002).

9. Seet, B. T. *et al.* Molecular determinants for CC-chemokine recognition by a poxvirus CC-chemokine inhibitor. *Proceedings of the National Academy of Sciences of the United States of America* **98**, (2001).

10. Nguyen, A. *et al.* Biophysical and Computational Studies of the vCCI:vMIP-II Complex. *International Journal of Molecular Sciences* **18**, 1778 (2017).

11. Kuo, N. W. *et al.* Structural insights into the interaction between a potent anti-inflammatory protein, viral CC chemokine inhibitor (vCCI), and the human CC chemokine, eotaxin-1. *Journal of Biological Chemistry* **289**, 6592–6603 (2014).

12. Beck, C. G. *et al.* The Viral CC Chemokine-binding Protein vCCI Inhibits Monocyte Chemoattractant Protein-1 Activity by Masking Its CCR2B-binding Site. *Journal of Biological Chemistry* **276**, 43270–43276 (2001).

13. Zhang, L. *et al.* Solution structure of the complex between poxvirus-encoded CC chemokine inhibitor vCCI and human MIP-1beta. *Proceedings of the National Academy of Sciences* **103**, 13985–13990 (2006).

14. Dabbagh, K. *et al.* Local blockade of allergic airway hyperreactivity and inflammation by the poxvirus-derived pan-CC-chemokine inhibitor vCCI. *Journal of immunology (Baltimore, Md. : 1950)* **165**, 3418–3422 (2000).

15. Yaron, J. R. *et al.* Deriving Immune Modulating Drugs from Viruses—A New Class of Biologics. *Journal of Clinical Medicine* **9**, 972 (2020).

16. Nelson, C. A., Epperson, M. L., Singh, S., Elliott, J. I. & Fremont, D. H. Structural conservation and functional diversity of the poxvirus immune evasion (PIE) domain superfamily. *Viruses* vol. 7 4878–4903 (2015).

17. Zhong, S., Macias, A. & MacKerell, A. Computational Identification of Inhibitors of Protein-Protein Interactions. *Current Topics in Medicinal Chemistry* **7**, 63–82 (2006).

18. Shaw, D. E. *et al.* Atomic-Level Characterization. **330**, 341–347 (2010).

19. Pan, A. C., Weinreich, T. M., Piana, S. & Shaw, D. E. Demonstrating an Order-of-Magnitude Sampling Enhancement in Molecular Dynamics Simulations of Complex Protein Systems. *Journal of Chemical Theory and Computation* **12**, (2016).

20. Dror, R. O., Dirks, R. M., Grossman, J. P., Xu, H. & Shaw, D. E. Biomolecular simulation: A computational microscope for molecular biology. *Annual Review of Biophysics* **41**, 429–452 (2012).

21. Wang, J., Alekseenko, A., Kozakov, D. & Miao, Y. Improved Modeling of Peptide-Protein Binding Through Global Docking and Accelerated Molecular Dynamics Simulations. *Frontiers in Molecular Biosciences* **6**, 1–10 (2019).

22. Elsässer, B. & Goettig, P. Mechanisms of proteolytic enzymes and their inhibition in Qm/mm studies. *International Journal of Molecular Sciences* vol. 22 (2021).

23. Proctor, N. K. *et al.* Towards a Better Understanding of Computational Models for Predicting DNA Methylation Effects at the Molecular Level. *Current Topics in Medicinal Chemistry* **20**, (2020).

24. Ando, D. *et al.* Nuclear pore complex protein sequences determine overall copolymer brush structure and function. *Biophysical Journal* **106**, (2014).

25. Pandolfi, R. J., Edwards, L., Johnston, D., Becich, P. & Hirst, L. S. Designing highly tunable semiflexible filament networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* **89**, (2014).

26. Bradley, R. & Radhakrishnan, R. Coarse-grained models for protein-cell membrane interactions. *Polymers* **5**, (2013).

27. Dror, R. O., Jensen, M. Ø., Borhani, D. W. & Shaw, D. E. Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations. *Journal of General Physiology* **135**, (2010).

28. Maragakis, P. *et al.* Microsecond molecular dynamics simulation shows effect of slow loop dynamics on backbone amide order parameters of proteins. *Journal of Physical Chemistry B* **112**, (2008).

29. Shaw, D. E. *et al.* Anton, a special-purpose machine for molecular dynamics simulation. *Communications of the ACM* **51**, (2008).

30. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242 (2000).

31. Arnold, P. L. & Fremont, D. H. Structural determinants of chemokine binding by an Ectromelia virus-encoded decoy receptor. *Journal of Virology* **80**, 7439–7449 (2006).

32. DeLano, W. L. Pymol: An open-source molecular graphics tool. *{CCP4} Newsletter On Protein Crystallography* **40**, (2002).

33. van der Spoel, D. *et al.* GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* vol. 26 1701–1718 (2005).

34. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. doi:10.1002/prot.22711.

35. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, (1983).

36. PDBePISA (Proteins, Interfaces, Structures and Assemblies). http://www.ebi.ac.uk/pdbe/pisa/ (2014).

37. van Rossum, G. & Drake, F. L. Python 3 Reference Manual; CreateSpace. *Scotts Valley, CA* (2009).

38. White, G. E. *et al.* Site-Directed Mutagenesis of the CC Chemokine Binding Protein 35K-Fc Reveals Residues Essential for Activity and Mutations That Increase the Potency of CC Chemokine Blockade Europe PMC Funders Group. *Mol Pharmacol* **80**, 328–336 (2011).

39. Couñago, R. M. *et al.* Structures of Orf Virus Chemokine Binding Protein in Complex with Host Chemokines Reveal Clues to Broad Binding Specificity. *Structure* **23**, (2015).

40. Dias, J. M. *et al.* Structural basis of chemokine sequestration by a tick chemokine binding protein: The crystal structure of the complex between Evasin-1 and CCL3. *PLoS ONE* **4**, (2009).

41. Liang, W. G. *et al.* Structural basis for oligomerization and glycosaminoglycan binding of CCL5 and CCL3. *Proceedings of the National Academy of Sciences of the United States of America* **113**, (2016).

42. Pace, C. N. *et al.* Contribution of hydrogen bonds to protein stability. *Protein Science* **23**, 652–661 (2014).

43. Jelesarov, I. & Karshikoff, A. Defining the Role of Salt Bridges in Protein Stability. in *Methods in molecular biology (Clifton, N.J.)* vol. 490 227–260 (2009).

44. Perez, A., Morrone, J. A., Simmerling, C. & Dill, K. A. Advances in free-energy-based simulations of protein folding and ligand binding. *Current Opinion in Structural Biology* **36**, 25–31 (2016).

45. DeRider, M. L., Zhang, L. & LiWang, P. J. Resonance Assignments and Secondary Structure of vCCI, a 26 kDa CC Chemokine Inhibitor from Rabbitpox Virus. *Journal of Biomolecular NMR* **36**, 22–22 (2006).

46. Clore, G. M. & Gronenborn, A. M. Three-dimensional structures of α and β chemokines. *The FASEB Journal* **9**, 57–62 (1995).

47. Chipot, C. & Pohorille, A. SPRINGER SERIES IN CHEMICAL PHYSICS 86 Free Energy Calculations Theory and Applications in Chemistry and Biology.

48. Williams-Noonan, B. J., Yuriev, E. & Chalmers, D. K. Free Energy Methods in Drug Design: Prospects of "alchemical Perturbation" in Medicinal Chemistry. *Journal of Medicinal Chemistry* **61**, 638–649 (2018).

49. Clark, A. J. *et al.* Free Energy Perturbation Calculation of Relative Binding Free Energy between Broadly Neutralizing Antibodies and the gp120 Glycoprotein of HIV-1. *Journal of Molecular Biology* **429**, 930–947 (2017).

50. He, X. *et al.* Fast, Accurate, and Reliable Protocols for Routine Calculations of Protein-Ligand Binding Affinities in Drug Design Projects Using AMBER GPU-TI with ff14SB/GAFF. *ACS Omega* **5**, 4611–4619 (2020).

51. Steinbrecher, T. *et al.* Predicting the Effect of Amino Acid Single-Point Mutations on Protein Stability—Large-Scale Validation of MD-Based Relative Free Energy Calculations. *Journal of Molecular Biology* **429**, 948–963 (2017).

52. Steinbrecher, T., Abel, R., Clark, A. & Friesner, R. Free Energy Perturbation Calculations of the Thermodynamics of Protein Side-Chain Mutations. *Journal of Molecular Biology* vol. 429 923–929 (2017).

53. Pérez-Benito, L., Casajuana-Martin, N., Jiménez-Rosés, M., van Vlijmen, H. & Tresadern, G. Predicting Activity Cliffs with Free-Energy Perturbation. *Journal of Chemical Theory and Computation* **15**, (2019).

54. Hansen, N. & van Gunsteren, W. F. Practical Aspects of Free-Energy Calculations: A Review. *Journal of Chemical Theory and Computation* **10**, 2632–2647 (2014).

55. Siebenmorgen, T. & Zacharias, M. Computational prediction of protein–protein binding affinities. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **10**, 1–18 (2020).

56. Plattner, N., Doerr, S., de Fabritiis, G. & Noé, F. Complete protein-protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. *Nature Chemistry* **9**, 1005–1011 (2017).

57. Martinon, F., Mayor, A. & Tschopp, J. The inflammasomes: Guardians of the body. *Annual Review of Immunology* vol. 27 (2009).

58. Pal, A. *et al.* Inhibition of NLRP3 inflammasome activation by cell-permeable stapled peptides. *Scientific Reports* **9**, (2019).

59. Bryan, N. B. *et al.* Differential splicing of the apoptosis-associated speck like protein containing a caspase recruitment domain (ASC) regulates inflammasomes. *Journal of Inflammation* **7**, (2010).

60. Dorfleutner, A., Chu, L. & Stehlik, C. Inhibiting the inflammasome: one domain at a time. *Immunological Reviews* **265**, 205–216 (2015).

61. Vajjhala, P. R., Mirams, R. E. & Hill, J. M. Multiple binding sites on the pyrin domain of ASC protein allow self-association and interaction with NLRP3 protein. *Journal of Biological Chemistry* **287**, (2012).