

UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

Studies of Protein Interactions and Knowledge-Based Drug Design: (A) The Electrostatic Nature of Recognition Between HIV-1 gp120 V3 Loop and Coreceptors CCR5/CXCR4, (B) Complement System Inhibition by Compstatin Family Peptides

Permalink

<https://escholarship.org/uc/item/7kb315gt>

Author

López de Victoria Suárez, Aliana

Publication Date

2012

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Studies of Protein Interactions and Knowledge-Based Drug Design: (A) The Electrostatic
Nature of Recognition Between HIV-1 gp120 V3 Loop and Coreceptors CCR5/CXCR4,
(B) Complement System Inhibition by Compstatin Family Peptides

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

Doctor of Philosophy

in

Bioengineering

by

Aliana López de Victoria Suárez

June 2012

Dissertation Committee

Dr. Dimitrios Morikis, Chairperson
Dr. David D. Lo
Dr. Jiayu Liao

Copyright by
Aliana López de Victoria Suárez
2012

The Dissertation of Aliana López de Victoria Suárez is approved:

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the guidance and support of several individuals who in many ways contributed to the preparation and completion of these studies. To all of you thanks!

My dissertation committee, especially my advisor, Dr. Dimitrios Morikis, for their patience, guidance and advice.

My colleagues, Chris A. Kieslich, Ron D. Gorham Jr. and Dr. Phanorios Tamamis for their support and the countless hours of discussion.

My funding sources, the U.S. Department of Education Graduate Assistance in the Areas of National Need (GAANN), the Graduate Research Mentorship Program (GRMP) Fellowship and the Dissertation Year Fellowship Award (DYFA) from Graduate Division at the University of California, Riverside (UCR).

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

1. López de Victoria A, Kieslich C, Rizo AK, Krambovitis E, Morikis D. Clustering of HIV-1 subtypes based on gp120 V3 loop electrostatic properties. *BMC Biophysics* 2012, **5**: 3.

2. López de Victoria A, Gorham RD Jr., Bellows-Peterson ML, Ling J, Lo DD, Floudas CA, Morikis D. A new generation of potent complement inhibitors of the compstatin family. *Chemical Biology and Drug Design* 2011, **77**: 431-440.

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

DEDICATION

I dedicate this dissertation to my family, especially,

to my daughter, Arodí, for been the driving force behind all this work;

to my parents, for teaching me the value of a good education;

to my partner in crime, Alberto, for believing in me.

ABSTRACT OF THE DISSERTATION

Studies of Protein Interactions and Knowledge-Based Drug Design: (A) The Electrostatic Nature of Recognition between HIV-1 gp120 V3 Loop and Coreceptors CCR5/CXCR4, (B) Complement System Inhibition by Compstatin Family Peptides

by

Aliana López de Victoria Suárez

Doctor of Philosophy, Graduate Program in Bioengineering
University of California, Riverside, June 2012
Dr. Dimitrios Morikis, Chairperson

Computational and experimental methods were used to understand (i) protein interactions involving the V3 loop of gp120 of HIV-1 with coreceptors in host cells and (ii) peptide analogs from the compstatin family to human C3. Computational methods, including molecular dynamics (MD) simulations and electrostatic calculations, provide quantitative predictions of dynamics and interactions at atomic resolution, while experimental methods, including surface plasmon resonance (SPR) and enzyme-linked immunosorbent assays (ELISA) are needed to confirm binding and inhibition.

HIV-1 entry into host cells is mediated by the interaction of the V3 loop of gp120 and coreceptors CCR5 or CXCR4 on host cell surfaces, with assistance of viral protein gp41 and cell receptor CD4. The mechanism of coreceptor selectivity is not well understood, given the sequence variability and structural flexibility of the V3 loop. Positive net charge is a persistent physicochemical property throughout HIV subtypes

and has been recognized as an influencing factor for cell entry. Electrostatic analyses of V3 loop structures with consensus sequences from HIV subtypes, show similar electrostatic potential characteristics, irrespective of sequence variability. Charge and other sequence-based criteria were combined to develop a scheme for determining coreceptor selection. In addition, MD simulations provide insight into loop dynamics, indicating that persistent salt bridges contribute in keeping the two loop strands in proximity, therefore providing a charged scaffold for electrostatic interactions with coreceptors, irrespective of structural variability.

Compstatin family peptides are inhibitors of the complement system and potential drug candidates against autoimmune and inflammatory diseases. Compstatin analogs are cyclic peptides that inhibit cleavage of human C3, therefore preventing further complement system activation. Introduction of tryptophan residues at the termini resulted in potent analogs, but suffering from reduced solubility. To balance hydrophobicity (important for binding) and polarity (important for solubility), additional analogs were designed guided by MD simulation results of bound analogs to C3. New analogs with polar substitutions at the N-terminus, including dipeptide sequence extensions and use of methylated tryptophan residues, were experimentally tested with ELISAs, demonstrating comparable inhibition to that of known analogs, but with improved solubility.

TABLE OF CONTENTS

Chapter 1: Protein Interactions	1
1.1 Background	1
1.2 Poisson–Boltzmann Electrostatics	7
1.3 Molecular Dynamics Simulations	8
1.4 References	12
Chapter 2: The Electrostatic Nature of Recognition between HIV-1 gp120 V3 Loop and Coreceptors CCR5/CXCR4	15
2.1 Background	15
2.2 Methods	18
2.2.1 Structural Templates	18
2.2.2 Docking Studies	19
2.2.3 HIV-1 Subtypes Studies Year 2009	21
2.2.4 Electrostatic Calculations	23
2.2.5 Clustering	25
2.2.6 Molecular Dynamics Simulations	26
2.2.6.1 Electrostatic Analyses of MD Trajectories	27
2.3 Results and Discussion	27
2.3.1 Docking Studies	27
2.3.2 V3 Loop Electrostatics and HIV-1 Subtypes Year 2009	28
2.3.3 V3 Loop Dynamics	36

2.4 Figures and Tables	42
2.5. References	78
Chapter 3: Complement System Inhibition by Compstatin Family Peptides	84
3.1 Background	84
3.2 Methods	91
3.2.1 Peptide Synthesis	91
3.2.2 Surface Plasmon Resonance	92
3.2.3 Normal Human Serum Kinetics	94
3.2.3 Human C3b Enzyme-linked Immunosorbent Assay	95
3.3 Results and Discussion	97
3.3.1 Surface Plasmon Resonance	97
3.3.2 Human C3b Enzyme-linked Immunosorbent Assay	99
3.3.2.1 Generation A	99
3.3.2.2 Generation B	101
3.4 Figures and Tables	106
3.5. References	123
Appendices	127
A.1 HIV-1 Subtypes Studies Year 1999	127
A.1.1 Consensus Sequences	127
A.1.2 Electrostatic Calculations	128
A.1.3 Sequence Clustering	130

A.1.4 Results and Discussion	130
A.2 V3 Loop Molecular Dynamics	133
A.2.1 Electrostatic Analyses	136
A.3 Surface Plasmon Resonance Protocol.....	136
A.4 Human C3b Enzyme-linked Immunosorbent Assay	141
A.4.1 Kinetics Assay	141
A.4.2 Compstatin Concentration Calculations	142
A.4.3 ELISA Protocol	143
A.4.4 GraphPad Prism Protocol	145
A.5 VMD/APBS Tutorial	146
A.6 Figures and Tables	150
A.7 References	174
Curriculum Vitae	176

LIST OF FIGURES

Chapter 2: The Electrostatic Nature of Recognition between HIV-1 gp120 V3 Loop and Coreceptors CCR5/CXCR4

Figure 2.1: Molecular models of the V3 loop.....	42
Figure 2.2:V3 loop residues influencing coreceptor selection	43
Figure 2.3: HIV-1 Classification.....	44
Figure 2.4: Conformational and orientational variability of CCR5-Nt	45
Figure 2.5: Docked structures with lowest total energy	46
Figure 2.6: Docked structures of 2B4C:CCR5-Nt base rotations.....	47
Figure 2.7: Docked structures of 2QAD:CCR5-Nt base rotations.....	48
Figure 2.8: Electrostatic clustering of mutants from structural template 2B4C	49
Figure 2.9: Electrostatic clustering of mutants from structural template 2QAD.....	50
Figure 2.10: Electrostatic clustering of HIV-1 subtypes, from consensus sequences and structural template 2QAD	51
Figure 2.11: Electrostatic clustering of HIV-1 subtypes, from consensus sequences and structural template 2QAD	53
Figure 2.12: Charge distribution clustering of HIV-1 subtypes, from consensus sequences and structural template 2QAD	55

Figure 2.13: Electrostatic clustering of HIV-1 subtypes, from consensus sequences and structural template 2B4C	56
Figure 2.14: Electrostatic clustering of HIV-1 subtypes, from consensus sequences and structural template 2B4C	58
Figure 2.15: Charge distribution clustering of HIV-1 subtypes, from consensus sequences and structural template 2B4C	60
Figure 2.16: Sequence clustering of HIV-1 subtypes, from consensus sequences, based on sequence similarity	61
Figure 2.17: Flow chart for HIV-1 coreceptor prediction	62
Figure 2.18: Root mean square deviation and Root mean square fluctuations	63
Figure 2.19: Secondary structure analysis for structural template 2B4C	64
Figure 2.20: Secondary structure analysis for structural template 2QAD	65
Figure 2.21: Contact plot for structural template 2B4C	66
Figure 2.22: Contact plot for structural template 2QAD	67
Figure 2.23: Electrostatic clustering of snapshots of the MD trajectories	68
Figure 2.24: Electrostatic clustering of snapshots of the MD trajectories	69
 Chapter 3: Complement System Inhibition by Compstatin Family Peptides	
Figure 3.1 Diagram of the complement system	106
Figure 3.2 Parent compstatin and W4/A9 analog	107

Figure 3.3 Diagram of the SPR procedure	108
Figure 3.4 Example of SPR data quality	109
Figure 3.5 Diagram of the ELISA procedure	110
Figure 3.6 Sensorgrams of Peptides IIA-VA from Generation A	111
Figure 3.7 Sensorgrams of Peptides VIA-IXA from Generation A	112
Figure 3.8 Human C3b ELISA data from Generation A	113
Figure 3.9 Human C3b ELISA data from Generation B	114
Figure 3.10 Human C3b ELISA data from Generation B	115

Appendices

Figure A.1 Electrostatic potential clustering of HIV-1 subtypes, from the 1999 consensus sequences and structural template 2B4C	150
Figure A.2 Electrostatic potential clustering of HIV-1 subtypes, from the 1999 consensus sequences and structural template 2QAD	152
Figure A.3 Electrostatic potential clustering of HIV-1 subtypes, from the 1999 consensus sequences and structural template 2B4C	154
Figure A.4 Electrostatic potential clustering of HIV-1 subtypes, from the 1999 consensus sequences and structural template 2QAD	156
Figure A.5 Charge distribution clustering of HIV-1 subtypes, from the 1999 consensus sequences and structural template 2B4C	158

Figure A.6 Charge distribution clustering of HIV-1 subtypes, from the 1999 consensus sequences and structural template 2QAD .	159
Figure A.7 Sequence clustering of HIV-1 subtypes from the 1999 consensus sequences .	160
Figure A.8 Electrostatic clustering of snapshots of the MD trajectories .	161
Figure A.9 Electrostatic clustering of snapshots of the MD trajectories .	163
Figure A.10 Cuvette for SPR and ELISA .	165
Figure A.11 Kinetic assay example of times and distribution. .	166
Figure A.12 Dilution plate and ELISA plate .	167

LIST OF TABLES

Chapter 2: The Electrostatic Nature of Recognition between HIV-1 gp120 V3 Loop and Coreceptors CCR5/CXCR4

Table 2.1: Alignment of HIV-1 V3 loop consensus sequences	70
Table 2.2: Docking of V3 loop with CCR5-Nt	71
Table 2.3: Docking of 2B4C with CCR5-Nt rotations	72
Table 2.4: Docking of 2QAD with CCR5-Nt rotations	74
Table 2.5: Occupancy of prevalent hydrogen bonds during the MD trajectories	76
Table 2.6: Occupancy of salt bridges during the MD trajectories	77

Chapter 3: Complement System Inhibition by Compstatin Family Peptides

Table 3.1 Compstatin analogs from Generation A	116
Table 3.2 Compstatin analogs from Generation B	117
Table 3.3 Kinetic analysis of analogs from Generation A	118
Table 3.4 IC ₅₀ results of analogs from Generation A	119
Table 3.5 IC ₅₀ results of analogs from Generation B	120
Table 3.6 Lipophilicity values of analogs from Generation B	121
Table 3.7 Lipophilicity algorithms comparison	122

Appendices

Table A.1 Alignment of HIV-1 V3 loop consensus sequences from 1999	168
Table A.2 SPR materials	170
Table A.3 Human C3b assays materials	171
Table A.4 Solutions formulas and calculations for Human C3b assays.....	172

CHAPTER 1

Protein Interactions

1.1 Background

Most proteins do not act alone; they have binding partners (ligands) to perform specific biological functions within an organism. Those binding partners can be other proteins, ions, sugars, small molecules, nucleic acids, etc. Proteins are involved in almost every biological function: DNA/RNA replication, gene regulation, molecule synthesis and degradation, transport, regulation of biochemical pathways, gated channels in cell membranes, signal transduction, immune responses, etc. The importance of protein interactions in biological functions has led to extensive research and the development of biophysical and biochemical methods to understand and detect how they function and interact.

Protein-ligand binding involves geometric complementarity and favorable energetic and entropic effects. Energetic effects consist of favorable physicochemical interactions, such as hydrophobic, hydrogen bonding, and electrostatic. These interactions occur mostly on the protein surface, in cavities or in pockets that are formed because of the protein structure. The properties of the cavities or pockets have been tailored for the binding of specific ligands through evolution. For example, if residues within a cavity are mostly hydrophobic, they create a non-polar region that may enable binding to hydrophobic ligands such as lipids. If residues within a cavity are

polar, they are amenable to the formation of inter-molecular hydrogen bonds with the ligand. If the residues within a cavity are charged, they generate electrostatic fields that attract ligands of opposite charge and are amenable to the formation of salt bridges. These interactions are highly specific and only the right ligand will result in the correct function.

Several properties are often studied to understand binding interfaces, such as the surface area buried upon complex formation, hydrogen bonds and salt bridge formation, residue composition and conservation, the shape and size of the interface, and secondary structure. Different families of proteins can have similar properties within their interfaces. Although research all of these properties contributed to the understanding of interfaces, they do not provide sufficient information about the process of binding (ei, kinetics and dynamics) and therefore it has been a challenge to predict interfaces and protein associations.

A study by Jones *et al* (1) compared residue types and their accessible surface areas (protein surface area that is accessible to solvent). Their study showed that hydrophobic and uncharged residues are mostly found at the interfaces of hetero-complexes, while charged residues were mostly found in non-interface areas. Several studies have measured the size of the interface to observe possible correlations with the formation of different types of protein complexes (reviewed in (2)). Those studies have explored the buried surface area (BSA) of homodimer proteins, protein-protein

complexes and crystal packing interfaces (only observed in protein crystals) within the Protein Data Bank (PDB, (3)) resulting in an average of 1600 Å² BSA for complexes, 3900 Å² BSA for homodimers, and 570 Å² BSA for crystal packing interfaces (reviewed in (2)). These results provide researchers with a way to distinguish biologically relevant complexes versus crystal complexes.

Initially it was proposed that protein interactions occur because of shape complementary, where two proteins would interact because their geometric shapes would fit into one another. This hypothesis treated proteins as rigid bodies and was referred to as the lock and key model, where the correct key (ligand) fits in the lock (protein), while wrongly shaped ligands do not fit. Because this model does not take into account the flexible nature of proteins and ligands a modification, called the induced fit model, was proposed. The induced fit model suggests that conformational changes occur upon binding in response to the interactions that are gained/lost during binding in an effort to maintain an energetic balance. Because of proteins inherent flexibility, one protein may be able to bind more than one ligand with different affinities, resulting in different functions. A more recent model, called conformational selection, suggests that before binding the free-ligand exist as an ensemble of conformations in equilibrium, from which one conformation (not necessarily the lowest-energy one) is selected to form a protein-ligand complex with the lowest free energy. The protein binding site also has dynamic character, but to a much lesser extent than the smaller ligand.

Proteins contain residues with ionizable groups that are dependent on their properties and environment. When those residues are found within interfaces, they can generate electrostatic fields that attract ligands of opposite charge. Therefore, electrostatic interactions are important for protein-protein interactions and play a role in many biological functions (4). Taking the influence on electrostatics into consideration for protein interactions, a two-step association model was proposed consisting of recognition and binding (5). The recognition step is driven by long range electrostatic interactions that expedite the formation of a non-specific encounter complex, where both molecules undergo reorientation of their binding sites. The binding step involves short/medium range electrostatic interactions that stabilize the final complex. Previous studies have proposed electrostatic mechanisms for the function of complement system regulator Factor H (6, 7), complement system receptor CR2 (8, 9) and complement system inhibiting proteins of viral or bacterial origin, such as vaccinia control protein (VCP) of the vaccinia virus (10), smallpox inhibitor of complement enzymes (SPICE) of the variola virus (10), Kaposi's sarcoma-associated herpesvirus inhibitor of complement activation (KAPOSICA) (11), and extracellular fibrinogen binding protein (Efb) and its homolog Ehp from the bacterium *S. aureus* (12). To delineate the contributions of each ionizable residue during association, electrostatic perturbations can be performed by creating computational alanine mutations, which remove a charged side chain altering the local charge and global electrostatic potential of the protein (13-17). This principle is

useful to depict intermolecular electrostatic interactions that are important for recognition and binding in protein-protein association.

The study of protein dynamics gives insight into how proteins adapt their structure to binding events. Among computational methods used to study dynamics is molecular dynamics (MD) simulations. MD simulations have been applied to biological systems for over 35 years, because they provide detailed atomic information on the structure, dynamics, and interactions of proteins, and enabling scientist to make correlations to function. Early MD studies by Mallik *et al* (18, 19) have demonstrated the dynamic character of compstatin, a fact that was proposed by the original structure determination study of compstatin by NMR (20) and subsequent NMR studies (21-24). Studies by Tamamis *et al* (25, 26) applied MD simulations to peptides of the compstatin family in complex with their target protein complement component 3 (C3) to determine structural differences and their importance in the function and inhibitory activities of the peptides. Another study employed MD simulations to delineate the species specificity of compstatin analogs using human or rat C3 (27). In addition, Tamamis *et al* have used MD simulations to design a hybrid mouse C3 with humanized binding site to alleviate the species specificity of compstatin family peptides (28).

Some interactions also require modifications of the protein in order to function properly. The most common modifications are phosphorylation (addition of phosphate groups), glycosylation (addition of glycans), lipidation (addition of lipid groups),

methylation (addition of methyl groups), and acetylation (addition of acetyl groups). These modifications also help regulate the function of proteins and their complexes. For example, several studies have shown that sulfation at Tyr present in the N-terminal of the G-protein coupled receptor (GPCR) CCR5 facilitated HIV entry into the cell (29, 30).

Anomalous protein interactions are associated with several diseases (reviewed in (31)). These diseases could be the result of protein mutations that cause conformational changes necessary for binding and lead to a destabilized complex, ligand binding occurring in the wrong place, or extremely high affinity that becomes toxic to the cell. In situations like these, understanding the binding mechanism between proteins and ligands can lead to the design of drugs that disrupt or reconcile the interaction.

Methods used to determine and characterize protein interactions involve a combination of experimental assays and computational tools (reviewed in (32, 33)). Each individual method can provide a piece of the puzzle, whereas the combination of methods and their results can give a more complete picture of the interaction.

Protein interactions are an essential part of cell function. The research presented here focuses (i) on improving the understanding of how the V3 loop of gp120 of HIV-1 selects and binds to the appropriate coreceptor in the host cell, and (ii) on optimizing the binding of compstatin family peptides to human C3 by developing analogs with better balance between their hydrophobic and polar characteristics.

1.2 Poisson–Boltzmann Electrostatics

Since electrostatic interactions are important for protein-protein interactions and play a role in many biological functions (4), electrostatic analyses are an essential part of the studies presented here.

The fundamental expression that describes electrostatic potential is Coulomb's law. Coulomb's law defines electrostatic potential as a function of charge q , distance r , vacuum permittivity ϵ_0 , and a relative dielectric permittivity of the medium with respect to vacuum dielectric constant ϵ_r , shown in the following expression:

$$V(r) = \frac{1}{4\pi\epsilon_0\epsilon_r} \frac{q}{r}$$

This expression is only appropriate for describing systems with a single dielectric medium, which is not the case when studying proteins in aqueous environment. Due to the complex nature of the protein environment, in which there exists a hydrophobic core with a dielectric constant closer to vacuum, surrounded by the high dielectric aqueous solvent, a more elaborate expression is needed. One such expression is the linearized-Poisson-Boltzmann equation (LPBE), which allows the calculation of the electrostatic potentials of systems of multiple dielectric constants, as well as accounts for the presence of mobile ions. The LPBE expresses electrostatic potential $\phi(r)$ as a function of charge $q(r)$, dielectric coefficient $\epsilon(r) = \epsilon_0\epsilon_r(r)$, and an ion accessibility function $\kappa(r)$, as described in the following equation (in SI units):

$$-\nabla \cdot \varepsilon(r) \nabla \varphi(r) + \varepsilon(r) \kappa^2(r) \varphi(r) = \sum_{i=1}^F q_i \delta(r - r_i)$$

The charge term $q(r) = \sum q_i \delta(r - r_i)$ accounts for fixed charges ez_i at positions $\delta(r - r_i)$, such as point and partial charges with values z_i , where e is the magnitude of electron charge. The amount and type of mobile counterions of valence $\pm Z_i$ is controlled by the ion accessibility function $\kappa(r)$, which is directly dependent on the ionic strength I . Expressions for these parameters are provided below,

$$\kappa^2(r) = \frac{2e^2 I}{\varepsilon(r) k_B T} \quad I = \frac{1}{2} \sum_{i=1}^M z_i^2 n_i^0$$

where k_B is the Boltzmann constant, T is the absolute temperature, and n_i^0 is the bulk number density of ion type i . The number density is the number of ions per unit volume (in units of m^{-3}), and relates to molar concentration c_i (in mol/L) as $n_i^0 = N_A c_i / 1000$.

Electrostatic potentials are related to electrostatic free energies by

$$G_{elec} = \frac{1}{2} \sum_{i=1}^N q_i \phi_i$$

The basic strategy for solving the LPBE is to place the protein in a three-dimensional grid, and then to assign a value for the charge, dielectric coefficient and ionic accessibility function at every grid point. Following the solution of LPBE a three-dimensional array describing the spatial distribution of the electrostatic potential is obtained.

1.3 Molecular Dynamics Simulations

Molecular dynamics is a useful tool to obtain information on structure, dynamics, and interactions of proteins. In MD simulations Newton's equations of motions are solved for each atom within the protein across a series of discrete time steps. Calculations of the force that acts on each atom allows for the determination of the acceleration of each atom. Integration of the equation results in a trajectory describing the positions, velocities and accelerations of each atom as they varied with time. MD simulations employ the use of a potential energy function (or force field) that models the basic interactions between atoms (forces). The potential energy function is dependent on the position of the atoms (that vary as a function of time) and contains bonded energy terms (describing atoms that are covalently bonded) and non-bonded energy terms (describing van der Waals and electrostatic forces). For these studies, an all-atom force field was used, where there are parameters for every atom type in the system.

The general potential energy function used in these studies is contained in CHARMM (34) implemented on NAMD (35), and is describe in the following equation.

$$\begin{aligned}
U(\vec{R}) = & \sum_{bonds} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{Urey-Bradley} K_{UB} (S - S_0)^2 \\
& + \sum_{dihedrals} K_\varphi (1 + \cos(n\varphi - \delta)) + \sum_{impropers} K_\omega (\omega - \omega_0)^2 \\
& + \sum_{non-bonded\ pairs} \left\{ \epsilon_{ij}^{min} \left[\left(\frac{R_{ij}^{min}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}^{min}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0\epsilon r_{ij}} \right\} \\
& + \sum_{residues} U_{CMAP(\varphi, \psi)}
\end{aligned}$$

The potential energy is the sum over each term representing the bonded and non-bonded contributions as a function of the atomic coordinates (34). The terms of the equation correspond to bonds (b), valence angles (θ), distance S between atoms A, C of a valence angle A-B-C (Urey-Bradley term, UB), dihedral angles (φ), improper angles (ω) and backbone corrections (CMAP, φ , ψ). The first five terms of the potential energy function account for bonded interactions, while the sixth term accounts for non-bonded interactions, and the last term accounts for backbone corrections. The parameters K_b , K_θ , K_{UB} , K_φ , and K_ω are force constants, while the variables with subscript 0 are the equilibrium values for the respective variable (34). The dihedral term is a sinusoidal expression, with n been the multiplicity of the dihedral angle and δ is the phase shift, while all the other terms are taken as harmonic. The protein backbone contains a numerical correction (CMAP) based on *ab initio* quantum mechanics calculations and structure-based potentials of mean force (34). This correction accounts for certain small

systematic errors in the description of the protein backbone by the all-atom CHARMM force field and improves the results obtained with MD simulations of proteins. The non-bonded terms include Coulombic interactions between point charges (q_i and q_j) and the Lennard–Jones (LJ) term, which accounts for attraction and repulsion of atoms. In cases of explicit solvent, where the protein is surrounded by solvent molecules (water and ions) with which the protein interacts, the relative dielectric constant (ϵ) is set to one (34).

1.4 References

1. Jones S, Thornton JM. Analysis of protein-protein interaction sites using surface patches. *J Mol Biol* 1997, **272**: 121-132.
2. Janin J. X-ray study of protein-protein complexes and analysis of interfaces. In: Zacharias M, editor. *Protein-Protein Complexes: Analysis, Modeling and Drug Design*. London, UK: Imperial College Press; 2010. p. 1-24.
3. Berman HM. The Protein Data Bank. *Nucleic Acids Res* 2000, **28**: 235-242.
4. Wu JZ, Morikis D. Molecular thermodynamics for charged biomacromolecules. *Fluid Phase Equilib* 2006, **241**: 317-333.
5. Mccammon JA, Northrup SH, Allison SA. Diffusional dynamics of ligand receptor association. *J Phys Chem* 1986, **90**: 3901-3905.
6. Kieslich CA, Vazquez H, Goodman GN, López de Victoria A, Morikis D. The effect of electrostatics on factor H function and related pathologies. *J Mol Graph Model* 2011, **29**: 1047-1055.
7. El-Assaad AM, Kieslich CA, Gorham RD, Jr., Morikis D. Electrostatic exploration of the C3d-FH4 interaction using a computational alanine scan. *Mol Immunol* 2011, **48**: 1844-1850.
8. Morikis D, Lambris JD. Physical methods for structure, dynamics and binding in immunological research. *TRENDS Immunol* 2004, **25**: 700-707.
9. Zhang L, Mallik B, Morikis D. Immunophysical exploration of C3d-CR2(CCP1-2) interaction using molecular dynamics and electrostatics. *J Mol Biol* 2007, **369**: 567-583.
10. Zhang L, Morikis D. Immunophysical properties and prediction of activities for vaccinia virus complement control protein and smallpox inhibitor of complement enzymes using molecular dynamics and electrostatics. *Biophys J* 2006, **90**: 3106-3119.
11. Pyaram K, Kieslich CA, Yadav VN, Morikis D, Sahu A. Influence of electrostatics on the complement regulatory functions of Kaposica, the complement inhibitor of Kaposi's sarcoma-associated herpesvirus. *J Immunol* 2010, **184**: 1956-1967.
12. Gorham RD, Jr., Kieslich CA, Morikis D. Complement inhibition by *Staphylococcus aureus*: electrostatics of C3d-EfbC and C3d-Ehp association. *Cell Molec Bioeng* 2011, **5**: 32-43.

13. Kieslich CA, Yang J, Gunopulos D, Morikis D. Automated computational protocol for alanine scans and clustering of electrostatic potentials: application to C3d–CR2 association. *Biotechnol Prog* 2010, **27**: 316-325.
14. Kieslich CA, Gorham RD, Jr., Morikis D. Is the rigid-body assumption reasonable? Insights into the effects of dynamics on the electrostatic analysis of barnase-barstar. *J Non-Cryst Solids* 2010, **357**: 707-716.
15. Gorham RD, Jr., Kieslich CA, Morikis D. Electrostatic clustering and free energy calculations provide a foundation for protein design and optimization. *Ann Biomed Eng* 2011, **39**: 1252-1263.
16. Gorham RD, Jr., Kieslich CA, Nichols A, Sausman NU, Foronda M, Morikis D. An evaluation of Poisson-Boltzmann electrostatic free energy calculations through comparison with experimental mutagenesis. *Biopolymers* 2011, **95**: 746-754.
17. Hakkoymaz H, Kieslich CA, Gorham RD, Jr., Gunopulos D, Morikis D. Electrostatic similarity determination using multi-resolution analysis. *Mol Inf* 2011, **30**: 733-746.
18. Mallik B, Lambris JD, Morikis D. Conformational inter-conversion of compstatin probed with molecular dynamics simulations. *Proteins* 2003, **53**: 130-141.
19. Mallik B, Morikis D. Development of a quasi-dynamic pharmacophore model for anti-complement peptide analogues. *J Am Chem Soc* 2005, **127**: 10967-10976.
20. Morikis D, Assa-Munt N, Sahu A, Lambris JD. Solution structure of compstatin, a potent complement inhibitor. *Protein Sci* 1998, **7**: 619-627.
21. Morikis D, Roy M, Sahu A, Troganis A, Jennings PA, Tsokos GC, Lambris JD. The structural basis of compstatin activity examined by structure-function-based design of peptide analogs and NMR. *J Biol Chem* 2002, **277**: 14942-14953.
22. Soulika AM, Morikis D, Sarrias MR, Roy M, Spruce LA, Sahu A, Lambris JD. Studies of structure-activity relations of complement inhibitor compstatin. *J Immunol* 2003, **171**: 1881-1890.
23. Morikis D, Soulika AM, Mallik B, Klepeis JL, Floudas CA, Lambris JD. Improvement of the anti-C3 activity of compstatin using rational and combinatorial approaches. *Biochem Soc T* 2004, **32**: 28-32.
24. Mallik B, Katragadda M, Spruce LA, Carafides C, Tsokos CG, Morikis D, Lambris JD. Design and NMR characterization of active analogues of compstatin containing non-natural amino acids. *J Med Chem* 2005, **48**: 274-286.

25. Tamamis P, Skourtis SS, Morikis D, Lambris JD, Archontis G. Conformational analysis of compstatin analogues with molecular dynamics simulations in explicit water. *J Mol Graph Model* 2007, **26**: 571-580.
26. Tamamis P, López de Victoria A, Gorham RD, Jr., Bellows-Peterson ML, Pierou P, Floudas CA, Morikis D, Archontis G. Molecular dynamics in drug design: new generations of compstatin analogs. *Chem Biol Drug Des* 2012, **79**: 703-718.
27. Tamamis P, Morikis D, Floudas CA, Archontis G. Species specificity of the complement inhibitor compstatin investigated by all-atom molecular dynamics simulations. *Proteins* 2010, **78**: 2655-2667.
28. Tamamis P, Pierou P, Mytidou C, Floudas CA, Morikis D, Archontis G. Design of a modified mouse protein with ligand binding properties of its human analog by molecular dynamics simulations: The case of C3 inhibition by compstatin. *Proteins* 2011, **79**: 3166-3179.
29. Farzan M, Mirzabekov T, Kolchinsky P, Wyatt R, Cayabyab M, Gerard NP, Gerard C, Sodroski J, Choe H. Tyrosine sulfation of the amino terminus of CCR5 facilitates HIV-1 entry. *Cell* 1999, **96**: 667-676.
30. Seibert C, Cadene M, Sanfiz A, Chait BT, Sakmar TP. Tyrosine sulfation of CCR5 N-terminal peptide by tyrosylprotein sulfotransferases 1 and 2 follows a discrete pattern and temporal sequence. *Proc Natl Acad Sci USA* 2002, **99**: 11031-11036.
31. Ryan DP, Matthews JM. Protein-protein interactions in human disease. *Curr Opin Struct Biol* 2005, **15**: 441-446.
32. Shoemaker BA, Panchenko AR. Deciphering protein-protein interactions. Part I. Experimental techniques and databases. *PLoS Comput Biol* 2007, **3**: e42.
33. Shoemaker BA, Panchenko AR. Deciphering protein-protein interactions. Part II. Computational methods to predict protein and domain interaction partners. *PLoS Comput Biol* 2007, **3**: e43.
34. Brooks BR, Brooks CL, III, Mackerell AD, Jr., Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, *et al.* CHARMM: the biomolecular simulation program. *J Comput Chem* 2009, **30**: 1545-1614.
35. Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K. NAMD2: Greater scalability for parallel molecular dynamics. *J Comp Phys* 1999, **151**: 283-312.

CHAPTER 2

The Electrostatic Nature of Recognition between HIV-1 gp120 V3 Loop and Coreceptors CCR5/CXCR4

2.1 Background

HIV-1 infects macrophages and T-lymphocyte cells. Its entry is mediated by the envelope glycoprotein gp120 associated with gp41. HIV-1 entry into the host cell involves binding of gp120 to CD4 receptor, inducing a conformational change in gp120 that exposes a region that can interact with a chemokine receptor: CCR5 or CXCR4, acting as coreceptors. Binding of gp120 to CD4 can also cause further conformational changes, exposing a region of gp41 that is involved in membrane fusion (1). CCR5 and CXCR4 belong to the G-protein couple receptor (GPCR) family, a large family of membrane proteins distinguished by seven transmembrane α -helices connected by three extracellular and three intracellular loops.

Gp120 is composed of 400-410 residues divided in five variable regions (V1-V5). The third variable region forms a loop, called the V3 loop, which is composed of 31-39 residues (Figure 2.1). The loop is closed by a disulfide bridge and is excessively positively charged. The V3 loop consists of three regions: the base (closer to the core of gp120), the tip (or crown) at the opposite end (further away from the gp120 core), and the stem between the base and the tip. The V3 loop is responsible for determining HIV tropism by

selecting the appropriate coreceptor CCR5 (infecting mostly macrophages) or CXCR4 (infecting mostly T-cells) (2-4). Viruses using CCR5 as coreceptor are called R5 viruses, while viruses using CXCR4 are called X4 viruses, and viruses that are capable of using both coreceptors are called R5X4.

The mechanism of how the virus selects one coreceptor versus the other is not well understood. However, several criteria within the V3 loop are known to influence coreceptor selection: (i) the absence of the glycosylation motif $N^6X^7T^8|S^8X^9$ (where $X \neq \text{Pro}$) shows a V3 loop preference towards X4 (4); (ii) the presence of one or more positively charged residues at positions 11, 24 and 25, known as the “11/24/25” rule, suggests a selectivity towards X4 (5); and (iii) the overall charge, as an increase of the net charge is associated with selectivity for X4 (6). Figure 2.2 shows the V3 loop of the available gp120 crystal structures including the side chains of the residues that are known to influence coreceptor selection.

The diversity of HIV-1 presents a problem in the development of effective treatments. Currently, HIV-1 is divided into three distinct genetic groups: M (major), N (non-major, non-outlier), and O (outlier); with group M being responsible for the majority of the infected population. This group is further divided based on the sequence variability of its *env* and *gag* genes (7) into ten subtypes or clades, named A through K, and circulation recombinant forms (CRFs). Figure 2.3 shows a flow chart of HIV-1 classification. Differences in coreceptor usage, geographical distribution and global prevalence have been demonstrated for several of the identified subtypes (8, 9).

Earlier X-ray studies of gp120 lacked sufficient electron densities at the V3 loop (10). Currently, there are two X-ray structures of gp120 complexes with the V3 loop intact, in which the V3 loop shows structural variability: PDB Code 2B4C (11) and 2QAD ((12), Figure 2.1), indicating that it is flexible and mobile. Several studies have suggested that the V3 loop interacts with the N-terminal extra-cellular domain of CCR5 (CCR5-Nt) and the extracellular loop 2 (ECL2, (11, 13, 14)) in a similar manner as with CXCR4 (15, 16), and that this interaction is mainly driven by electrostatics (15, 17-20). It has been known that electrostatic interactions are essential for many biological functions (21).

However, the understanding of the mechanism of interaction between gp120 and CCR5 (or CXCR4) is still incomplete. Post-translational modifications through the addition of sulfate groups in several tyrosines of the N-terminal of CCR5 are essential for the interaction with gp120 (22). Tyrosine sulfation also occurs in the N-terminal of CXCR4, however no significant difference in virus entry was found when compared to non-sulfated CXCR4 (23). The research presented here contributes to the understanding of the role of structure, dynamics, and underlying physicochemical properties in the interaction between gp120 and the coreceptors using computational methods.

2.2 Methods

2.2.1 Structural Templates

The currently available crystal structure of gp120 with intact V3 loop, with PDB codes 2B4C (11) and 2QAD (12), were obtained from the Protein Data Bank (PDB, (24)). In 2B4C, the gp120 with V3 isolate JR-FL was bound to CD4 and the antigen-binding fragment (Fab) of the X5 antibody. In 2QAD, gp120 with V3 isolate YU2, was complexed with CD4 and a 412d, a functionally sulfated antibody. Antibodies 412d and X5 are neutralizing antibodies, which protect a cell from antigens, by neutralizing (or inhibiting) its biological effect. Both antibodies are in contact with regions within the V3 loop, with antibody 412d (in 2QAD) being closer to the base of the loop while antibody X5 is closer to the stem of the loop. The coordinates of the V3 loop were retained, from both structures. Both V3 loop structures start at position 296 and end at position 331. In the case of 2B4C four residues have double conformations, from which conformation A was retained. In both structures residue positions 310-311 are left blank, whereas residue position 322 is occupied by two residues; as a result the total length of the peptide is 35 residues. Atoms and residues were renumbered starting from position 1 and ending with position 35, using Swiss-PDB Viewer (SPDBV, (25)). The initial positions of all atoms were taken from the crystallographic structure of each V3 loop. The sequence of both V3 loop structures was saved in FASTA format for further use.

The available NMR structure of the N-terminal of CCR5 was obtained from the

PBD (Code 2RLL, (12)). The structure contains 9 residues, from position 7 to 15. Atoms and residues were renumbered starting from position 1 and ending with position 9 using SPDBV (25). Tyrosines at positions 10 and 14 contained sulfate groups, which were excluded in these studies.

2.2.2 Docking Studies

The molecular docking studies were performed using the Grid Docking approach of Affinity, part of the Docking Module of Insight II (Accelrys Inc., San Diego, CA). The Grid Docking approach consists of placing the receptor and ligand in a grid where the ligand is moved, followed by calculating energies and checking if the resulted complex is acceptable within the specified parameters. The docking takes places by employing the Monte Carlo method (26), where random conformational changes are performed in the complex, and the energy of the new conformation is calculated and accepted/rejected based on energy cut off values.

The V3 loop from the available crystal structures (11, 12) and the available NMR structure for CCR5-Nt (12) were used. The Builder Module of Insight-II was used to add hydrogen atoms and assign charges. A binding subset, defined as a selection of amino acids in the receptor (or the larger molecule) that interacts with the ligand (or the smaller molecule) was selected. For these studies, the V3 loop is defined as the receptor, with binding subsets defined based on the residue position: (i) residues 1-6 and 30-35, base subset; (ii) residues 5-13 and 21-30, stem subset; or (iii) residues 11-23,

tip subset. Placement of CCR5-Nt is dependent on the residue subset selected on each V3 loop, resulting in three ligand-receptor complexes for each structure. The default parameters were used for the grid setup, with the CVFF force field and the solvation grid option selected. The default docking parameters were used. An initial energy minimization using the conjugate gradient method was used to eliminate any bad contacts present.

Based on initial placement the Docking Module produced possible conformations of the complex and scored them using on a potential energy function that consist of Coulombic and van der Waals terms. The resulting files contained the complexes, their calculated energies and statistical information. Optimization was performed by 90° rotations of CCR5-Nt in each axis. The rotations were carried using the base, the stem, and the tip subsets of each V3 loop, as defined above. Two rotations were performed for each axis for each binding subset: the last docked structure for the first rotation on each axis and subset was used as the starting structure for the next run, but CCR5-Nt was rotated 90° in that particular axis (second rotation). For example, CCR5-Nt was placed at the base of the V3 loop and rotated 90 ° in the x-axis (first rotation), after which energy calculations and complexes were obtained; the last complex from the first rotation was used to rotate CCR5-Nt 90 ° in the x-axis (second rotation) and energy calculations and complexes were acquired.

2.2.3 HIV-1 Subtypes Studies Year 2009

HIV-1 sequences are deposited in the HIV Databases of the Los Alamos National Laboratory (<http://www.hiv.lanl.gov>). Using tools within the database we extracted consensus sequences for the V3 loop of HIV-1. The residue sequences between and including the first and last cysteines of the V3 loop were selected. The Sequence Search Interface Tool was first used to obtain nucleotide sequences for HIV-1 subtypes. Within this search tool, the parameters selected were: subtype (for example, subtype A), virus (HIV-1), and genomic region (V3). The search result file is the input file for the ElimDupes Tool, which compares all the sequences and eliminates any duplicates. A cutoff of 93% DNA sequence identity of the *env* gene was used. The unique sequences file was used as the input file for the HIValign Tool, which aligns the sequences based on curated alignments within the database using the Hidden Markov Model (HMM) method. Several options were selected for this tool: align the sequences by HMM, codon-align the sequences, and translate to amino acid. The Simple Consensus Maker Tool was then used to obtain a consensus sequence, with the resulting file from HIValign used as the input file. The default parameters were kept, resulting in an alignment sequence with the first sequence identified as the consensus.

This procedure was done for each subtype and groups N and O. The results of the consensus sequence alignment, created using ClustalW2 (27), are shown in Table 2.1. Subtype A includes sub-subtypes A1 and A2, subtype F includes sub-subtypes F1 and

F2, and subtype CPX includes the 11 cpx subtypes available in the database. The consensus for subtype D resulted in 33 residue sequence because of gaps at positions 24-25. To equalize the length of the D subtype with the 35-residue length of the rest of the subtypes, residue frequencies were calculated at positions 24-25 of the D subtype and the residues with the second highest frequency in the alignments (gaps being the highest frequency) were chosen. These residues were lysine at position 24 and asparagine at position 25. Subtype J and group O contained two residues with the exact same frequency at a particular location. In the case of subtype J, the residues were isoleucine and leucine; and for group O, the residue was glutamic acid and a gap (introduced by the alignment). Isoleucine was selected for subtype J and glutamic acid was selected for group O. Subtypes B and K have the same consensus sequences, and subtypes CPX and H also share identical consensus sequences. Subtypes AB, AE, AG, and CPX are circulating recombinant forms (CRFs).

The program Modeller (28) was used to create homology models of all subtypes, using the two crystal structures as templates, with the modifications described above. The default optimization and refinement protocol of Modeller was used to generate single models, optimized with conjugate gradients and molecular dynamics-based simulated annealing.

Clustering of subtype sequences was performed using the score matrix generated by ClustalW2 as the input distance file to create a clustering dendrogram using the linkage function of MatLab (The MathWorks Inc., Natick, MA).

2.2.4 Electrostatic Calculations

The use of similarity measures for clustering of electrostatic and other physicochemical properties is a topic of chemistry and drug design research (29-35). Clustering of electrostatic potential depicts electrostatic similarities of proteins, which can be correlated to biological properties and functions of proteins and protein complexes. The Analysis of Electrostatic Potentials Of Proteins (AESOP) computational framework (33-37) was used, which provides a platform for elucidating the role of electrostatics and, more specifically, the role of each ionizable amino acid in protein association. This is accomplished using theoretical alanine scan or other mutagenesis in which electrostatic properties are perturbed by systematically removing each ionizable amino acid, one at a time (33, 35-38). The effects of these perturbations are quantified using electrostatic similarity clustering and electrostatic free energies of association (34, 39) to give insight into the contributions of each ionizable amino acid in both recognition and binding (33, 37). Parent proteins are superimposed to achieve the same orientation for all proteins and mutants.

Poisson-Boltzmann electrostatic calculations and hierarchical clustering analysis

were performed as previously described (33, 34, 36, 37). The program PDB2PQR (40) was used to prepare the V3 loop coordinates for electrostatic calculations by including van der Waals radii and partial charges for all atoms according to the PARSE force field (41). Electrostatic potentials were calculated using the Adaptive Poisson Boltzmann Solver (APBS, (42)) and the linearized form of the Poisson-Boltzmann equation. A box with $129 \times 129 \times 129$ grid points was used. The box dimension for the V3 loop alanine scan were: $70 \text{ \AA} \times 70 \text{ \AA} \times 70 \text{ \AA}$ and $50 \text{ \AA} \times 50 \text{ \AA} \times 50 \text{ \AA}$ for 0 and 150 mM, respectively, for 2B4C; and $65 \text{ \AA} \times 70 \text{ \AA} \times 70 \text{ \AA}$ and $50 \text{ \AA} \times 50 \text{ \AA} \times 50 \text{ \AA}$ for 0 and 150 mM, respectively, for 2QAD. For the HIV-1 subtypes the box dimensions were $70 \text{ \AA} \times 70 \text{ \AA} \times 75 \text{ \AA}$ and $50 \text{ \AA} \times 50 \text{ \AA} \times 55 \text{ \AA}$ for 0 and 150 mM, respectively, for subtypes from the template 2B4C; and $60 \text{ \AA} \times 70 \text{ \AA} \times 75 \text{ \AA}$ and $50 \text{ \AA} \times 50 \text{ \AA} \times 50 \text{ \AA}$ for 0 and 150 mM, respectively, for subtypes from the template 2QAD. Different box sizes were used for 0 mM and 150 mM calculations to assure maximum resolution while including an optimal number of grid points with electrostatic potential values within and about $\pm 1 k_B T/e$. The molecular surface was calculated using a probe sphere with a radius of $1.4 \text{ \AA}$, representing the radius of a water molecule. The dielectric coefficients were set to 2 and 78.54 for the protein interior and solvent, respectively. The ion accessibility surface was calculated using a probe sphere with a radius of $2.0 \text{ \AA}$, representing monovalent counterions. Calculations were repeated with ionic strengths corresponding to 0 mM salt concentration (representing Coulombic interactions within the protein unscreened by solvent counterions and implicitly approximating the effects of dynamics) and 150

mM (representing physiological ionic strength in serum).

Electrostatic similarity distances (ESDs) were calculated according to

$$ESD_{a,b} = \frac{1}{N} \sum_{i,j,k} \frac{|\Phi_a(i,j,k) - \Phi_b(i,j,k)|}{\max\{|\Phi_a(i,j,k)|, |\Phi_b(i,j,k)|\}}$$

where Φ_a and Φ_b are the electrostatic potentials of proteins a and b at grid point (i, j, k) and N is the total number of grid points. This error-type relation compares the spatial distributions of electrostatic potentials of pairs of proteins. The normalization factor of the denominator assures small values in the vicinity of the 0-2 range, with 0 corresponding to identical spatial distributions of electrostatic potentials and 2 corresponding to totally different spatial distributions of electrostatic potentials. Visualization of the spatial distributions of electrostatic potentials, as isopotential contour surfaces, was performed using the program UCSF Chimera (43).

2.2.5 Clustering

The ESD shown above was also applied to cluster subtype sequences based on charge distribution maps using APBS, for the HIV-1 consensus sequences. Hierarchical clustering analyses were performed using the hclust function of R. The clustered data were plotted using the language and statistical computing environment R (Foundation for Statistical Computing: Vienna, Austria, 2009; <http://www.R-project.org>). The

clustering analyses allow for the classification of similarities/dissimilarities based on the common property of electrostatic potentials.

2.2.6 Molecular Dynamics Simulations

The two V3 loop crystal structures were used as templates, with the modifications described above to performed molecular dynamics simulations (MD). The V3 loop atomic charges, van der Waals and stereochemical parameters were taken from the CHARMM22 all-atom force field (44), including a CMAP backbone ϕ/ψ energy correction (45) and indole parameters (46). The disulfide bond between the first and last cysteine of the V3 loop was fixed using the patch DISU. The V3 loops were immersed in a water box of $45 \times 47 \times 64$ Å.

The structures were subject to 1000 steps of energy minimization prior to the simulations. All simulations were conducted with the program NAMD, versions 2.6 and 2.7 (47). A 100 ns simulation was run, per V3 loop structure, using explicit solvent model, under constant number of particles, pressure (1 atm), and temperature (298 K) conditions. The molecular dynamic time step was set to 2 ps. The total charge of the simulation was set to zero by the addition of sodium and chloride ions. Coordinates were sampled every 10 ps, to generate 10000 snapshots during the trajectory. Analysis of the trajectories was done by a series of in house scripts, using R programming language with analysis packages bio3d (48) in conjunction with Chimera (43).

2.2.6.1 Electrostatic Analyses of MD Trajectories

Electrostatic analyses using 100 snapshots from each trajectory, one snapshot per nanosecond, were performed as described in section 2.2.4 Electrostatic Analyses. The box dimensions for calculation of electrostatic potentials were $65 \text{ \AA} \times 65 \text{ \AA} \times 65 \text{ \AA}$, for both 0 and 150 mM ionic strength. Classical multidimensional scaling was performed using the `cmdscale` function of R. Multidimensional scaling converts a matrix of dissimilarities (ESD distances in our case) into a dimensional representation of a set of points, such that the distances between points are approximately equal to the dissimilarities.

2.3 Results and Discussion

2.3.1 Docking Studies

Molecular docking of the V3 loop and CCR5-Nt was performed using the Affinity Module of Insight II. Multiple runs were performed by positioning CCR5-Nt at the base, stem, and tip of the V3 loop for each set and energies were calculated. Table 2.2 shows the potential energy calculated for the initial docking studies, for structures 2B4C and 2QAD.

The electrostatic potential energies of the docked structures are always negative, independently of where the CCR5-Nt was initially placed, demonstrating that

the interaction between the V3 loop and CCR5-Nt is driven by electrostatics. The docked structure with the lowest total energy was observed when CCR5-Nt was at the base of the loop, in both structures. Conformational and orientational variability for CCR5-Nt is observed in the docked structures, as expected in docking studies (Figure 2.4). Figure 2.5 shows the docked structures with lowest energy for both structures.

Optimization was performed by 90° rotations of CCR5-Nt in each axis for each subset selected. Consistent with the result observed in Table 2.2, the electrostatic potential energies of all docked structures were always negative, independently of the binding subset and the rotation of CCR5-Nt. Tables 2.3-2.4 show the three lowest total potential energy docked structures, for CCR5-Nt rotations, for 2B4C and 2QAD, respectively. The lowest total potential energy structures were found when CCR5-Nt was at the base of the V3 loop, in both V3 loops, demonstrating the importance of the base of the loop. In 2B4C the lowest total potential energy was found when CCR5-Nt is rotated in the y-axis (Figure 2.6), while in 2QAD the lowest total potential energy was found when CCR5-Nt is rotated in the x-axis (Figure 2.7). These results demonstrate conformational and orientational variability for CCR5-Nt. These findings are in accordance with the docking studies carried out by Huang *et al* (12) with 2QAD-gp120 and CCR5-Nt (including the sulfotyrosines) that found the most favorable interaction to be at the base of the V3 loop. In 2B4C, negative potential energy values are also observed at the stem of the loop, this suggested that CCR5-Nt may interact with the base of the loop and the residues in the stem closest to the base might also be involved.

2.3.2 V3 Loop Electrostatics and HIV-1 Subtypes Year 2009

HIV is characterized by its ability to frequently mutate as evidenced by the large number of different isolates and by sequence diversity. Despite its hypervariable nature, V3 retains a basic function, to interact and modulate its preferential usage of CCR5 and CXCR4, a crucial step in the process of infection and indeed for the survival of the virus (18, 49). The combined electrostatic potentials of the amino acids in the V3 loop and their distribution in all HIV-1 subtypes, for which the tropism and V3 amino acid sequence are known, were explored in order to exploit canonical rules that might exist.

Electrostatic potential calculations and clustering analyses of the spatial distributions of electrostatic potentials for the gp120 V3 loop and several HIV-1 subtypes was performed. Electrostatic interaction is expected because the V3 loop typically has a positive charge whereas the interacting N-terminal domain of the coreceptors CCR5 and CXCR4, usually has a negative charge. It is actually the property of charge that many researchers have investigated to shed light into the V3 loop:CCR5/CXCR4 interaction. However, the electrostatic potential is responsible for recognition of two proteins if they have excess of opposite net charges. The underlying hypothesis is that if the electrostatic potentials and charges mediate protein-protein association, and if association mediates viral entry, we can deduce correlations to virulence by studying the specific properties of electrostatic potentials and charges, such as type (positive/negative), strength, and spatial distributions. These types of correlations are indications of where to look for causalities and may be helpful in

predicting viral attributes.

Alanine scans involving each ionizable residue within each V3 loop X-ray structure help understand their contribution to the electrostatic potential. Figures 2.8-2.9 show dendrograms for each V3 loop, with clustering based on similarities in spatial distributions of their electrostatic potentials, using two different ionic strengths 0 mM (depicting largest magnitudes of Coulombic interactions within each structure which are unscreened by solvent ions, Figures 2.8A and 2.9A) and 150 mM (corresponding to physiological ionic strength, Figures 2.8B and 2.9B). Mutants of residues of the same charge cluster together showing similar effects on the electrostatic potential, irrespective of their location in the sequence. Residues that are involved in intramolecular salt bridges within the V3 loop cluster together, depicting their electrostatic potential similarities (Arg3, Arg9, Arg31, Asp29, and Glu29; in finer subclusters or a supercluster). The salt bridges were explored during the MD simulations and are further discussed in section 2.3.3 V3 Loop Dynamics. Histidines, which are neutral in these studies, cluster together with the parent protein and have no electrostatic effect. The underlying assumption is that if electrostatics is important for binding, a mutation that affects electrostatics also affects binding.

Figure 2.10 shows the dendrogram that clusters the calculated spatial distributions of electrostatic potentials of V3 loop subtypes. These calculations were performed using 0 mM ionic strength, depicting largest magnitude of Coulombic

interactions within each structure. The calculations were performed using a homology model derived from the structure 2QAD and the HIV-1 subtype consensus sequences available in the year 2009 (described in Methods, Table 2.1). Clustering was performed by pairwise comparison of the electrostatic potentials of all subtypes listed in Table 2.1. V3 loop subtypes with similar spatial distribution of electrostatic potential cluster together. The V3 loops studied have positive net charge, with the exception of group O, which has a -1 net charge (Figure 2.10). The predominant net charge is +3, appearing in 9 subtypes (A, AE, AG, B, C, D35, G, F, K) and in the sequences of the two crystal structures, 2QAD and 2B4C, which belong to subtype B (Figure 2.10). From the remaining subtypes, group N has a net charge of +1 and AB, D, H, J, and CPX have net charge of +2 (Figure 2.10). The +2/+3 subtypes form a supercluster together. Within a cluster, subtypes with the same net charge form finer subclusters that are found for subtypes that discriminate according to the spatial distribution of electrostatic potentials. For example, from the +2 subtypes: AB and J cluster together; H and CPX cluster together (they are identical); and D clusters on its own. Overall, the +2 subtypes form the following cluster (with subclusters in brackets/parentheses): {[(J, AB), (H, CPX)], D} (Figure 2.10). Similarly, the +3 subtypes form the following cluster: { [(((G, AG), (K, B)), (2QAD, 2B4C), C), A], [(F, AE), D35]} (Figure 2.10). The +1 group N clusters on its own and forms a larger supercluster with the +2/+3 subtypes, whereas the -1 group O clusters entirely on its own (Figure 2.10).

In the dendrogram, generated with more realistic electrostatic potential

calculations using 150 mM ionic strength (physiological ionic strength), similar overall clustering with local variations are observed (Figure 2.11). For example, the +3 subtypes form the following cluster (with subclusters in brackets/parentheses): {[(F, AE), (D35, A)], [((G, AG), (K, B)), (2QAD, 2B4C)], C}. The +2 subtypes form individual clusters (D), (H, CPX), and (J, AB) within the +2/+3 supercluster. The +1 group N clusters on its own and forms a larger supercluster with the +2/+3 subtypes, whereas the -1 group O clusters entirely on its own (Figure 2.11). Coulombic interactions within the V3 loops are screened by solvent ions, which results in less obvious differences in the spatial distributions of electrostatic potentials when inspected visually (e.g., compare isopotential contours of Figure 2.10 to Figure 2.11). Nevertheless, persistent electrostatic clustering patterns are observed for the various subtypes, despite differences in their V3 loop sequences.

The clustering of the distribution of charges in space for each subtype and 2QAD as the structural template is shown in Figure 2.12. Some clusters within this dendrogram can be found in Figures 2.10-2.11 (e.g., H and CPX). However, the subtypes are mostly mixed within the +1/+2/+3 supercluster. In general, charge distribution does not depict subtle differences between the subtypes. This is because charges are localized in the structure and are independent from each other. However, electrostatic potentials, generated by these charges, have additional features. First, electrostatic potentials account for dielectric and ionic screening. Because of the latter, we observe differences in the magnitudes and shapes of electrostatic potentials in Figures 2.10-2.11. Second,

electrostatic potentials account for spatial enhancements (additive effect of potentials with same signs) or spatial cancellations (subtractive effect of potentials with opposite signs).

To assess the degree that V3 loop dynamics affect its electrostatic properties, at least using two extreme conformations of the crystal structures, clustering analyses for electrostatic potentials and charge were performed, using the 2B4C structure as template. Electrostatic potential clustering at 0 mM ionic strength (Figure 2.13) is similar to the corresponding data of the 2QAD structure (Figure 2.10). However, there are differences in the 150 mM data (Figure 2.11 and 2.14). For example, the +2 subtypes are scrambled within the +3 subtype clusters. The difference between the 150 mM clustering data from the two crystal structures originates from their conformational variability, which results in different charge distributions and different enhancements or cancellations of positive/negative electrostatic potential distributions. Such differences are not observed in the 0 mM data, because of lack of ionic screening, resulting in more uniform distribution of the dominant electrostatic potential (here being positive with the exception of subtype O). As in the case of 2QAD, in the 2B4C clustering of spatial distributions of charges does not depict the fine clustering of electrostatic potential similarities/dissimilarities (compare Figures 2.13-2.14 to Figure 2.15). Also, as in 2QAD, the 2B4C electrostatic clustering is more detailed, containing refined charge-related information not present in sequence clustering (compare Figures 2.13-2.14 to Figure 2.16).

Figures 2.10 and 2.13 also present correlations between the observed clusters and available epidemiological data on global prevalence and geographic distribution (year 2004, (8)), and coreceptor selectivity. Subtype C is responsible for almost 50% of the infected population (8). In the 0 mM data subtype C forms a cluster together with subtypes A, G, AG, K and B, together accounting for ~85% of the infected population (Figures 2.10 and 2.13). In the 150 mM data subtype C forms a cluster together with subtypes G, AG, K, and B, together accounting for ~73% of the infected population (subtype A, corresponding to ~12.3% of the infected population, moved to a neighboring cluster; Figure 2.11) when 2QAD is the structural template. However in the case of 2B4C subtype C cluster with the same subtype as in the 0 mM data (Figure 2.13). Geographic distributions (8) are also quoted in Figures 2.10 and 2.13.

Figure 2.16 shows clustering of the sequences of the gp120 V3 loops from the subtypes used to generate the data in Figures 2.10-2.15. This dendrogram does not, in general, depict the charge or the electrostatic potential differences of the various V3 loops. Obvious examples are the clusters (K, B, CPX, H) and (D35, D) that mix sequences with +2 and +3 net charges. These observations suggest that electrostatic clustering is more detailed, containing more refined charge-related information, than sequence clustering.

Because there are no X4-tropic consensus sequences in the 2009 data (Figures 2.10 and 2.13), sequence, glycosylation, and charge rules were used to present a

predictive scheme for coreceptor selectivity. The coreceptor selection by HIV-1 is known to be influenced by the charge of the V3 loop, amino acid types at specific locations, and the presence of glycosylation sites. Differences in coreceptor selection by HIV-1 subtypes have been shown by experimental studies (4, 5, 7, 50), and computationally predicted (50-52), although the effectiveness of the predictions is not conclusive. Based on previous studies and renewed thinking with respect to net charge, criteria known to affect coreceptor selection were used (as discussed in section 2.1) to develop a scheme to predict coreceptor selection, shown in Figure 2.17. This research suggests that if the $N^6X^7T^8S^8X^9$ glycosylation motif is present and any of the amino acids at positions 11, 24, and 25 are positively charged, coreceptor preference will be governed by the net charge of the V3 loop sequence. If the net charge of the V3 loop is > 5 , the virus will show preference toward CXCR4. Experimental studies have suggested that a high charge in the V3 is associated with loss of the glycosylation site and utilization of CXCR4 (4). However, if the net charge of the V3 loop is ≤ 5 , the virus will show preference for CCR5. Coreceptor selection will be affected by the presence and number of acidic chemical groups like sialic acids, in the glycans. Typically the glycans can have up to four sialic acids, each adding one negative charge to the loop (53). Thus, the presence of glycans may reduce the net charge of sequences with amino acid net charge of > 5 to ≤ 5 . This means that a sequence classified as X4-tropic based on amino acid net charge, can be reclassified as R5-tropic using net charge based on the amino acids and glycans. Because the number of sialic acids is not known, sequences falling in this category are classified

as X4-, R5-or dual-tropic (Figure 2.17). It should be distinguished that at lower V3 loop net charges (+3, +4), no effect was seen with alteration of N-glycosylation (4). This research suggest that if glycosylation takes place, it lowers the positive net charge even more and thus the sequence remains within the R5-tropic definition according to the scheme of Figure 2.17. The flow chart of Figure 2.17 was tested with experimental data for a series of R5- and X4-tropic sequences (4, 50) and consistency was found between the predicted and experimentally-derived tropisms. All consensus sequences studied here, and the sequence of 2QAD crystal structure, are R5-tropic according to the scheme of Figure 2.17, perhaps because CCR5 is the first viral preference for cell infection prior to switching to CXCR4. It is likely that as CCR5 receptors are being depleted, the virus evolves through mutational pressure by increasing the positive charge of the V3 loop for more efficient recognition of cells with CXCR4 receptors. A statistical model that includes the aforementioned characteristics was developed using V3 loop sequences with known coreceptor selection, as a tool for sequence based diagnostic for HIV (54).

2.3.3 V3 Loop Dynamics

We have performed MD simulations in search of persistent structural and physicochemical characteristics during two 100 ns trajectories; each initiated using different crystal structure (2B4C and 2QAD, respectively). Root mean square deviation

(RMSD) and root mean square fluctuation (RMSF) for each V3 loop structure are shown in Figure 2.18. Both structures are very flexible throughout the trajectories (RMSD values of 4 – 9 Å) and have similar RMSF profiles. The RMSF plots reveal that the GPG region, a highly conserved motif in the V3 loop, shows the highest fluctuations; this motif is found at the tip of the loop (further away from gp120) and as a result, it is expected to be highly mobile. In addition, the disulfide bridge and the residues closed to it are highly mobile (and perhaps flexible) in both systems. The 2B4C is somewhat more dynamic compared to 2QAD due to slightly larger fluctuations observed within the residues 24-26 in the former compared to the same region in the latter.

The secondary structure throughout the trajectories was determined using STRIDE (55) incorporated in bio3d, and is shown in Figures 2.19 and 2.20. The two available crystal structures of gp120 with intact V3 loop show different secondary structure (Figure 2.1): 2B4C contains a shorter β -hairpin at the tip, while 2QAD is a more rigid β -hairpin with longer beta strands, making the latter a thinner structure. It is the presence of β -hairpin structure and a β -turn (see below) that suggest that the dynamic character of the tip, as evidenced by RMSF values, is owed to mobility (as a whole) rather than flexibility (local fluctuations). Residues 2 and 34 are involved in β -bridge present in both structures owing to the two successive hydrogen bonds among His2 N – Thr34 O and Thr34 N – His2 O.

Within the MD simulations, after approximately the first 12 ns, extended β -hairpins are eliminated in 2B4C or reduced to a β -bridge among residues Thr8 – Ile26 (hydrogen bonds among atoms Ile27 N – Asn7 O and Arg9 N – Glu25 O) or Lys10 – Ile26 in 2QAD (hydrogen bonds among atoms Ile27 N – Arg9 O and Ser11 N – Glu25 O). The elimination and/or reduction of β -sheet content within the residue moiety 8-26 can be partly attributed to the deletion of the rest of the gp120 protein whose existence might offer increased stability to the loop. The initial 15-18 β -turn of the tip is entirely reproduced throughout both simulations, whereas fused 19-22 and 23-26 β -turns are mainly observed in 2B4C. β -turns within the stem residue moieties 4-10 and 27-32 are reproduced in both systems; both β -turn moieties in 2B4C and the latter moiety in 2QAD experience frequent interchanges with helical elements. Transient helices are also observed in both systems, mostly after the β -strands are lost. However closer to the end of both V3 loop trajectories, no defined secondary structure is observed. These results, in conjunction with the RMSD plots, reiterate the high dynamic character of the V3 loop.

Table 2.5 presents the hydrogen bonds with high occupancy, for both V3 loops, during the trajectories. The hydrogen bonding interactions stabilize the β -sheet elements and some of the β -turns encountered during the simulations. In addition, Table 2.6 presents the salt bridges found during the trajectories; salt bridges were calculated with a cut off of 5 Å. In 2B4C, a very stable salt bridge occurs between Arg9-Asp29 for 85.6% of the trajectory, located at the stem of the loop. Another salt bridge between Arg3-Asp29 occurs 22.6% of the time and is located at the base of the loop.

These results show that the salt bridge partner of Asp29 changes through the trajectory, however only 7.9% of the time both salt bridges occur simultaneously, making it a bifurcated salt bridge, compared with 88.5% of the time when only one salt bridge is present. In the case of 2QAD, two salt bridges are observed: Arg31-Asp29 for 91% of the time, present at the base of the loop, and Arg9-Glu25 for 57.5%, present at the stem of the loop. The salt bridge involving Arg9-Glu25 is located within the β -sheet of 2QAD. Occupancy of both salt bridges present occurred 59.1% of the time, while 30.3% of the time only one salt bridge was present.

Interestingly, salt bridges linking the two opposite stem regions exist in both systems, Arg9-Asp29 in 2B4C and Arg9-Glu25 in 2QAD. The specificity for the Arg9 partner could be attributed to the conformational selectivity that is caused by mutations. A characteristic pattern featuring a positive charge in position 9 in combination with a negative charge in position 25 and 29 is observed with relatively high propensity in V3 loops recognizing CCR5 coreceptors ((3, 5) unpublished). In addition, a study by Rosen *et al* (56) demonstrates that Arg9 and Glu25 are close in proximity, thus facilitating the β -sheet formation. Therefore, these charged residue interactions are likely a contributing factor in the stabilization of structure and motions of the V3 loop as well as the coreceptor selectivity.

Interactions between all amino acids side chains were calculated, averaged through the trajectory, and plotted. The contact plots are shown in Figures 2.21 and

2.22. For both systems the residues that formed the transient helices closely interact together, independently of the presence of the hydrogen bonds that form the helices. More interactions are present in structure 2QAD than 2B4C (compare Figure 2.22 with Figure 2.21), due to their differences in secondary structure. In 2B4C, Phe20 interacts with a cluster of residues (Ile12, Ile14, Arg18, Ala19, Tyr21) helping to stabilize the turn in conjunction with a hydrogen bond between Gly15-Arg18 and π -cation interaction between Arg18-Phe20. In 2QAD, the network of interactions among residues 7-13 with 22-28, not present in 2B4C, are responsible for its thinner structure, while in 2B4C the Phe20 side chain points to the inside of the loop, causing it to be wider.

Further analysis of both systems has shown that the motions observed within both trajectories are fairly similar, thus involving the same regions within both structures (57). Principal component analysis shows that the first principal component accounts for most of the observed motions within both structures (54.7% in 2B4C and 48.5% in 2QAD, (57)). The results suggest that as the stem regions approach, the tip and base come further apart, and vice versa. Also, as the stem regions approach, the Arg9-Asp29 (in 2B4C) and Arg9-Glu25 (in 2QAD) salt bridges are formed. In the opposite case, when the stem regions are further apart, the Arg3-Asp29 salt bridge in 2B4C forms, while the Asp29-Arg31 salt bridge in 2QAD is present throughout the various motions. These findings, in addition to the presence of salt bridges involving common players in both systems, suggest that despite different sequences and secondary structures, there

are persistent motions and charged interactions, which may be related to binding and coreceptor selection.

Electrostatic analysis using 100 snapshots from each trajectory, one snapshot per nanosecond, was performed as described in section 2.2.6.1 Electrostatic Analyses of MD Trajectories. Figures 2.23-2.24 show the plots that cluster the spatial distributions of electrostatic potentials of the snapshots, using 0 mM and 150 mM ionic strength, respectively. Overlap between the electrostatic potential of snapshots is observed at both ionic strengths. For example, in Figure 2.23 snapshots from the beginning of the 2B4C simulation (red, around 5-30 ns) are found together with snapshots towards the beginning and end of the 2QAD simulation (green, 18-36 ns and 86-100 ns). A similar trend is observed in Figure 2.24. The overlap of snapshots depicts instances in which both V3 loops have very similar electrostatic potentials. Despite sequence and structural variability, transient similarity of electrostatic potentials during the MD trajectories reaffirms the notion that electrostatics plays a common role in the interaction between different V3 loop sequences and coreceptors.

2.4 Figures and Tables

Figure 2.1: Molecular models of the V3 loop. (A) Stick representation of backbone and side chains using structure 2B4C. (B) Ribbon representation of backbone using structure 2B4C. (C) Stick representation of backbone and side chains using structure 2QAD. (D) Ribbon representation of backbone using structure 2QAD. The color code for (A) and (C) is: blue, positively charged; red, negatively charged; green, polar; gray, nonpolar. The color code for (B) and (D) is according to secondary structure.

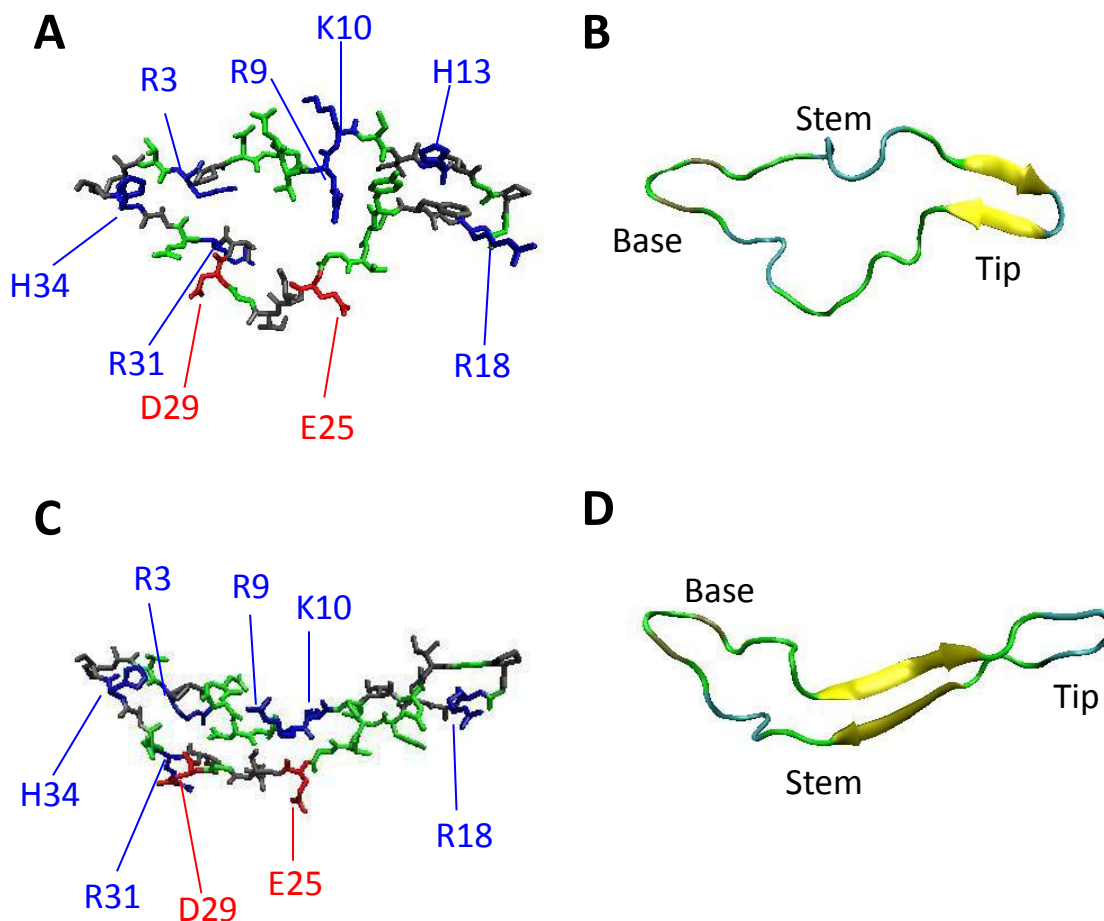


Figure 2.2: V3 loop residues influencing coreceptor selection for 2B4C (A) and 2QAD (B). Backbone is shown in grey with important side chain residues color coded: yellow denotes residues 1 and 35, forming the disulfide bridge; blue denotes residues involve in the glycosylation motif; red denotes residues involved in the “11/24/25” rule; green denotes the conserved GPG motif at the tip of the loop.

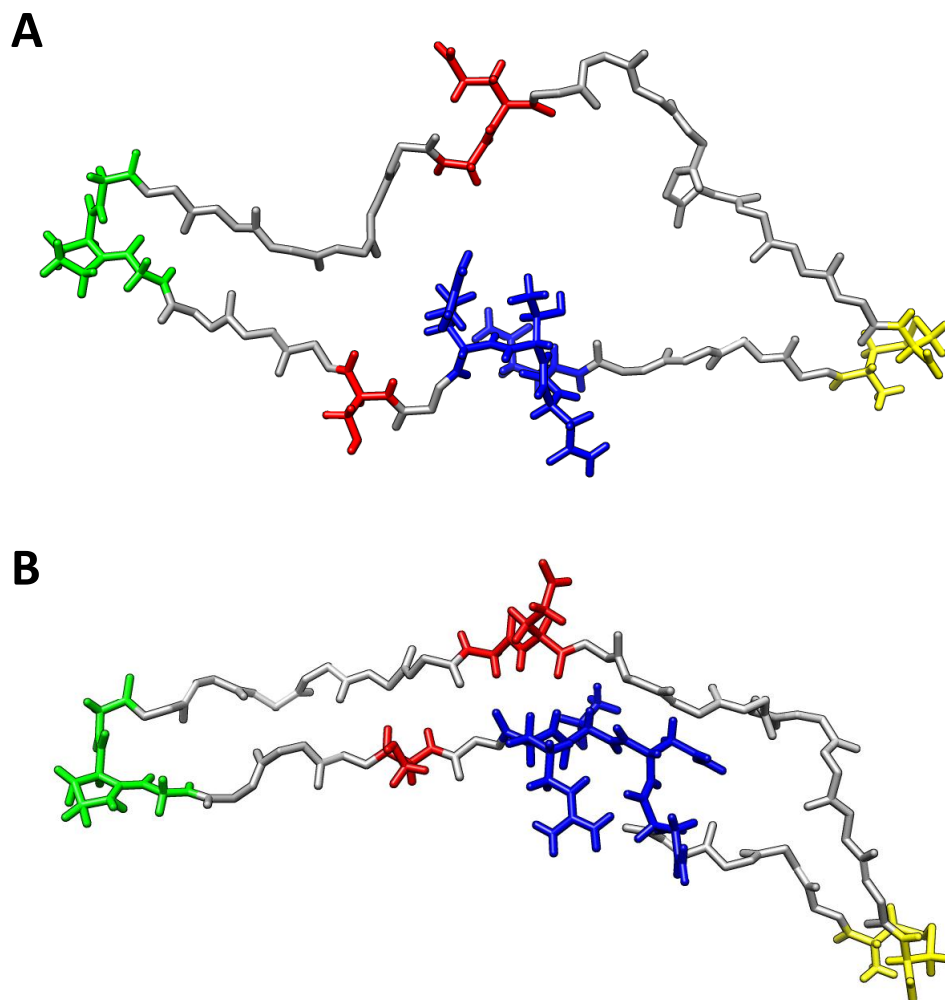


Figure 2.3: HIV-1 Classification. The groups are major (M), non-major and non-outlier (N) and outlier (O). CRFs stand for circulation recombinant forms.

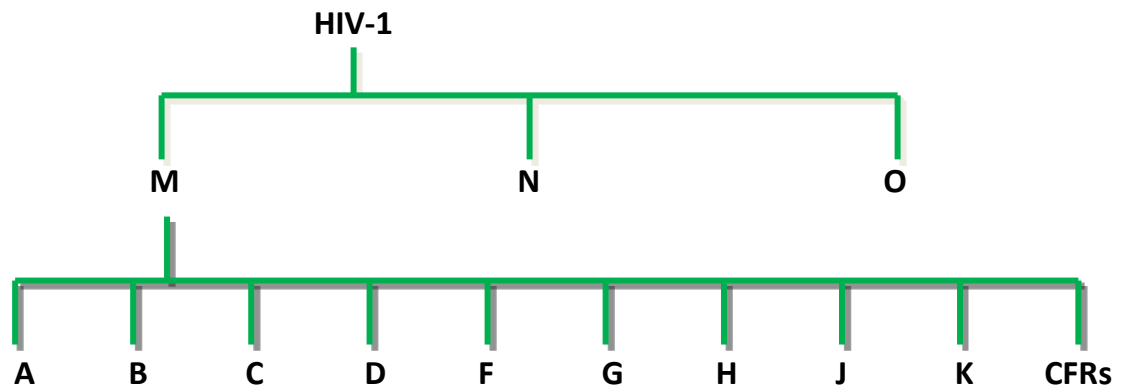


Figure 2.4: Conformational and orientational variability of CCR5-Nt. Structures are shown in ribbon representation with 2B4C at the left and 2QAD at the right. CCR5-Nt at the base of the loop is shown in (A) 2B4C and (D) 2QAD. CCR5-Nt at the stem of the loop is shown in (B) 2B4C and (E) 2QAD. CCR5-Nt at the tip of the loop is shown in (C) 2B4C and (F) 2QAD.

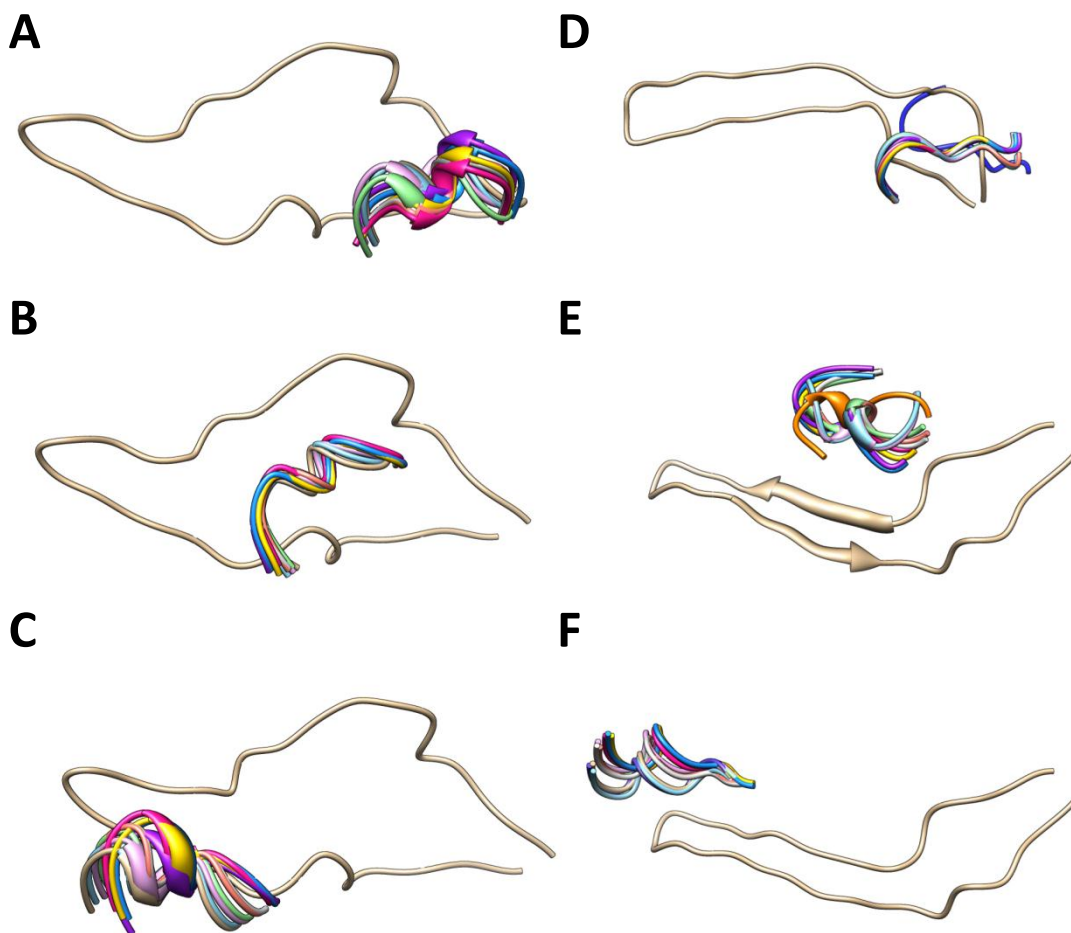


Figure 2.5: Docked structures with the lowest total energy. Backbone is shown in ribbon representation with side chains shown in wire representation. Disulfide bond involving residues 1 and 35 is colored brown. (A) 2B4C colored green with 2RLL colored blue. (B) 2QAD colored purple with 2RLL colored grey.

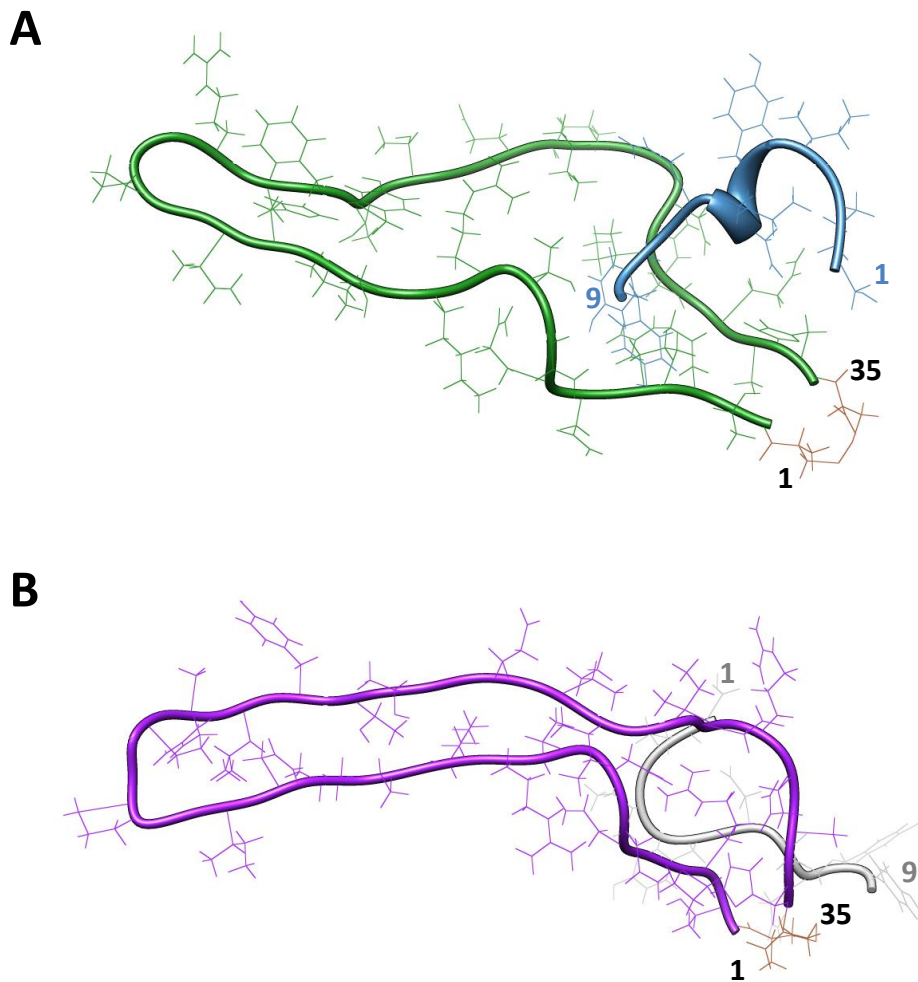


Figure 2.6: Docked structures of 2B4C:CCR5-Nt base rotations in the y-axis, with the lowest total energy. 2B4C backbone is shown in ribbon representation colored grey with side chains shown in wire representation. Disulfide bond involving residues 1 and 35 is colored brown. Three lowest energy structures from (A) first rotation in y-axis with CCR5-Nt colored green for structure #8, red for #9 and blue for #10; and (B) second rotation in y-axis with CCR5-Nt colored orange for structure #7, cyan for #9 and pink for #10. CCR5-Nt structures are numbered as in Table 2.3.

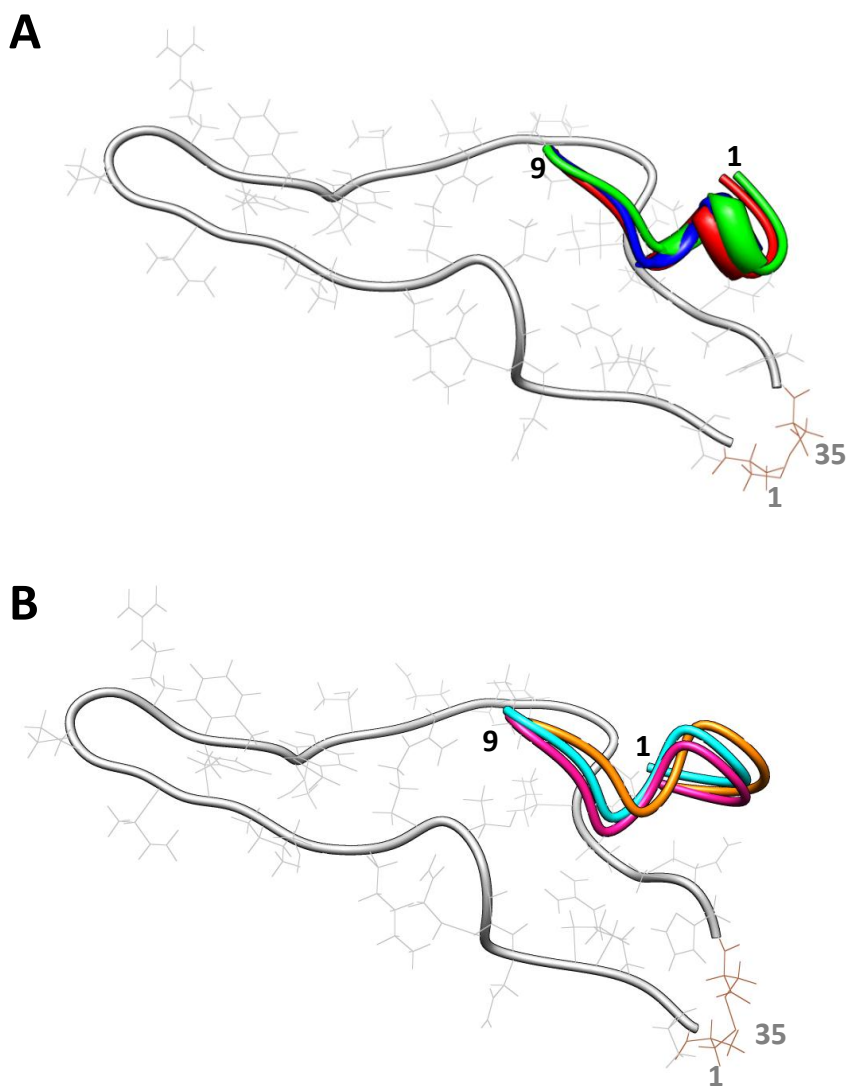
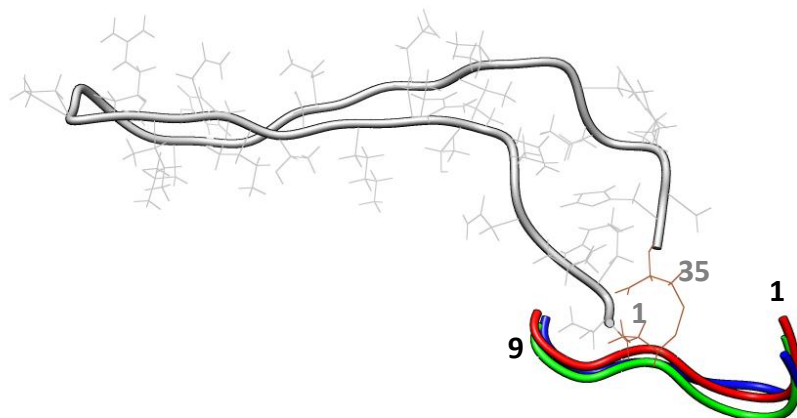


Figure 2.7: Docked structures of 2QAD:CCR5-Nt base rotations in the x-axis, with the lowest total energy. 2QAD backbone is shown in ribbon representation colored grey with side chains shown in wire representation. Disulfide bond involving residues 1 and 35 is colored brown. Three lowest energy structures from (A) first rotation in y-axis with CCR5-Nt colored red for structure #7, green for #8 and blue for #9; and (B) second rotation in y-axis with CCR5-Nt colored orange for structure #3, cyan for #4, and pink for #5. CCR5-Nt structures are numbered as in Table 2.4.

A



B

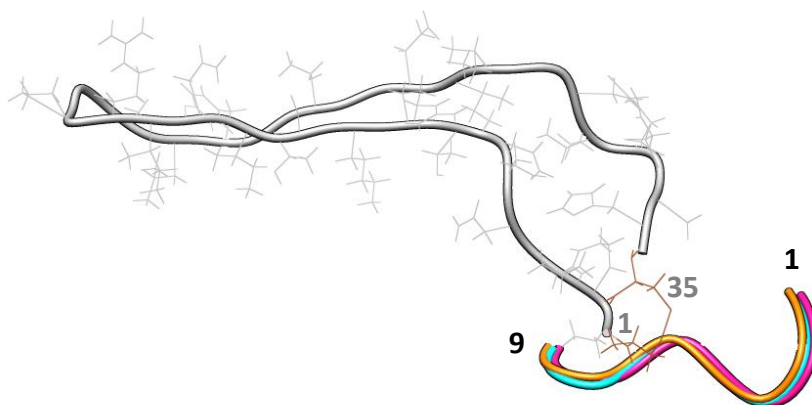


Figure 2.8: Electrostatic clustering of V3 loop mutants for structural template 2B4C. The vertical axis denotes residue and sequence number that was mutated into alanine. The horizontal axis of the dendrogram represents electrostatic similarity distance. Electrostatic potentials were calculated using ionic strength corresponding to (A) 0 mM and (B) 150 mM. Isopotential contours are plotted at ± 1 $k_B T/e$, with blue and red corresponding to positive and negative electrostatic potentials, respectively. The net charge is indicated in the figure. Green circles in the branches of the dendrogram denote intersection points between net charges.

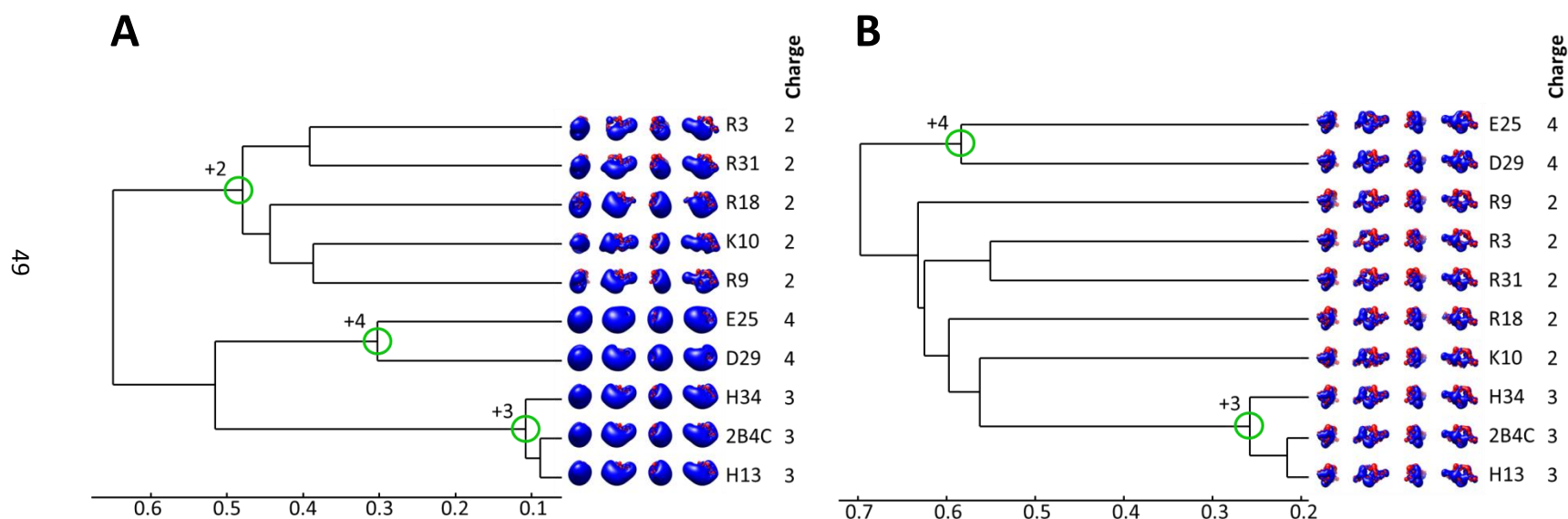


Figure 2.9: Electrostatic clustering of V3 loop mutants for structural template 2QAD. The vertical axis denotes residue and sequence number that was mutated into alanine. The horizontal axis of the dendrogram represents electrostatic similarity distance. Electrostatic potentials were calculated using ionic strength corresponding to (A) 0 mM and (B) 150 mM. Isopotential contours are plotted at ± 1 $k_B T/e$, with blue and red corresponding to positive and negative electrostatic potentials, respectively. The net charge is indicated in the figure. Green circles in the branches of the dendrogram denote intersection points between net charges.

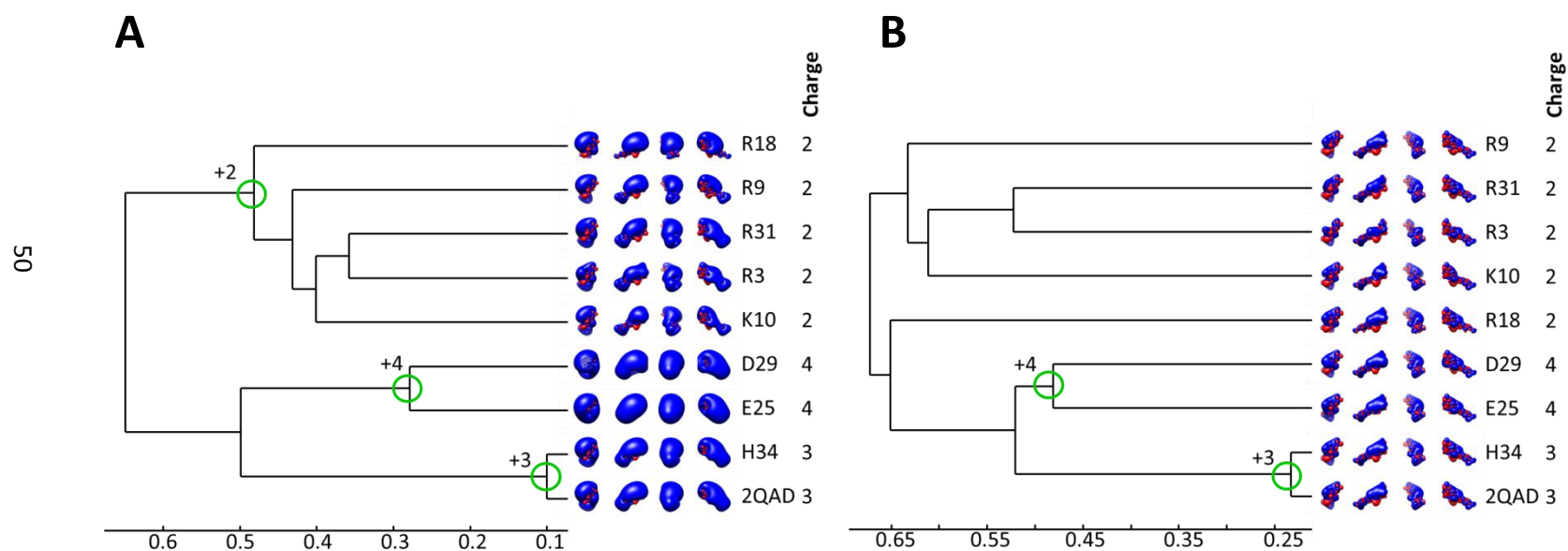


Figure 2.10: Electrostatic clustering of HIV-1 subtypes, from consensus sequences and structural template 2QAD. The horizontal axis of the dendrogram represents electrostatic similarity distance. Electrostatic potentials were calculated using ionic strength corresponding to 0 mM salt concentration. Isopotential contours are presented in four different orientations, corresponding to rotations about the vertical axis (indicated in the figure). Isopotential contours are plotted at $\pm 1 k_B T/e$, with blue and red corresponding to positive and negative electrostatic potentials, respectively. The net charge, global prevalence, geographic distribution, and coreceptor selectivity are indicated in the figure for each subtype. N/A denotes that information was not available. The orange boxes highlight clusters with HIV-1 subtypes that have similar electrostatic potential and same charge. Green circles in the branches of the dendrogram denote intersection points between net charges or infected population. The symbol # refers to the global prevalence of Subtype D, which includes D and D35 combined.

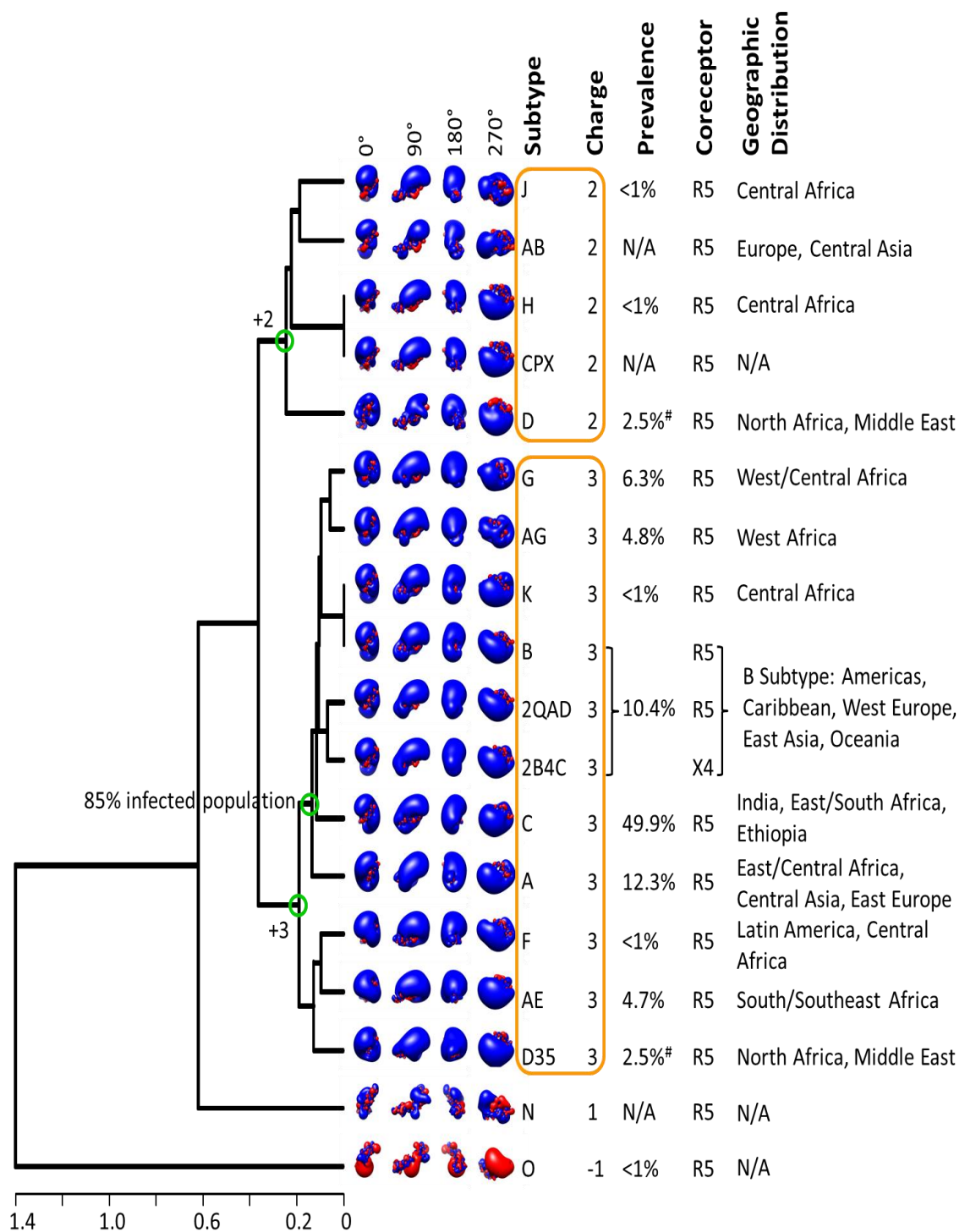


Figure 2.11: Electrostatic clustering analysis of HIV-1 subtypes, from consensus sequences and structural template 2QAD. The horizontal axis of the dendrogram represents electrostatic similarity distance. Electrostatic potentials were calculated using ionic strength corresponding to 150 mM salt concentration. Isopotential contours are presented in four different orientations, corresponding to rotations about the vertical axis (indicated in the figure). Isopotential contours are plotted at $\pm 1 k_B T/e$, with blue and red corresponding to positive and negative electrostatic potentials, respectively. The orange box highlights clusters with HIV-1 subtypes that have similar electrostatic potential and same charge. Green circles in the branches of the dendrogram denote intersection points between net charges or infected population.

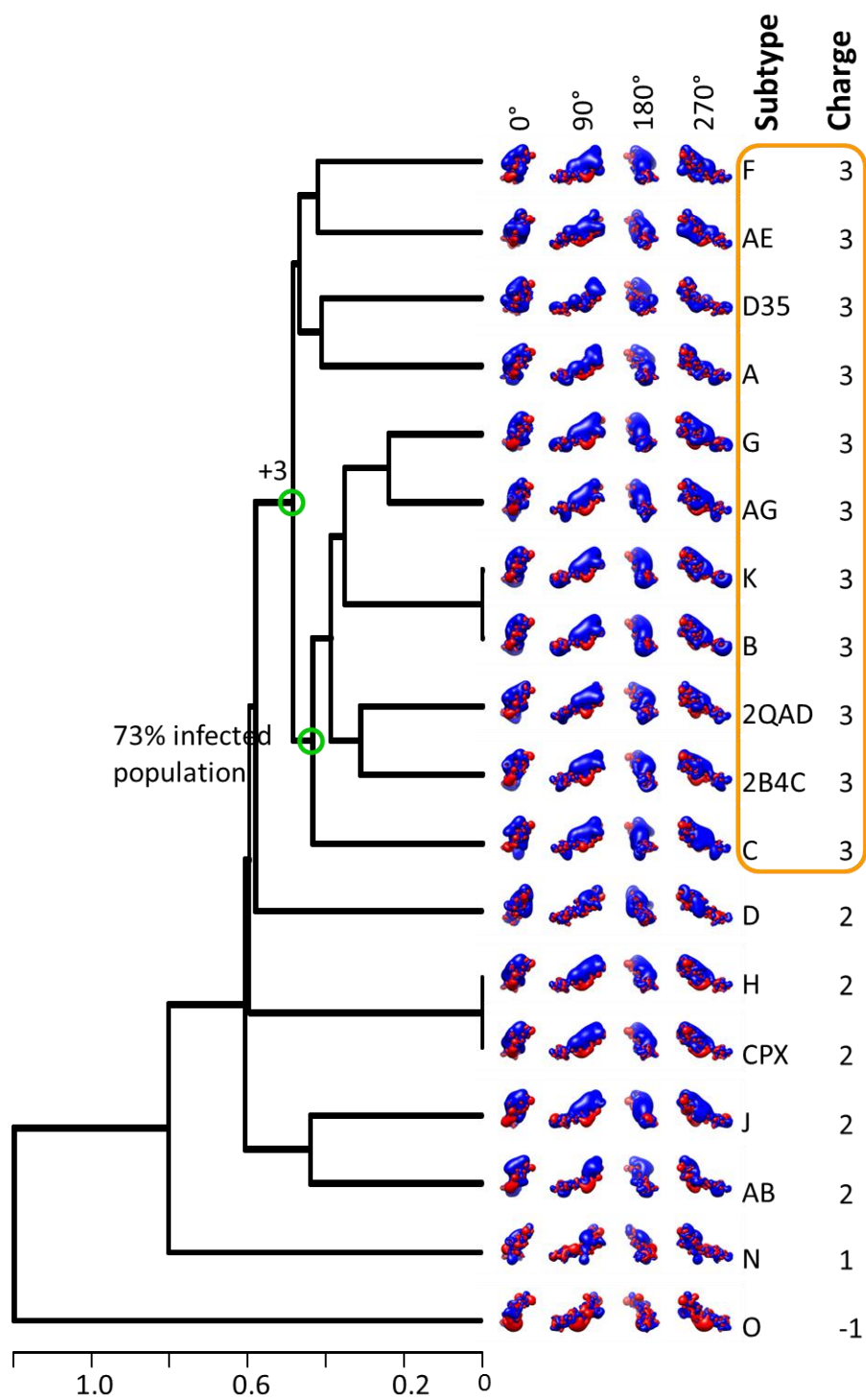


Figure 2.12: Charge distribution clustering of HIV-1 subtypes, from consensus sequences and structural template 2QAD. The horizontal axis of the dendrogram represents charge similarity distance. The net charge, global prevalence, coreceptor selectivity and geographies of each subtype are indicated in the figure for each subtype. N/A denotes that information was not available. Green circles in the branches of the dendrogram denote intersection points between net charges or infected population. The symbol * refers to the global prevalence of Subtype B which includes the two crystal structural templates (from 2QAD and 2B4C). The symbol # refers to the global prevalence of Subtype D which includes D and D35, combined.

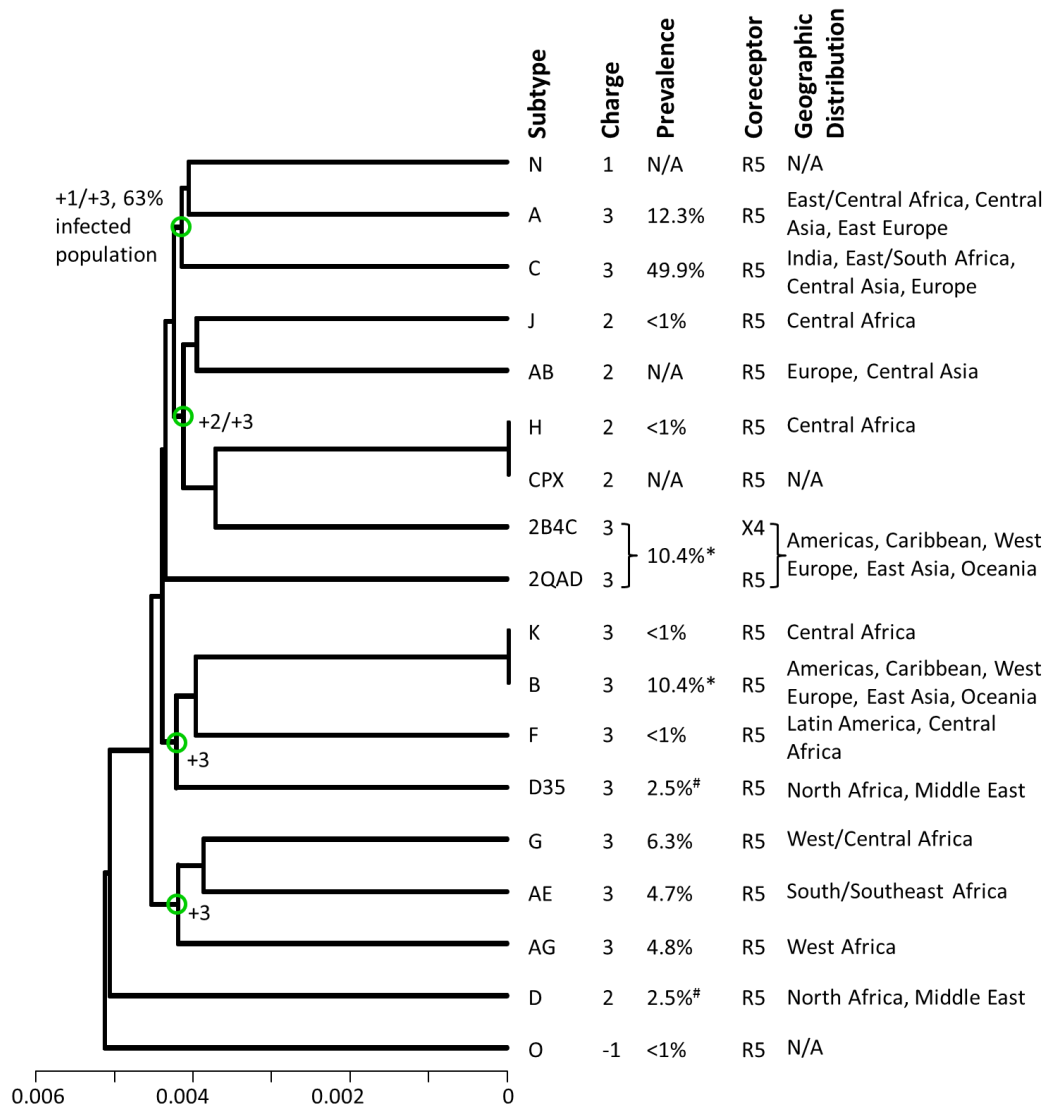


Figure 2.13: Electrostatic potential clustering of HIV-1 subtypes, from consensus sequences and structural template 2B4C. The horizontal axis of the dendrogram represents electrostatic similarity distance. Electrostatic potentials were calculated using ionic strength corresponding to 0 mM salt concentration. Isopotential contours are presented in 4 different orientations, corresponding to rotations about the vertical axis. Isopotential contours are plotted at $\pm 1 k_B T/e$, with blue and red corresponding to positive and negative electrostatic potentials, respectively. The net charge, global prevalence, coreceptor selectivity, and geographic distribution are indicated in the figure for each subtype. N/A denotes that information was not available. The orange boxes highlight clusters with HIV-1 subtypes that have similar electrostatic potential and same charge. Green circles in the branches of the dendrogram denote intersection points between net charges or infected population. The * refers to the global prevalence of subtype B, which include the two structural templates (2B4C and 2QAD). The # refers to the global prevalence of subtype D, which include D and D35.

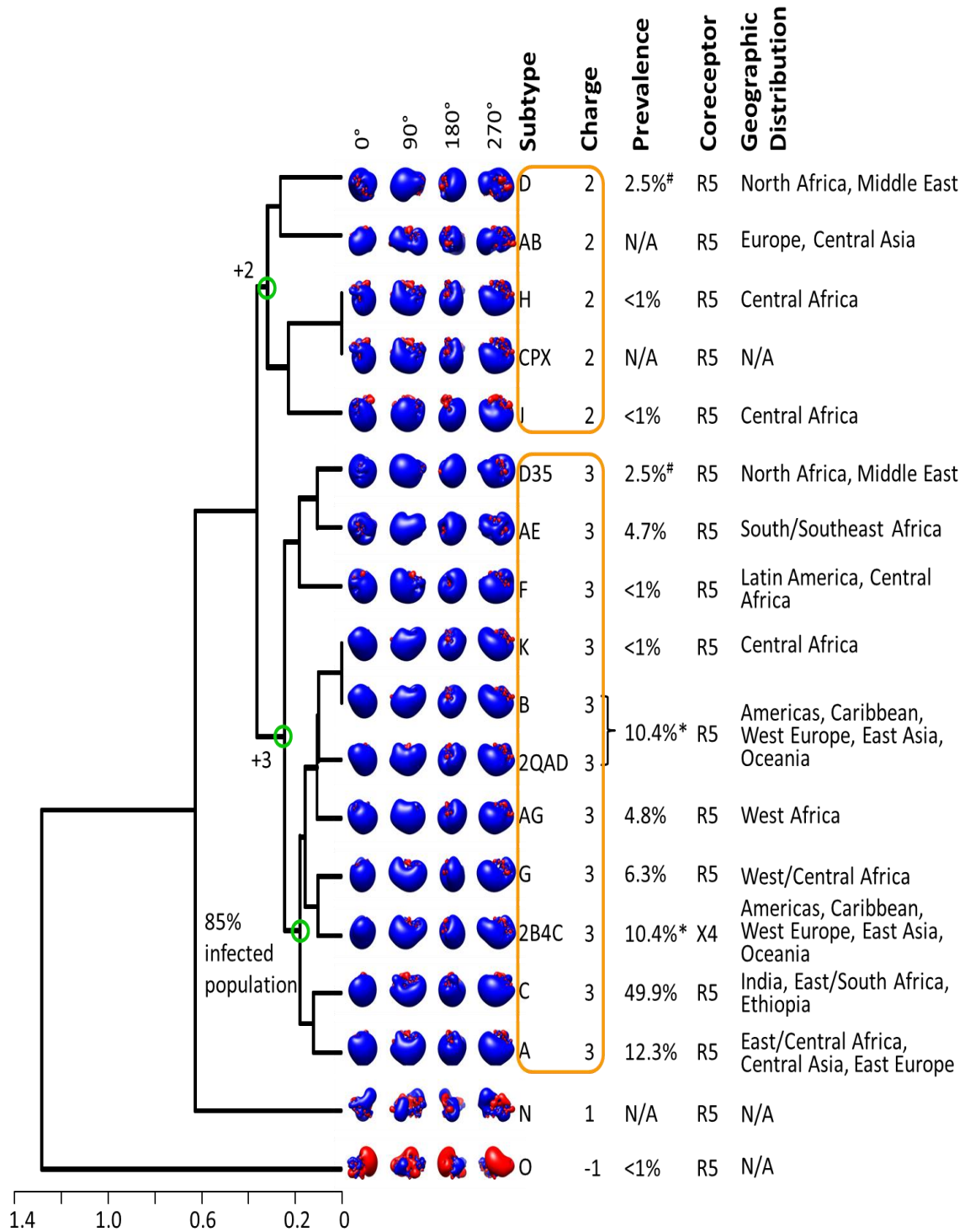


Figure 2.14: Electrostatic potential clustering of HIV-1 subtypes, from consensus sequences and structural template 2B4C. The horizontal axis of the dendrogram represents electrostatic similarity distance. Electrostatic potentials were calculated using ionic strength corresponding to 150 mM salt concentration. Isopotential contours are presented in 4 different orientations, corresponding to rotations about the vertical axis. Isopotential contours are plotted at $\pm 1 k_B T/e$, with blue and red corresponding to positive and negative electrostatic potentials, respectively. The orange box highlight clusters with HIV-1 subtypes that have similar electrostatic potential and same charge. Green circles in the branches of the dendrogram denote intersection points between net charges or infected population.

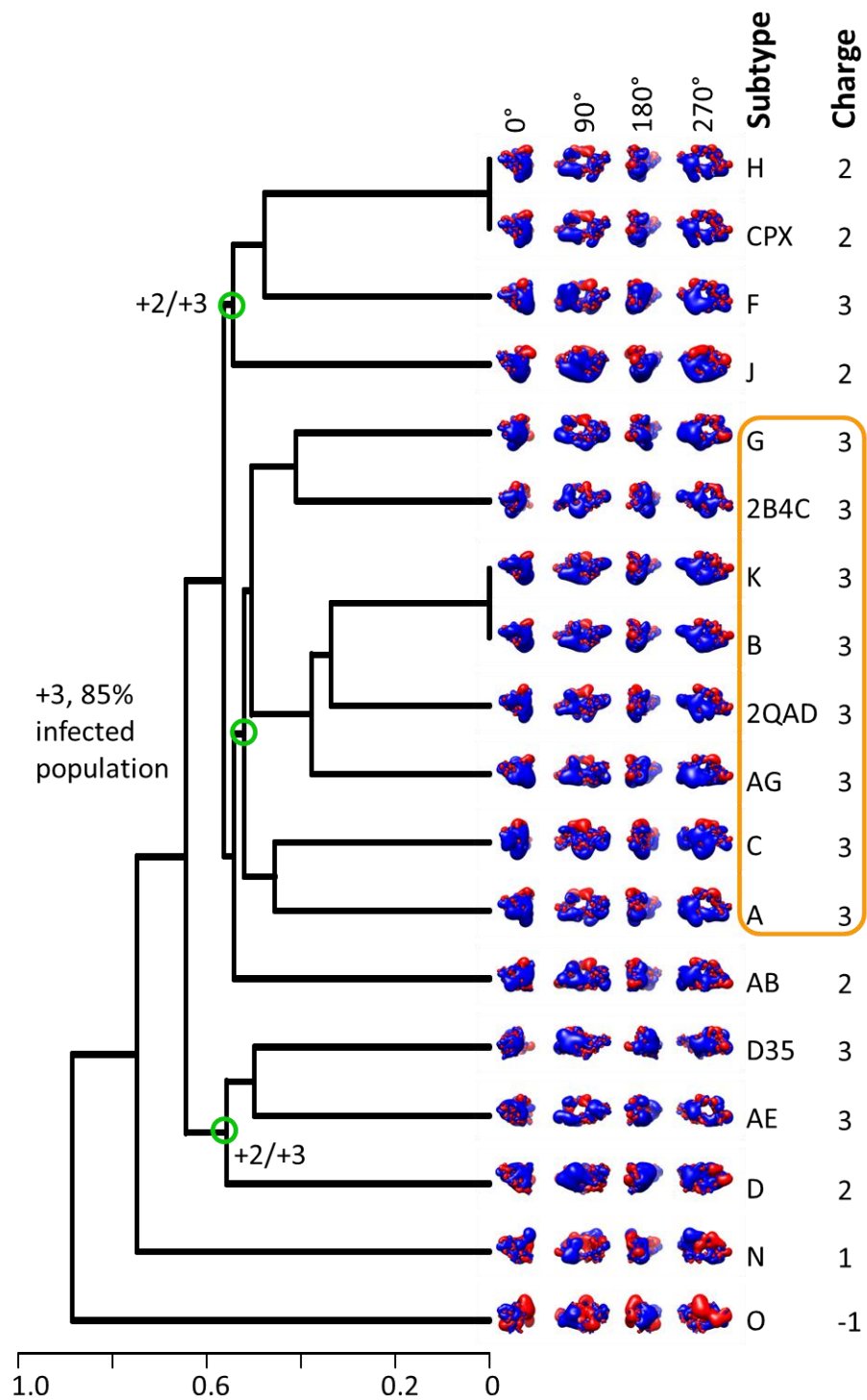


Figure 2.15: Charge distribution clustering of HIV-1 subtypes, from consensus sequences and structural template 2B4C. The horizontal axis of the dendrogram represents charge similarity distance. The net charge, global prevalence, coreceptor selectivity and geographical distribution are indicated in the figure for each subtype. N/A denotes that information was not available. Green circles in the branches of the dendrogram denote intersection points between net charges or infected population. The * refers to the global prevalence of subtype B, which include the two structural templates (2B4C and 2QAD). The # refers to the global prevalence of subtype D, which include D and D35.

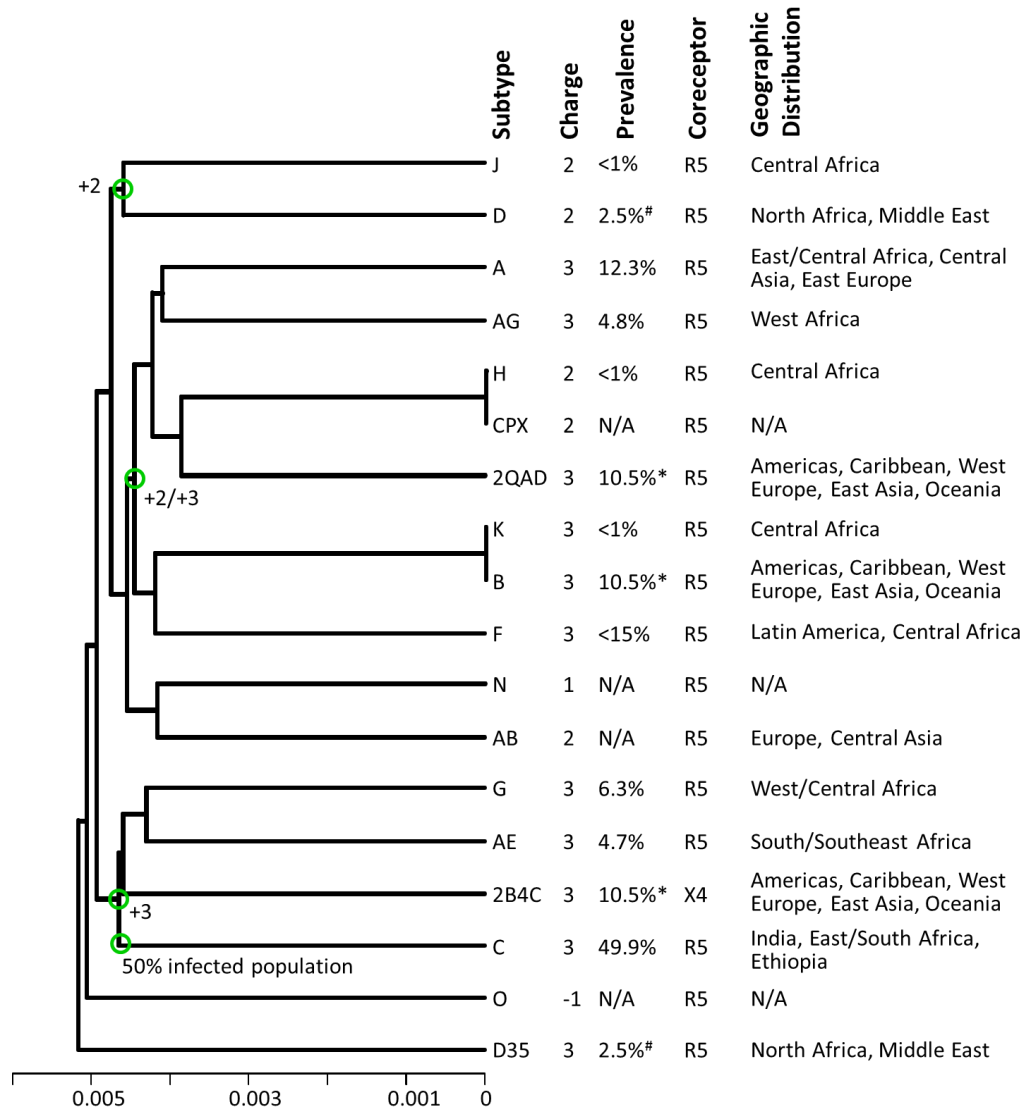


Figure 2.16: Sequence clustering of HIV-1 subtypes, from consensus sequences, based on sequence similarity. The horizontal axis of the dendrogram represents sequence similarity distance. Global prevalence, coreceptor selectivity and geographic distribution of each subtype are indicated in the figure. N/A denotes that information was not available. The green box highlights sequences that belong to Subtype D, while the orange box highlights the two crystal structural templates (from 2QAD and 2B4C), which belong to subtype B. The * refers to the global prevalence of Subtype B which includes the two crystal structure templates.

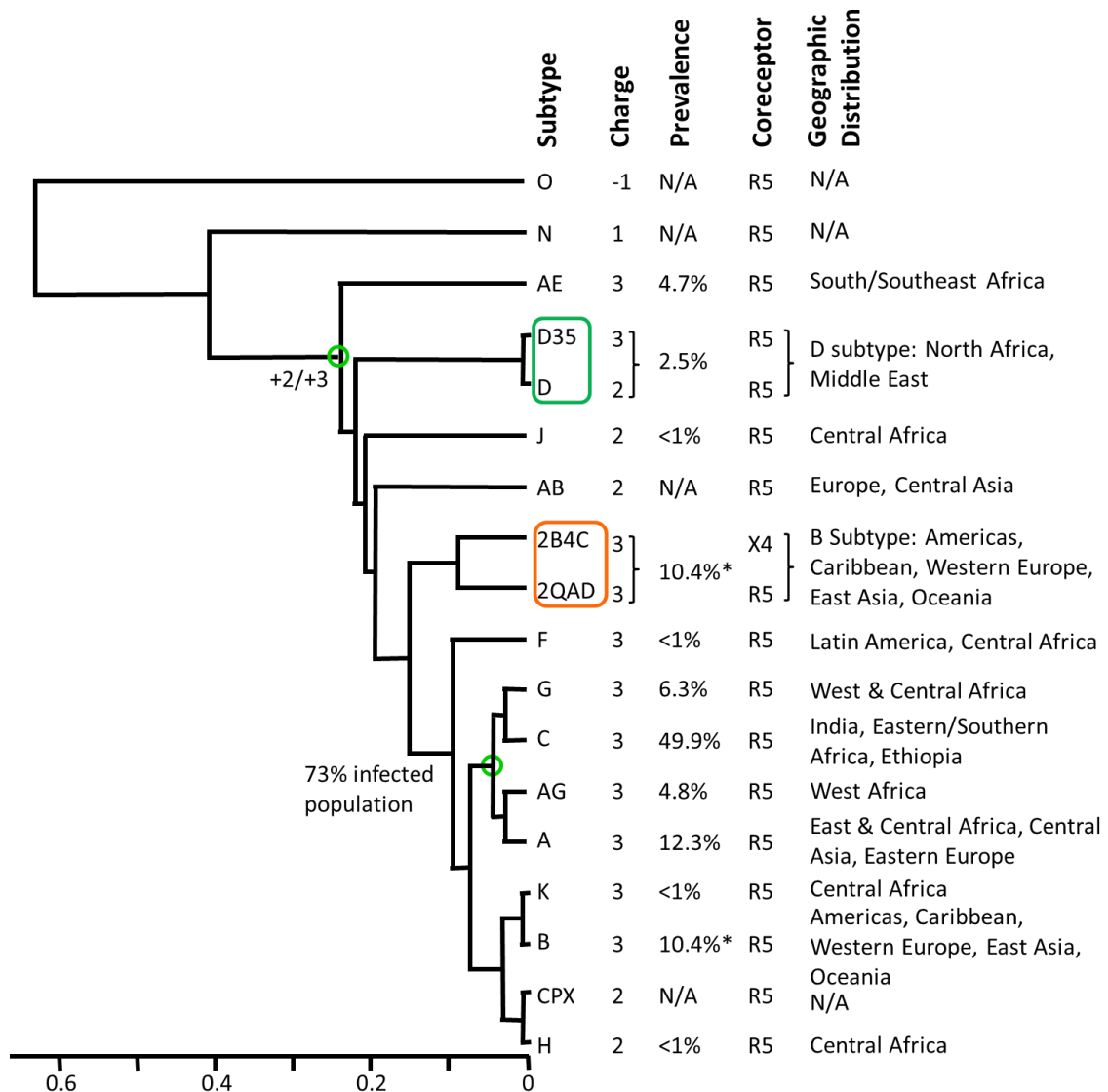


Figure 2.17: Flow chart for prediction of HIV-1 coreceptor selectivity based on V3 loop sequence and charge properties. This scheme is based on the presence of the $N^6X^7T^8|S^8X^9$ sequence/glycosylation motif (4), the presence of a positive amino acid at sequence positions 11, 24, and 25 (the 11/24/25 rule) (5), and the net charge. The presence of acidic chemical groups in the glycosylation patterns (e.g., sialic acids) could affect the charge of the V3 loop, thus affecting the coreceptor selection. Therefore, the virus can use CXCR4, CCR5 or both receptors for cell entry (dual tropic).

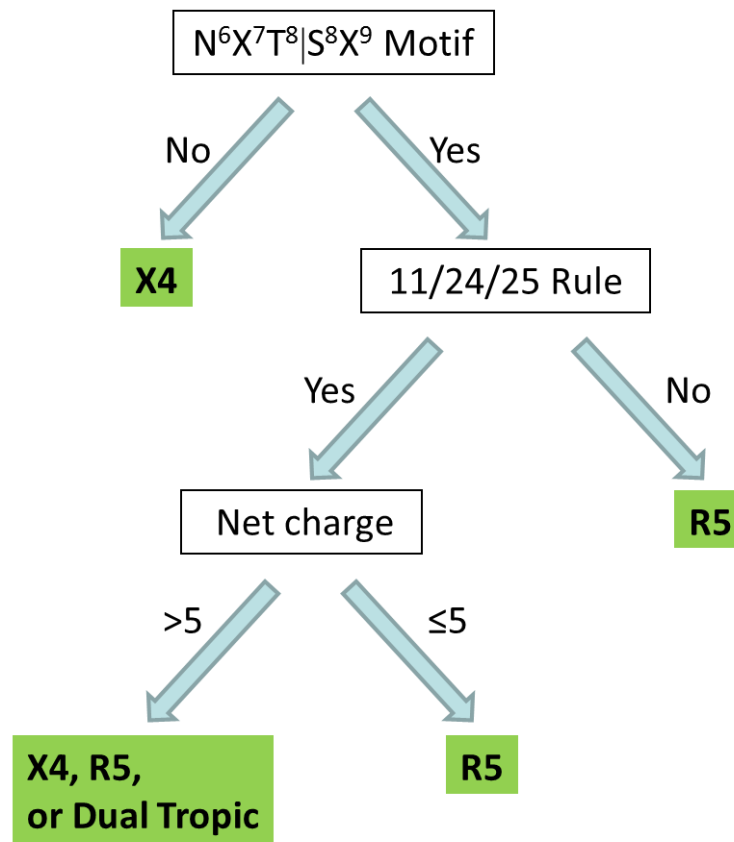


Figure 2.18: Root mean square deviation (RMSD) and root mean square fluctuations (RMSF), in Å, for the MD trajectories. (A) RMSD for 2B4C, (B) RMSF for 2B4C, (C) RMSD for 2QAD, and (D) RMSF for 2QAD. The colored code for the RMSF is: black, C-alphas; red, backbone atoms; green, side chain atoms; and blue, all the atoms per residue.

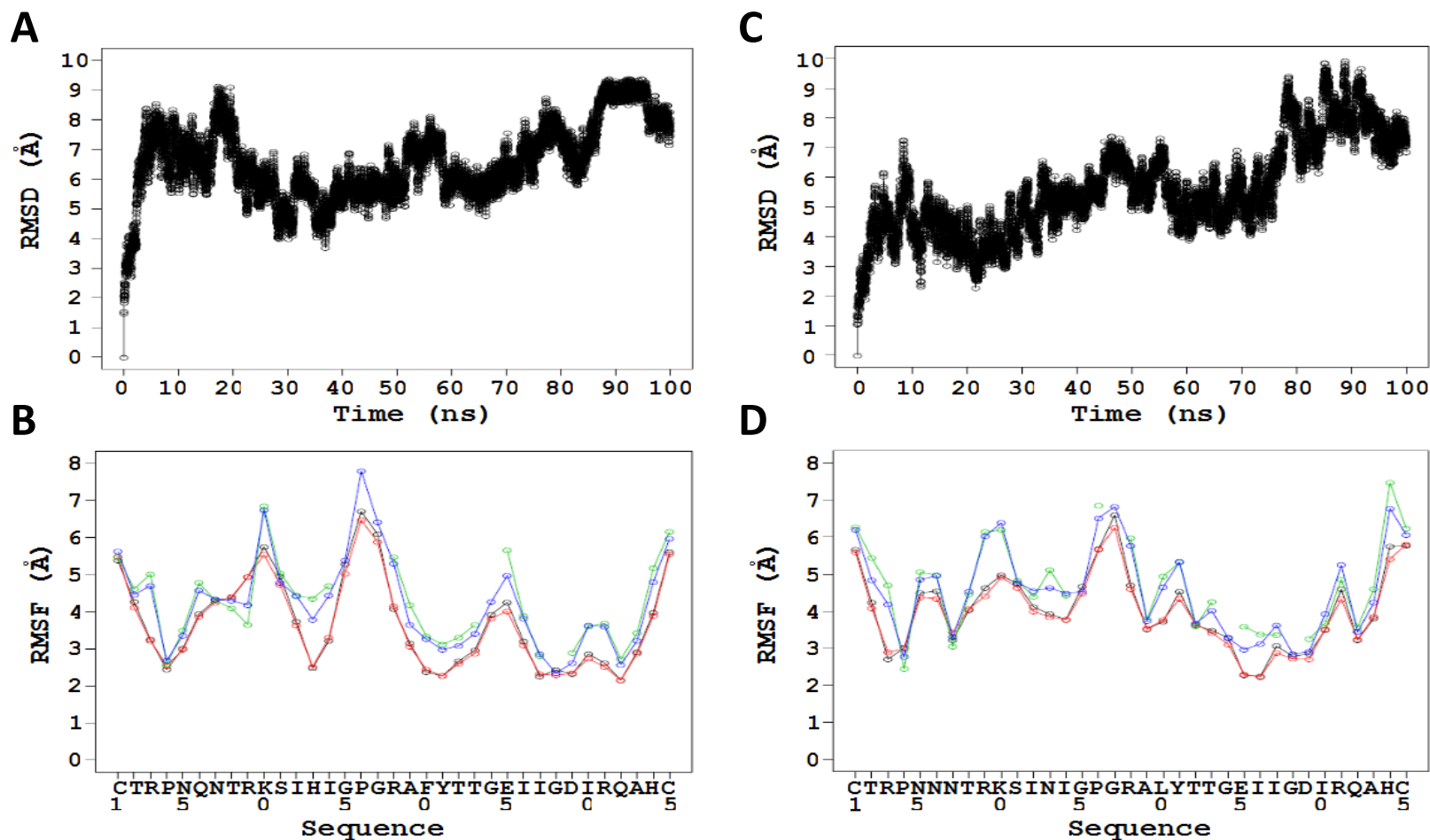


Figure 2.19: Secondary structure analysis for structural template 2B4C, calculated through the trajectory. The colored code is: orange, no defined structure; red, turn; yellow, beta strands; and white, helices. The y-axis represent time, which each line representing 1ns, starting at 0ns at the right and ending at 100ns at the left.

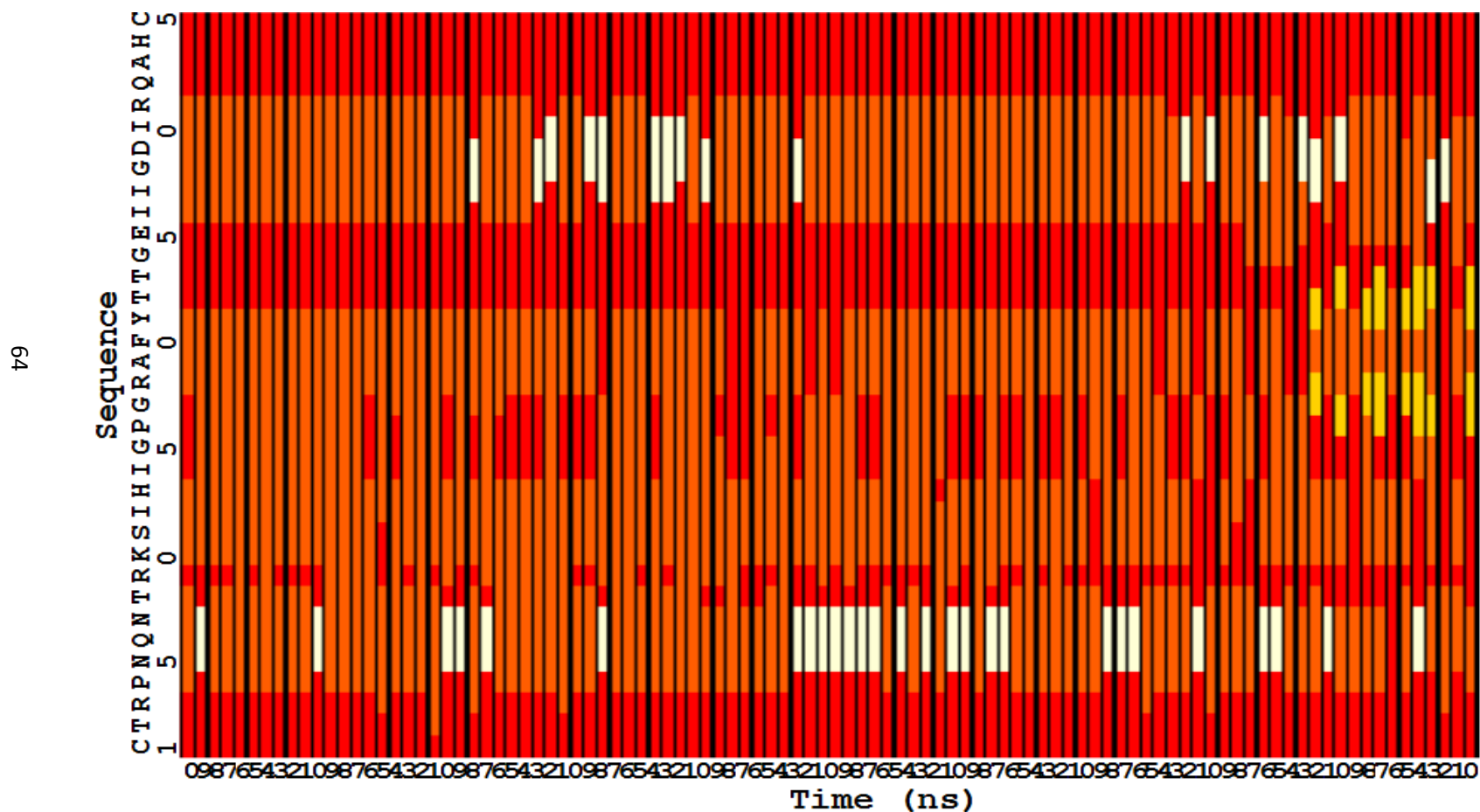


Figure 2.20: Secondary structure analysis for structural template 2QAD, calculated through the trajectory. The colored code is: orange, no defined structure; red, turn; yellow, beta strands; and white, helices. The y-axis represent time, which each line representing 1ns, starting at 0ns at the right and ending at 100ns at the left.

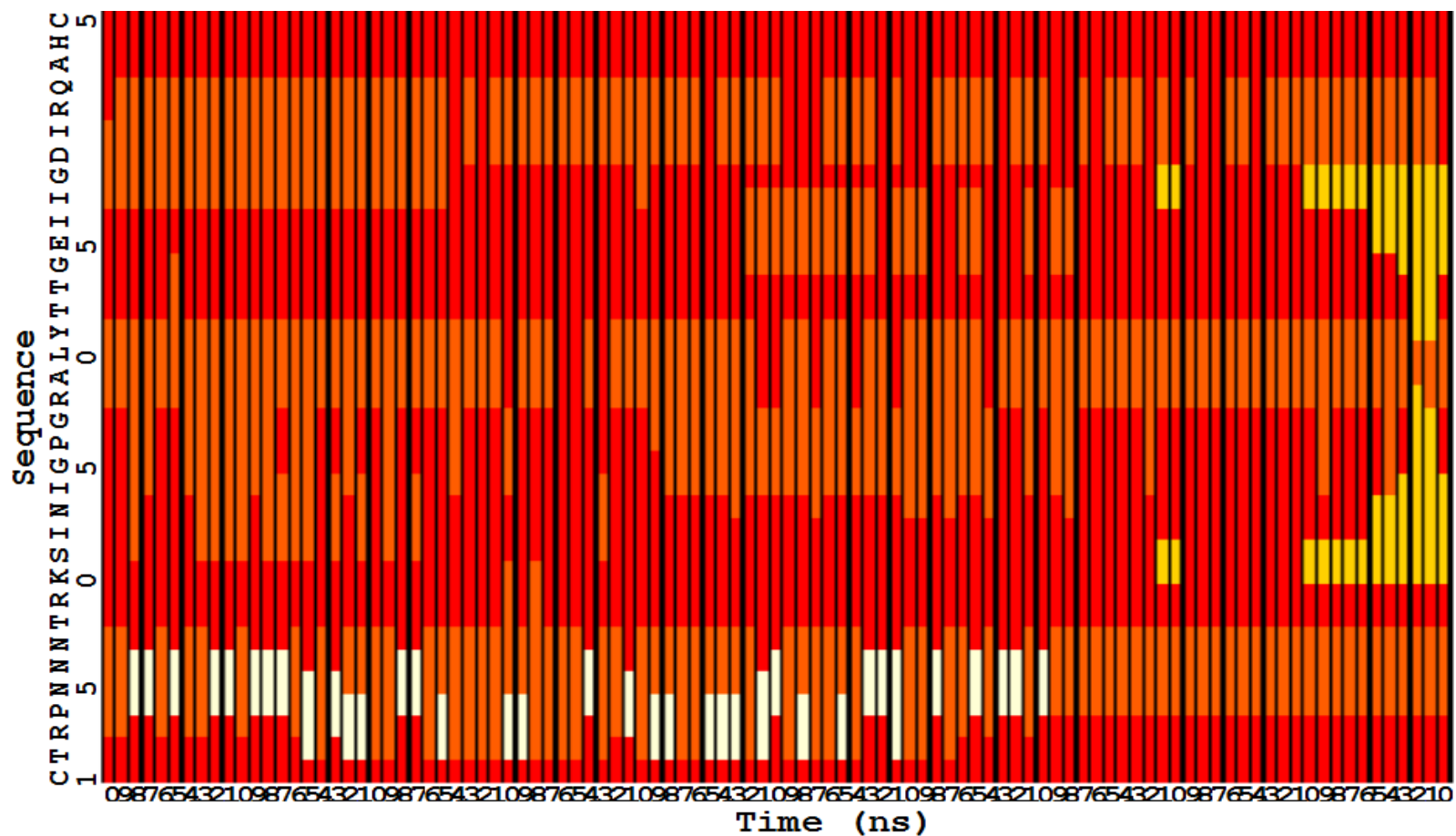


Figure 2.21: Contact plot for structural template 2B4C. Interactions between residues calculated at 4 Å and average through the trajectory. The colored code is shown at the right, with 0% occupancy shown as white and 100% occupancy shown as dark purple. Axes denote the residue number in sequence. The purple box represents the residues interacting at the turn while the pink box represents the residues that formed the transient helices.

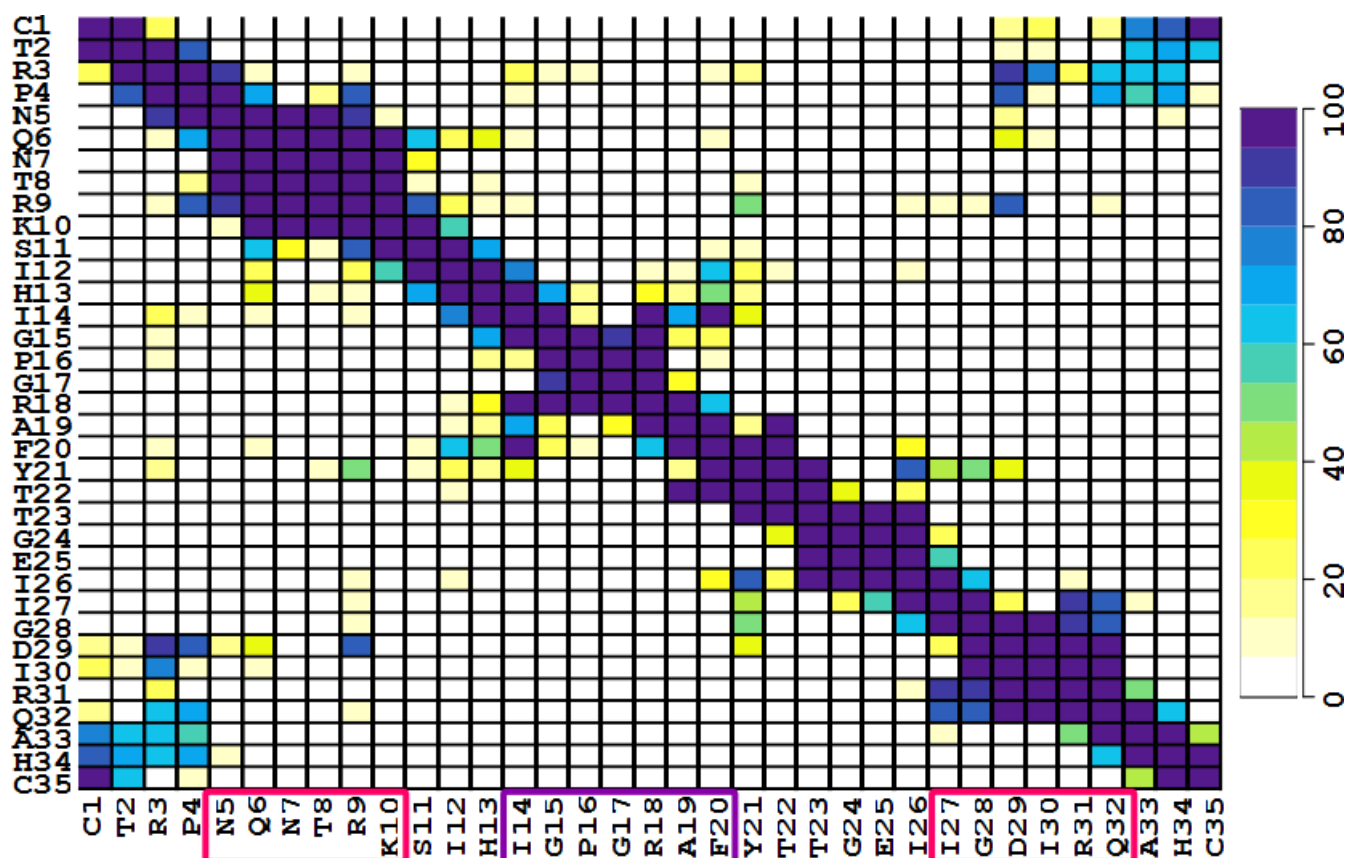


Figure 2.22: Contact plot for structural template 2QAD. Interactions between residues calculated at 4 Å and average through the trajectory. The colored code is shown at the right, with 0% occupancy shown as white and 100% occupancy shown as dark purple. Axes denote the residue number in sequence. The purple box represents the residues interacting at the turn while the pink box represents the residues that formed the transient helices. The orange box represents the residues that formed the beta strands.

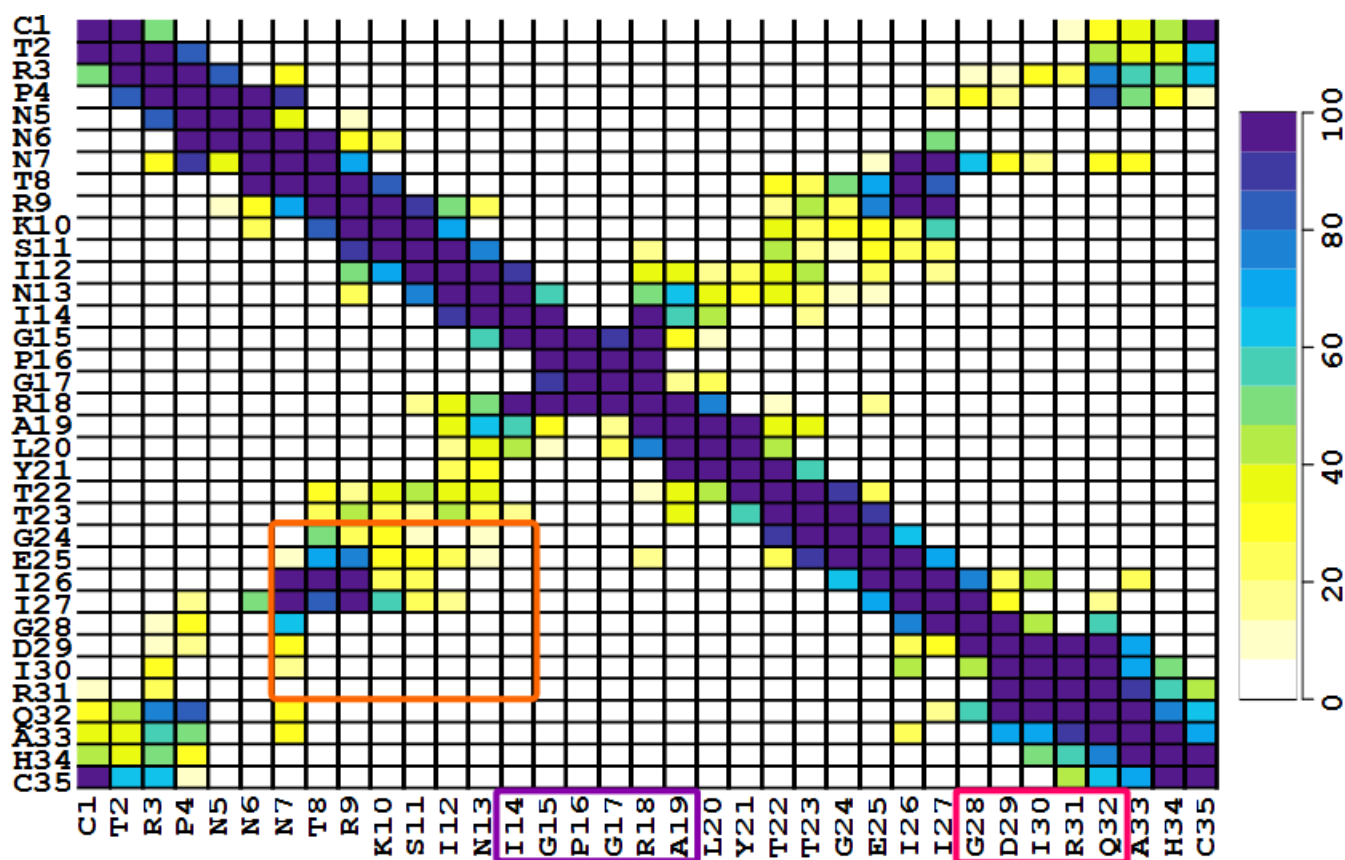


Figure 2.23: Electrostatic clustering of snapshots of the MD trajectories of the V3 loop with ionic strength corresponding to 0 mM concentration. The graph was generated using multidimensional scaling as explained in section 2.2 Methods. The axes denote coordinates that reflect the dissimilarities between snapshots. The snapshots are labeled as numbers corresponding to time in ns, with the red labels corresponding to the trajectory of 2B4C as starting structure and the green labels corresponding to the trajectory of 2QAD as starting structure.

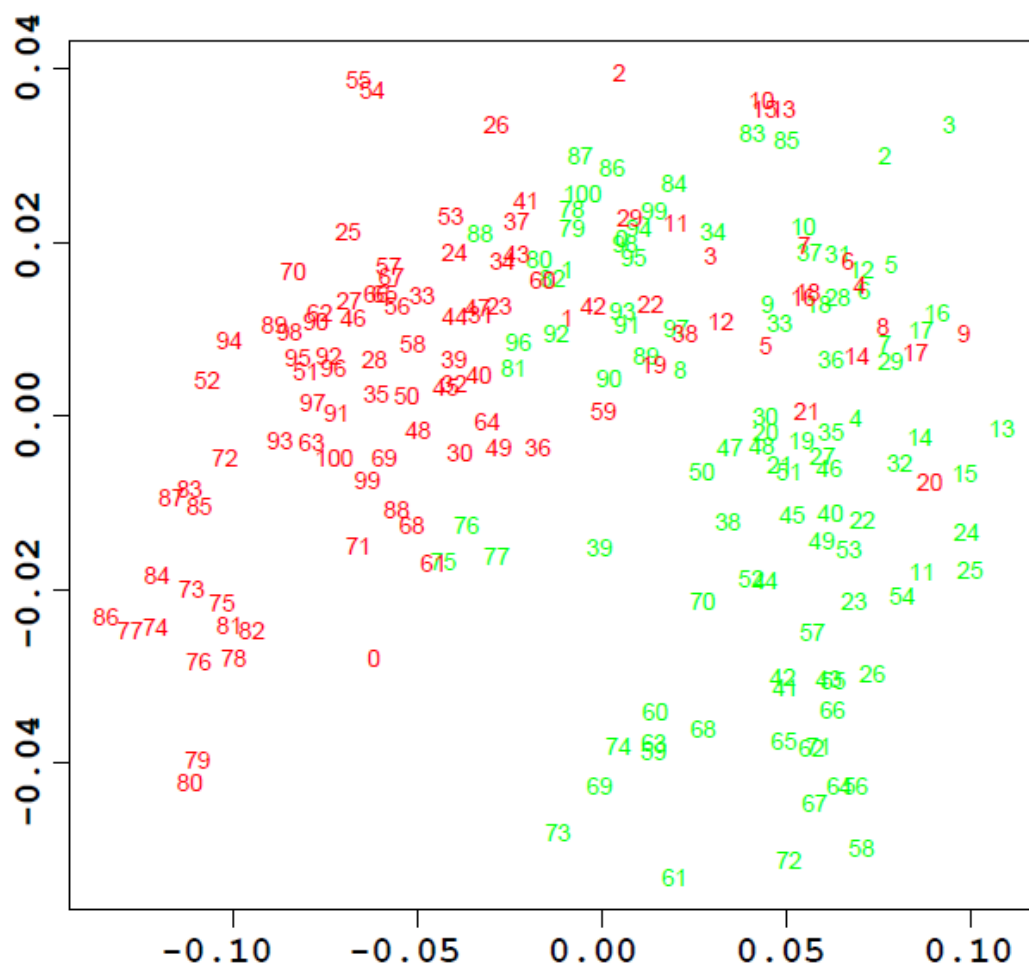


Figure 2.24: Electrostatic clustering of snapshots of the MD trajectories of the V3 loop with ionic strength corresponding to 150 mM concentration. The graph was generated using multidimensional scaling as explained in section 2.2 Methods. The axes denote coordinates that reflect the dissimilarities between snapshots. The snapshots are labeled as numbers corresponding to time in ns, with the red labels corresponding to the trajectory of 2B4C as starting structure and the green labels corresponding to the trajectory of 2QAD as starting structure.

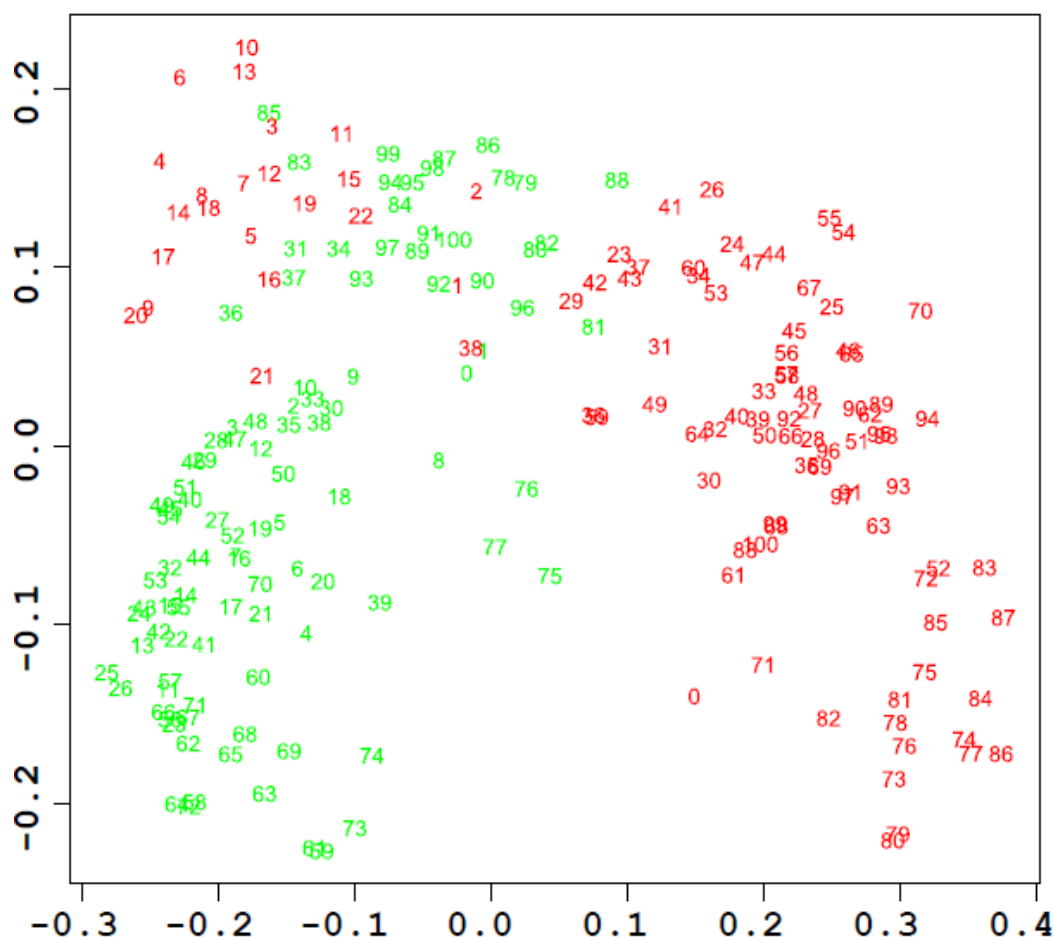


Table 2.1: Alignment of HIV-1 V3 loop consensus sequences. V3 loop subtype consensus sequences were obtained using the tools available within the Los Alamos National Laboratory Database (<http://www.hiv.lanl.gov>), except for sequences 2B4C and 2QAD, which are from the crystal structures deposited at the PDB. Some subtype consensus sequences were identical: CPX with H, and B with K. Subtypes AB, AE, AG, and CPX are circulation recombinant forms (CRFs). The construction of D35 from D is described in Methods. Gaps are introduced by the alignment.

Subtype	Sequence																								Charge										
	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4		5									
2B4C	C	T	R	P	N	Q	N	T	R	K	S	I	H	I	G	P	G	R	A	F	Y	T	T	G	-	3									
2QAD	C	T	R	P	N	N	N	T	R	K	S	I	N	I	G	P	G	R	A	L	Y	T	T	G	-	3									
A	C	T	R	P	N	N	N	T	R	K	S	V	R	I	G	P	G	Q	A	F	Y	A	T	G	-	3									
AB	C	I	R	P	G	N	N	T	R	T	S	I	R	I	G	P	G	Q	T	F	Y	A	T	G	-	2									
AE	C	T	R	P	S	N	N	T	R	T	S	I	T	I	G	P	G	Q	V	F	Y	R	T	G	-	3									
AG	C	T	R	P	N	N	N	T	R	K	S	V	R	I	G	P	G	Q	T	F	Y	A	T	G	-	3									
B	C	T	R	P	N	N	N	T	R	K	S	I	H	I	G	P	G	R	A	F	Y	A	T	G	-	3									
C	C	T	R	P	N	N	N	T	R	K	S	I	R	I	G	P	G	Q	T	F	Y	A	T	G	-	3									
CPX	C	T	R	P	N	N	N	T	R	K	S	I	H	I	G	P	G	Q	A	F	Y	A	T	G	-	2									
D	C	T	R	P	Y	N	N	T	R	Q	S	T	H	I	G	P	G	Q	A	L	Y	T	T	-	-	2									
D35	C	T	R	P	Y	N	N	T	R	Q	S	T	H	I	G	P	G	Q	A	L	Y	T	T	K	-	3									
F	C	T	R	P	N	N	N	T	R	K	S	I	H	L	G	P	G	Q	A	F	Y	A	T	G	-	3									
G	C	T	R	P	N	N	N	T	R	K	S	I	R	I	G	P	G	Q	A	F	Y	A	T	G	-	3									
H	C	T	R	P	N	N	N	T	R	K	S	I	H	I	G	P	G	Q	A	F	Y	A	T	G	-	2									
J	C	I	R	P	A	N	N	T	R	K	G	I	H	I	G	P	G	Q	V	L	Y	A	T	G	-	2									
K	C	T	R	P	N	N	N	T	R	K	S	I	H	I	G	P	G	R	A	F	Y	A	T	G	-	3									
N	C	T	R	P	G	N	N	T	G	G	Q	V	Q	I	G	P	A	M	T	F	Y	N	I	E	-	1									
O	C	E	R	P	G	N	N	T	V	Q	E	I	K	I	G	P	-	M	A	W	Y	S	M	G	L	E	N	N	S	R	A	A	Y	C	-1

Table 2.2: Docking of V3 loop with CCR5-Nt at each subset. Potential energy is in kcal/mol. Lowest energy structure is highlighted green.

2B4C			2QAD		
<i>CCR5-Nt at the Base</i>			<i>CCR5-Nt at the Base</i>		
Structure #	Total Potential Energy	Electrostatic	Structure #	Total Potential Energy	Electrostatic
1	8.6	-462.7	1	-201.7	-405.2
2	-64.4	-508.1	2	-205.2	-405.8
3	-112.0	-548.3	3	-209.2	-409.7
4	-146.7	-582.2	4	-213.7	-410.4
5	-217.3	-596.7	5	-214.9	-408.3
6	-215.9	-612.6	6	-181.5	-409.9
7	-220.8	-617.1	7	-181.4	-410.2
8	-226.4	-612.8	8	-183.8	-411.1
9	-249.2	-643.9	9	-186.2	-411.4
10	-253.0	-645.3	10	-188.8	-413.6
<i>CCR5-Nt at the Stem</i>			<i>CCR5-Nt at the Stem</i>		
Structure #	Total Potential Energy	Electrostatic	Structure #	Total Potential Energy	Electrostatic
1	576.6	-262.8	1	276.1	-207.0
2	202.2	-303.8	2	272.6	-255.2
3	86.0	-386.9	3	249.0	-259.9
4	36.7	-454.7	4	226.1	-247.8
5	1.0	-465.7	5	211.9	-254.6
6	-24.9	-490.4	6	184.5	-271.0
7	-24.6	-480.2	7	177.7	-271.2
8	-26.0	-472.7	8	166.5	-274.1
9	-37.9	-481.5	9	154.9	-279.4
10	-51.7	-492.2	10	148.8	-283.9
<i>CCR5-Nt at the Tip</i>			<i>CCR5-Nt at the Tip</i>		
Structure #	Total Potential Energy	Electrostatic	Structure #	Total Potential Energy	Electrostatic
1	348.0	-29.3	1	256.6	-98.6
2	344.5	-27.3	2	186.4	-125.2
3	337.8	-27.2	3	172.1	-133.2
4	336.7	-25.5	4	170.3	-137.8
5	330.0	-24.3	5	158.8	-139.8
6	321.3	-29.7	6	144.0	-141.3
7	314.8	-28.3	7	127.1	-154.6
8	309.2	-35.8	8	123.6	-156.4
9	305.2	-41.9	9	114.0	-155.2
10	304.3	-40.2	10	110.5	-152.8

Table 2.3: Docking of 2B4C with CCR5-Nt rotations at each subset. Tabulated potential energy for the three structures with the lowest values. The last structure of the first rotations was used as the starting point for the second rotations. Potential energy is in kcal/mol.

2B4C:CCR5-Nt at the Base					
<i>First Rotation</i>			<i>Second Rotation</i>		
Structure #	Total Potential Energy	Electrostatic	Structure #	Total Potential Energy	Electrostatic
<i>90° x-axis</i>			<i>90° x-axis</i>		
6	-247.6	-645.9	6	-219.4	-641.8
9	-245.0	-630.3	8	-221.1	-646.1
10	-245.4	-640.8	9	-221.4	-643.6
<i>90° y-axis</i>			<i>90° y-axis</i>		
8	-265.3	-673.7	7	-266.5	-711.2
9	-271.8	-682.1	9	-269.9	-715.8
10	-281.1	-687.1	10	-273.1	-715.5
<i>90° z-axis</i>			<i>90° z-axis</i>		
8	-242.4	-626.7	8	-237.1	-674.6
9	-246.4	-627.6	9	-239.6	-676.3
10	-246.8	-625.0	10	-243.0	-678.8
2B4C:CCR5-Nt at the Stem					
<i>First Rotation</i>			<i>Second Rotation</i>		
Structure #	Total Potential Energy	Electrostatic	Structure #	Total Potential Energy	Electrostatic
<i>90° x-axis</i>			<i>90° x-axis</i>		
8	-147.2	-538.8	7	-163.8	-553.8
9	-152.5	-534.6	8	-163.5	-551.3
10	-153.0	-535.4	9	-163.9	-552.7
<i>90° y-axis</i>			<i>90° y-axis</i>		
8	-107.4	-516.1	8	-212.3	-576.3
9	-118.4	-527.9	9	-210.8	-577.4
10	-140.9	-556.6	10	-214.4	-582.7
<i>90° z-axis</i>			<i>90° z-axis</i>		
8	-76.5	-464.4	8	-184.7	-581.3
9	-82.4	-466.1	9	-211.1	-600.0
10	-89.6	-471.0	10	-216.6	-600.1

2B4C:CCR5-Nt at the Tip

<i>First Rotation</i>			<i>Second Rotation</i>		
Structure #	Total Potential Energy	Electrostatic	Structure #	Total Potential Energy	Electrostatic
<i>90° x-axis</i>			<i>90° x-axis</i>		
7	218.8	-129.6	8	223.5	-121.9
9	217.6	-129.5	9	217.1	-129.4
10	218.4	-128.7	10	200.3	-143.9
<i>90° y-axis</i>			<i>90° y-axis</i>		
8	219.5	-127.4	8	218.5	-125.7
9	216.5	-129.8	9	217.0	-124.9
10	214.5	-130.6	10	215.0	-123.4
<i>90° z-axis</i>			<i>90° z-axis</i>		
8	306.5	-30.2	8	297.8	-28.0
9	305.9	-39.0	9	295.3	-24.8
10	301.9	-40.5S	10	297.0	-30.3

Table2.4: Docking of 2QAD with CCR5-Nt rotations at each subset. Tabulated potential energy for the three structures with the lowest values. The last structure of the first rotations was used as the starting point for the second rotations. Potential energy is in kcal/mol.

2QAD:CCR5-Nt at the Base					
<i>First Rotation</i>			<i>Second Rotation</i>		
Structure #	Total Potential Energy	Electrostatic	Structure #	Total Potential Energy	Electrostatic
<i>90° x-axis</i>			<i>90° x-axis</i>		
7	-226.6	-455.7	3	-418.1	-462.9
8	-232.1	-459.1	4	-420.3	-462.7
9	-232.4	-457.9	5	-423.9	-461.9
<i>90° y-axis</i>			<i>90° y-axis</i>		
7	-219.0	-436.1	3	-251.7	-438.5
9	-218.8	-432.6	4	-289.9	-461.4
10	-227.7	-444.9	5	-306.2	-471.4
<i>90° z-axis</i>			<i>90° z-axis</i>		
7	-190.9	-438.6	7	-103.3	-615.4
8	-195.3	-439.7	8	-103.3	-613.4
10	-191.2	-447.1	9	-107.0	-617.6
2QAD:CCR5-Nt at the Stem					
<i>First Rotation</i>			<i>Second Rotation</i>		
Structure #	Total Potential Energy	Electrostatic	Structure #	Total Potential Energy	Electrostatic
<i>90° x-axis</i>			<i>90° x-axis</i>		
7	73.8	-332.3	7	59.9	-339.9
9	76.2	-346.0	8	60.8	-340.0
10	66.1	-333.4	10	50.3	-354.4
<i>90° y-axis</i>			<i>90° y-axis</i>		
7	60.4	-337.7	8	110.5	-346.3
9	57.7	-337.5	9	103.0	-359.2
10	55.2	-339.9	10	100.3	-358.6
<i>90° z-axis</i>			<i>90° z-axis</i>		
7	146.2	-338.9	8	137.0	-342.7
8	146.9	-340.0	9	121.1	-360.2
10	146.6	-356.8	10	118.6	-351.5

2QAD:CCR5-Nt at the Tip

<i>First Rotation</i>			<i>Second Rotation</i>		
Structure #	Total Potential Energy	Electrostatic	Structure #	Total Potential Energy	Electrostatic
<i>90° x-axis</i>			<i>90° x-axis</i>		
8	118.4	-133.9	6	74.1	-138.9
9	110.8	-150.4	7	73.3	-138.0
10	92.5	-149.0	9	73.2	-140.2
<i>90° y-axis</i>			<i>90° y-axis</i>		
6	144.8	-147.2	7	137.9	-187.1
7	142.8	-149.2	8	138.5	-186.3
9	141.2	-144.6	10	138.4	-183.8
<i>90° z-axis</i>			<i>90° z-axis</i>		
8	114.3	-142.8	3	147.2	-121.4
9	107.2	-143.2	4	141.3	-123.1
10	108.3	-145.1	5	135.5	-129.6

Table 2.5: Occupancy of prevalent hydrogen bonds during the MD trajectories, involving side chains and/or backbone, for both structural templates used. Bold font highlight bonds common in both structures.

Structure: 2B4C							Structure: 2QAD						
Residue #1			Residue #2			%	Residue #1			Residue #2			%
ARG	9	NH2	ASP	29	OD1	68.1 ^a	ARG	31	NH2	ASP	29	OD2	74.6 ^a
ARG	9	NH2	ASP	29	OD2	65.2 ^a	ARG	31	NH2	ASP	29	OD1	73.6 ^a
HIS	34	N	THR	2	O	58.6	ILE	27	N	ASN	7	O	70.2
THR	2	N	HIS	34	O	57.3	ARG	9	N	GLU	25	O	68.5
ASN	7	N	ASN	5	OD1	52.4	ARG	3	NH1	CYS	35	OXT	48.4
ARG	9	NH1	ASP	29	OD2	50.1	ARG	3	NE	CYS	35	O	47.5
SER	11	N	GLN	6	O	47.1	ARG	31	NE	ASP	29	OD1	45.2
ARG	9	NH1	ASP	29	OD1	44.1	ARG	31	NE	ASP	29	OD2	42.3
ARG	9	N	ASN	5	O	41.8	ALA	33	N	ASP	29	O	41.0
LYS	10	N	GLN	6	O	40.4	ARG	31	N	ASP	29	OD1	39.3
ARG	31	N	GLY	28	O	39.8	ARG	31	N	ASP	29	OD2	37.0
GLN	32	N	GLY	28	O	35.4	CYS	35	N	GLN	32	O	34.8
GLN	32	N	ASP	29	O	32.3	THR	2	N	HIS	34	O	32.8
LYS	10	N	ASN	7	O	32.1	ARG	18	N	GLY	15	O	32.7^b
THR	8	N	ASN	5	OD1	27.8	GLN	32	N	ASP	29	O	32.5
ARG	9	NE	SER	11	OG	26.0	ARG	3	NH1	CYS	35	O	32.3
GLN	32	NE2	GLY	28	O	25.6	ARG	9	NH1	GLU	25	OE1	31.2 ^a
ARG	18	N	GLY	15	O	23.8^b	HIS	34	N	THR	2	O	30.6
ARG	3	NH1	GLN	32	O	22.4	GLY	24	N	THR	22	OG1	30.4
ASN	7	ND2	ASN	5	OD1	21.8	ARG	9	NE	SER	11	OG	28.3
GLU	25	N	THR	23	OG1	21.4	ARG	9	NH1	GLU	25	OE2	27.1 ^a
							HIS	34	N	ILE	30	O	25.8
							GLY	28	N	ASN	7	O	23.2
							ILE	27	N	ARG	9	O	22.8
							THR	8	N	ASN	6	OD1	22.7
							GLY	15	N	ARG	18	O	21.6
							SER	11	N	GLU	25	O	21.4

^a Salt bridges (see Table 2.6).

^b Hydrogen bond present in the turn region (at the tip off each structure).

Table 2.6: Occupancy of salt bridges during the MD trajectories, for each structural template used.

Structure: 2B4C							Structure: 2QAD						
Residue #1			Residue #2			%	Residue #1			Residue #2			%
ARG	9	CZ	ASP	29	CG	85.6	ARG	31	CZ	ASP	29	CG	91.0
ARG	3	CZ	ASP	29	CG	22.6	ARG	9	CZ	GLU	25	CD	57.5

2.5 References

1. Chan DC, Fass D, Berger JM, Kim PS. Core structure of gp41 from the HIV envelope glycoprotein. *Cell* 1997, **89**: 263-273.
2. Dittmar MT, McKnight A, Simmons G, Clapham PR, Weiss RA, Simmonds P. HIV-1 tropism and co-receptor use. *Nature* 1997, **385**: 495-496.
3. Xiao LH, Owen SM, Goldman I, Lal AA, DeJong JJ, Goudsmit J, Lal RB. CCR5 coreceptor usage of non-syncytium-inducing primary HIV-1 is independent of phylogenetically distinct global HIV-1 isolates: Delineation of consensus motif in the V3 domain that predicts CCR-5 usage. *Virology* 1998, **240**: 83-92.
4. Pollakis G, Kang S, Kliphuis A, Chalaby MI, Goudsmit J, Paxton WA. N-linked glycosylation of the HIV type-1 gp120 envelope glycoprotein as a major determinant of CCR5 and CXCR4 coreceptor utilization. *J Biol Chem* 2001, **276**: 13433-13441.
5. Cardozo T, Kimura T, Philpott S, Weiser B, Burger H, Zolla-Pazner S. Structural basis for coreceptor selectivity by the HIV type 1 V3 loop. *AIDS Res Hum Retrov* 2007, **23**: 415-426.
6. Edo-Matas D, van Dort KA, Setiawan LC, Schuitemaker H, Kootstra NA. Comparison of in vivo and in vitro evolution of CCR5 to CXCR4 coreceptor use of primary human immunodeficiency virus type 1 variants. *Virology* 2011, **412**: 269-277.
7. Tscherning C, Alaeus A, Fredriksson R, Bjorndal A, Deng HK, Littman DR, Fenyo EM, Albert J. Differences in chemokine coreceptor usage between genetic subtypes of HIV-1. *Virology* 1998, **241**: 181-188.
8. Hemelaar J, Gouws E, Ghys PD, Osmanov S. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS* 2006, **20**: W13-23.
9. Spira S, Wainberg MA, Loemba H, Turner D, Brenner BG. Impact of clade diversity on HIV-1 virulence, antiretroviral drug sensitivity and drug resistance. *J Antimicrob Chemother* 2003, **51**: 229-240.
10. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* 1998, **393**: 648-659.

11. Huang CC, Tang M, Zhang MY, Majeed S, Montabana E, Stanfield RL, Dimitrov DS, Korber B, Sodroski J, Wilson IA, *et al.* Structure of a V3-containing HIV-1 gp120 core. *Science* 2005, **310**: 1025-1028.
12. Huang CC, Lam SN, Acharya P, Tang M, Xiang SH, Hussan SS, Stanfield RL, Robinson J, Sodroski J, Wilson IA, *et al.* Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4. *Science* 2007, **317**: 1930-1934.
13. Napier KB, Wang ZX, Peiper SC, Trent JO. CCR5 interactions with the variable 3 loop of gp120. *J Mol Model* 2007, **13**: 29-41.
14. Liu S, Fan S, Sun Z. Structural and functional characterization of the human CCR5 receptor in complex with HIV gp120 envelope glycoprotein and CD4 receptor by molecular modeling studies. *J Mol Model* 2003, **9**: 329-336.
15. Brelot A, Heveker N, Montes M, Alizon M. Identification of residues of CXCR4 critical for human immunodeficiency virus coreceptor and chemokine receptor activities. *J Biol Chem* 2000, **275**: 23736-23744.
16. Wu B, Chien EYT, Mol CD, Fenalti G, Liu W, Katritch V, Abagyan R, Brooun A, Wells P, Bi FC, *et al.* Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonist. *Science* 2010, **330**: 1066-1071.
17. Cornier EG, Persuh M, Thompson DAD, Lin SW, Sakmar TP, Olson WC, Dragic T. Specific interaction of CCR5 amino-terminal domain peptides containing sulfotyrosines with HIV-1 envelope glycoprotein gp120. *Proc Natl Acad Sci USA* 2000, **97**: 5762-5767.
18. Krambovitis E, Zafiropoulos A, Baritaki S, Spandidos DA. Simple electrostatic interaction mechanism in the service of HIV-1 pathogenesis. *Scand J Immunol* 2004, **59**: 231-234.
19. Morikis D, Rizos AK, Spandidos DA, Krambovitis E. Electrostatic modeling of peptides derived from the V3-loop of HIV-1 gp120: implications of the interaction with chemokine receptor CCR5. *Int J Mol Med* 2007, **19**: 343-351.
20. Baritaki S, Zafiropoulos A, Sioumpara M, Politis M, Spandidos DA, Krambovitis E. Ionic interaction of the HIV-1 V3 domain with CCR5 and deregulation of T lymphocyte function. *Biochem Biophys Res Commun* 2002, **298**: 574-580.
21. Wu J, Morikis D. Molecular thermodynamics for charged biomacromolecules. *Fluid Phase Equilib* 2006, **241**: 317-333.

22. Farzan M, Mirzabekov T, Kolchinsky P, Wyatt R, Cayabyab M, Gerard NP, Gerard C, Sodroski J, Choe H. Tyrosine sulfation of the amino terminus of CCR5 facilitates HIV-1 entry. *Cell* 1999, **96**: 667-676.
23. Farzan M, Babcock GJ, Vasilieva N, Wright PL, Kiprilov E, Mirzabekov T, Choe H. The role of post-translational modifications of the CXCR4 amino terminus in stromal-derived factor 1 alpha association and HIV-1 entry. *J Biol Chem* 2002, **277**: 29484-29489.
24. Berman HM. The Protein Data Bank. *Nucleic Acids Res* 2000, **28**: 235-242.
25. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PDB Viewer: An environment for comparative protein modeling. *Electrophoresis* 1997, **18**: 2714-2723.
26. Insight II Modeling Environment. June 2000 ed. San Diego, CA, USA: Molecular Simulations Inc.
27. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* 2007, **23**: 2947-2948.
28. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci* 2007, **Chapter 2**: 2.9.9-2.9.30.
29. Carbo R, Leyda L, Arnau M. How similar is a molecule to another? An electron-density measure of similarity between 2 molecular-structures. *Int J Quantum Chem* 1980, **17**: 1185-1189.
30. Hodgkin EE, Richards WG. Molecular similarity based on electrostatic potential and electric-field. *Int J Quantum Chem* 1987, **32**: 105-110.
31. Reynolds CA, Burt C, Richards WG. A linear molecular similarity index *Quant Struct-Act Rel* 1992, **11**: 34-35.
32. Petke JD. Cumulative and discrete similarity analysis of electrostatic potentials and fields. *J Comput Chem* 1993, **14**: 928-933.
33. Kieslich CA, Yang J, Gunopulos D, Morikis D. Automated computational protocol for alanine scans and clustering of electrostatic potentials: application to C3d–CR2 association. *Biotechnol Prog* 2010, **27**: 316-325.

34. Kieslich CA, Gorham RD, Jr., Morikis D. Is the rigid-body assumption reasonable? Insights into the effects of dynamics on the electrostatic analysis of barnase-barstar. *J Non-Cryst Solids* 2010, **357**: 707-716.
35. Hakkoymaz H, Kieslich CA, Gorham RD, Jr., Gunopulos D, Morikis D. Electrostatic similarity determination using multi-resolution analysis. *Mol Inf* 2011, **30**: 733-746.
36. Gorham RD, Jr., Kieslich CA, Nichols A, Sausman NU, Foronda M, Morikis D. An evaluation of Poisson-Boltzmann electrostatic free energy calculations through comparison with experimental mutagenesis. *Biopolymers* 2011, **95**: 746-754.
37. Gorham RD, Jr., Kieslich CA, Morikis D. Complement inhibition by *Staphylococcus aureus*: electrostatics of C3d-EfbC and C3d-Ehp association. *Cell Molec Bioeng* 2011, **5**: 32-43.
38. El-Assaad AM, Kieslich CA, Gorham RD, Jr., Morikis D. Electrostatic exploration of the C3d-FH4 interaction using a computational alanine scan. *Mol Immunol* 2011, **48**: 1844-1850.
39. Cheung AS, Kieslich CA, Yang J, Morikis D. Solvation effects in calculated electrostatic association free energies for the C3d-CR2 complex and comparison with experimental data. *Biopolymers* 2010, **93**: 509-519.
40. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res* 2007, **35**: W522-525.
41. Sitkoff D, Sharp KA, Honig B. Accurate calculation of hydration free energies using macroscopic solvent models. *J Phys Chem* 1994, **98**: 1978-1988.
42. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 2001, **98**: 10037-10041.
43. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF chimera - A visualization system for exploratory research and analysis. *J Comput Chem* 2004, **25**: 1605-1612.
44. MacKerell AD, Jr., Bashford D, Bellott M, Dunbrack RL, Jr., Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, *et al.* An all-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 1998, **102**: 3586-3616.

45. Mackerell AD, Jr., Feig M, Brooks CL, III Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* 2003, **25**: 1400-1415.
46. Macias AT, MacKerell AD, Jr. CH/ π interactions involving aromatic amino acids: refinement of the CHARMM tryptophan force field. *J Comput Chem* 2005, **26**: 1452-1463.
47. Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K. NAMD2: Greater scalability for parallel molecular dynamics. *J Comp Phys* 1999, **151**: 283-312.
48. Grant BJ, Rodrigues AP, ElSawy KM, McCammon JA, Caves LS. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* 2006, **22**: 2695-2696.
49. Krambovitis E, Spandidos DA. HIV-1 infection: is it time to reconsider our concepts? *Int J Mol Med* 2006, **18**: 3-8.
50. Huang W, Eshleman SH, Toma J, Fransen S, Stawiski E, Paxinos EE, Whitcomb JM, Young AM, Donnell D, Mmiro F, *et al.* Coreceptor tropism in human immunodeficiency virus type 1 subtype D: high prevalence of CXCR4 tropism and heterogeneous composition of viral populations. *J Virol* 2007, **81**: 7885-7893.
51. Harrigan PR, Geretti AM. Genotypic tropism testing: evidence-based or leap of faith? *AIDS* 2011, **25**: 257-264.
52. Masso M, Vaisman II. Accurate and efficient gp120 V3 loop structure based models for the determination of HIV-1 co-receptor usage. *BMC Bioinformatics* 2010, **11**: 494.
53. West R, Hawkins E, Jackson P, editors. Preparation of modified cyanine dyes for the electrophoretic analysis of glycoproteins. *Sixth International Conference on Methods and Application of Fluorescence Spectroscopy, Imaging and Probes* 1999; Paris, France.
54. Kieslich CA, Shin D, López de Victoria A, González-Rivera G, Morikis D. A Predictive Model for HIV-1 co-receptor selectivity and disease progression. *Submitted* 2012.
55. Frishman D, Argos P. Knowledge-based protein secondary structure assignment. *Proteins* 1995, **23**: 566-579.
56. Rosen O, Sharon M, Quadat-Akabayov SR, Anglist J. Molecular switch for alternative conformations of the HIV-1 V3 region: implications for phenotype conversion. *Proc Natl Acad Sci USA* 2006, **103**: 13950-13955.

57. López de Victoria A, Tamamis P, Kieslich CA, Morikis D. Insights into the structure and correlated motions of two HIV-1 gp120 V3 loops using molecular dynamics simulations. *Submitted* 2012.

CHAPTER 3

Complement System Inhibition by Compstatin Family Peptides

3.1 Background

The complement system is part of the human innate immunity that constantly scrutinizes and clears the body of foreign pathogens, and also invokes an inflammation response to tissue injuries. Complement components (ie, proteins) are found circulating in the serum as inactive precursors waiting for an activation signal. The activation of the complement system can occur through three pathways: (i) the classical pathway, (ii) the lectin pathway or, (iii) the alternative pathway. All pathways converge in the cleavage of complement component 3 (C3) to C3a and C3b, after which the complement cascade continues until the formation of the membrane attack complex (MAC) resulting in cell lysis. Figure 3.1 shows a simple model of the complement system. The complement system is tightly regulated by a series of proteins and reactions, to minimize the possibility of attacking host cells. However, under or over regulation of the complement system can cause damage to host cells leading to autoimmune or inflammatory conditions, such as age-related macular degeneration, asthma, hemolytic anemia and transplant rejection, among others (1-4). Thus therapeutic inhibition or activation control of the complement system can be beneficial to treat several diseases. Since C3 is

an essential component and a common protein between all pathways, it is an excellent candidate for drug development.

The compstatin family consists of peptides that bind to C3 and inhibit the activation of the complement system by hindering the cleavage of C3 to C3a and C3b (5). Compstatin was first discovered using a phage-displayed random peptide library for binding against C3b (6) and optimized over several years using knowledge-based design and experimental and computational combinatorial methods (reviewed in (5-11)). The design/optimization process involves the understanding of physicochemical properties that underlie its structure and inhibitory activity, and the optimization of specific amino acids that result in enhanced inhibitory activity.

The compstatin family peptides consist of analogs with 13 amino acids. Eleven of the amino acids form a cyclic chain through a disulfide bridge and the other two amino acids are located outside the cyclic peptidic ring. The parent compstatin sequence is I[CVVQDWGHHRC]T, where brackets denote cyclization between Cys2 and Cys12, and is shown in Figure 3.2A. Initial characterization of compstatin showed that the disulfide bond between Cys2 and Cys12 is essential for inhibitory activity (6). The determination of the three-dimensional solution structure of compstatin by nuclear magnetic resonance (NMR), in combination with alanine scans, established the first sequence-structure-activity relations (12). The NMR structure of parent compstatin reveals a type I β -turn involving residues 5-8 (12). Alanine substitutions of residues Val3, Gln5, Asp6, Gly8, and Trp7 resulted in lower activity, with four of those residues involved in the β -

turn, thus demonstrating their significance in the peptide activity and structure stability (12). On the other hand, alanine substitutions of residues Val4, His9, His10, and Arg11 resulted in minimal change to the peptide inhibitory activity (12), suggesting that those positions could be optimized. A hydrophobic cluster involving residues Val3, Val4, and Trp7 was also observed. NMR and inhibitory activity studies of compstatin analogs with rationally-designed mutations, aimed at introducing local perturbations affecting structure and activity, were responsible for identifying the physicochemical properties that were important for structural stability and inhibitory activity (7, 13). Those studies revealed a hydrophobic clusters at the termini involving residues 1-4 and 12-13, because those residues are close in proximity due to the cyclic nature of the peptide (13). In combination, these studies proposed the sequence template X[CVXQDWGXXXC]X, which distinguishes 7 amino acids indispensable for activity and 6 amino acids that were amenable to further optimization (noted by X), with the former maintaining their dominant parent physicochemical properties (7, 13). Acetylation of the N-terminal of compstatin protects Ile1 cleavage by proteases and causes a 3-fold increase in activity (14).

Application of a global optimization computational method, using the solution structure of compstatin and the aforementioned sequence template, was responsible for pinpointing the need for an aromatic amino acid at position 4, such as Tyr and Trp, in order to optimized the activity of parent compstatin (15). This substitution led to the sequence of the W4/A9 analog, I[CVWQDWGAHRC]T, the most active sequence with

natural amino acids (Figure 3.2B). The W4/A9 analog opened the way for the inclusion of non-natural amino acids at position 4, which further increased the activity of compstatin family peptides (16-18). A study incorporated Trp analogs at position 4, resulting in the sequence of MeW4/A9 analog, I[CV(MeW)QDWGAHRC]T (where MeW denotes 1-methyl tryptophan hereafter (18)). This study also suggests that reducing the hydrophobic character of the residue at position 4 and at position 7 are unfavorable for activity (18).

Other major breakthroughs involve the first study of compstatin dynamics using molecular dynamics (MD) simulations (19), a pharmacophore study of compstatin family peptides using quasi-dynamic pharmacophore models (20), the crystallographic structure of the complex between C3c and the most potent compstatin analog comprised of natural amino acids (W4/A9, Figure 3.2C, (21)). The crystal structure of C3c:W4/A9 complex revealed the compstatin binding site and that the β -turn is extended into residues 8-11 (compared to parent compstatin). It also showed an interaction rich complex, involving hydrogen bonds and hydrophobic interactions. In summary, Ile1 and Cys2 participate in hydrogen bonds with Asn390; Val3 is in a hydrophobic pocket created by Met346, Pro347, and Leu454; Trp4 participates in hydrogen bond with Gly345 and Arg456; Gln5 participates in hydrogen bonds with Leu455 and Met457; Trp7 side chain is found between residues 455–458 and 488–491, and involved in a hydrogen bond with Met457 while being closed to Gln5; Ala9 and His10 participate in hydrogen bonds with Asp491, while His10 is also in hydrophobic

interactions with Leu454 and Leu492 (21). All interactions are contained within four sections of C3c involving residues 344-349, 388-393, 454-462, and 488-492 (22-24). Based on the binding observed in the crystal structure, Janssen *et al* (21) proposed that compstatin sterically hinders the access of the C3 convertase to C3, thus blocking further complement activation.

An atomistic MD simulations of W4/A9 complexes with human and rat C3 was performed to understand the species-specificity binding of compstatin (23). The results revealed that rat C3c undergoes rearrangement, thus disrupting several polar and non-polar interactions (partial or total loss) with compstatin, and reducing the stability of the complex. As a result of those studies, a transgenic mouse C3 was designed with human features at the binding site (24) aiming in maintaining the necessary interactions for binding and stabilization of the complex.

A study employing *de novo* design computational frameworks to design new compstatin analogs with predicted binding abilities for C3c and inhibitory activities against complement activation revealed new key positions in the sequence of compstatin (25). This study identified the presence of Trp at position 13 and a variety of combinations of polar amino acids at positions 9-11. Based on these new key features, three new analogs were designed using the W4/A9 sequence as the template, with Trp at position 13 instead of Thr. Since the two end amino acids are hanging outside the cyclization ring (positions 1 and 13) and thus have a backbone which is conformationally

less restrained than the rest of the peptide, position 1 and/or 13 were also mutated to Trp. By inserting another Trp at position 10, we could symmetrically arrange (in sequence) Trp amino acids at positions (i, i + 3) (Peptide XA). This arrangement could probably enhance the possibility of Trp ring stacking, which could stabilize the peptide structure and perhaps binding to C3. Three top ranking sequences from the computational framework and the four Trp analogs were analyzed using surface plasmon resonance (SPR) to test for binding and determine kinetics, and enzyme-linked immunosorbent assay (ELISA) to determine the concentration at which half of the C3 present is inhibited by the peptide (26). Collectively, these analogs are hereafter called Generation A (Table 3.1). The insertion of Trp after the cyclization (positions 1 and/or 13) introduces the ability for new interactions with C3 (aromatic, π -cation, π -stacking, hydrogen bonds). To corroborate this possibility, MD simulations were performed using the Trp analogs of Generation A and human C3 (22). The MD simulations used the available crystal structure of human C3c:W4/A9 complex as the template, with appropriate mutations to W4/A9, in explicit solvent. The results showed that the Trp1 side chain interacts with His10 (analog) and Leu492 (C3c), while Trp13 resulted in a π -cation interaction with Arg11 (analog).

The MD study by Tamamis *et al* (22) introduced analogs with different modifications: (i) Arg at position 1, inspired by *de novo* design results, introducing the ability of new polar interactions, and in addition improving solubility; and (ii) di-Serine extensions at the N-terminal, inspired by the results of the species-specificity

simulations (23) and the original phage-displayed studies (6), introducing hydrogen bond capabilities. The results for the Arg1 analog showed that it maintained the interactions observed in the human C3c:W4/A9 complex, and a new salt bridge was observed with Asp349 in C3c. The analog with the di-Serine extension also maintained the interaction observed in the human C3c:W4/A9 complex and a new hydrogen bond was observed between Ser at position 0 with Asn390 (C3c).

Inspired by the species-specificity MD results and using the most potent analog with non-natural amino acid as a template (maintaining the hydrophobic character of position 4), several new analogs were designed. These new analogs incorporate Arg or MeW at position 1, and a combination of Arg, Ser and/or Trp at the N-terminal extension. The substitutions in these newly created analogs aim to maintain the inhibitory activity of the MeW4/A9 analog by keeping the hydrophobic characteristics of compstatin and improving solubility, by incorporating polar amino acids at the N-terminal extension. As a result, Generation B was born with a total of eight analogs, including five with the presence of MeW at position 4 (Table 3.2). These analogs were tested using ELISA to determine the concentration at which half of the C3 present is inhibited by the peptide (27).

A main aspect in the design of effective compstatin analogs is the balance between hydrophobicity and polarity. Enhanced hydrophobicity at key sequence positions (ie, positions 4 and 7 that participates in hydrophobic contacts with C3) can provide enhanced binding, whereas increasing the polarity at key sequence positions (ie,

position 1 or 13 that are outside the cyclic ring, positions 10-11 that have minimal effect on activity) can enhance the solubility of the free peptides.

Solubility is important for the storage and delivery of biopharmaceuticals administered in solution. The new analogs presented here have improved balance between hydrophobicity and polarity.

Herein, the compstatin analogs are referred to by their analog number and generation: Peptides IA-XA for the analogs in Generation A (Table 3.1) and Peptides IB-XIB for the analogs in Generation B (Table 3.2), irrespective of the presence of blocking groups at the termini or pegylation and biotinylation extensions at the C-terminus for the SPR experiments. The specifics of each sequence for the experimental studies are given in Tables 3.1-3.2, and 3.4, and are also discussed in section 3.2 Methods (below).

3.2 Methods

3.2.1 Peptide Synthesis

All compstatin analogs were synthesized by Abgent Inc. (San Diego, CA). The peptide sequences and mass spectral analyses of the experimentally tested compstatin analogs are shown in Table 3.1 for Generation A and Table 3.2 for Generation B.

Generation A consisted of six newly designed analogs and three control analogs. From the latter, two are positive controls, parent compstatin and W4/A9, and one is a negative control, linear compstatin (inactive). At the C-terminal, the analogs were

pegylated with an 8-PEG block spacer, followed by a Lys, and biotinylated (for SPR immobilization), followed by an NH₂ block. Three analogs were acetylated at the N-terminal (Table 3.1).

Generation B consisted of eight newly designed analogs and two positive control analogs. The positive controls are W4/A9 and MeW4/A9 (1-methylated tryptophan at position 4). All analogs were acetylated at the N-terminus and NH₂ blocked at the C-terminus (Table 3.2).

3.2.2 Surface Plasmon Resonance Assays

Binding of compstatin analogs to C3 was determined by Surface Plasmon Resonance (SPR) with Biacore X100 (GE healthcare, Piscataway, NJ) according to previous studies (16, 28, 29) with several modifications. The PEG spacer in the peptides increases mobility, solubility, and accessibility, and also decreases non-specific interactions. Streptavidin sensor chip SA was used to immobilize biotinylated compstatin peptides as the ligands. Phosphate buffer saline (PBS) containing 0.05% Tween-20 was used as the running buffer and the assay was carried out at 25°C.

Human C3 was purchased from Complement Technology Inc. (Tyler, TX). Approximately 300 µL of C3 were placed in a Microcon MWCO 100 kD (Millipore, Billerica, MA) and centrifuged at 10000 rcf for 15 minutes at 4°C. After discarding the flow-through, 100 µL of running buffer was added to the C3 and the solution was

centrifuged again under the same conditions. This process was repeated 3 times. C3 was recovered after the last run by adding 100 μ L of running buffer at a time, for 5 times. Concentration was determined using Beer-Lambert's Law with an extinction coefficient of $\epsilon^{1\%}$ at 280 nm of 10.3 (obtained from the Complement Technology Inc. report).

The compstatin analogs were immobilized onto the chip through their carboxy terminal, as previously suggested (16). The cyclic compstatin analogs were immobilized in the flow cell 2 (Fc2) and the linear analog (control) was immobilized in flow cell 1 (Fc1), both to a reading of 1000 response units (RU) with a flow rate of 10 μ L/min. The binding was measured at 30 μ L/min by adding several concentrations of human C3 (analyte) for 120 s and dissociation was monitored for 180 s, followed by regeneration with 10 mM NaOH for 60 s. Repeats of each experiment were performed for 180 s of association and 240 s of dissociation, followed by regenerations with 10 mM NaOH for 30 s and 0.05% SDS for 30 s. The solutions were placed in the Biacore sample rack in the following positions: (1) C3 0 nM, (2) C3 25 nM, (3) C3 50 nM, (4) C3 100 nM, (5) C3 200 nM, (6) C3 400 nM, (7) C3 800 nM, (8) C3 1600 nM, (9) running buffer, (10) 10 mM NaOH, (11) 0.05% SDS. A representation of the SPR method can be found in Figure 3.3.

The data obtained from Fc1 was subtracted from Fc2, resulting in a sensorgram. Figure 3.4A shows an example of the obtained data for Fc1 and Fc2 for a compstatin analog before the sensorgram was analyzed. The sensorgram was adjusted in the x-axis to start at the injection of the first sample, and in the y-axis to start at the baseline.

Figure 3.4B shows an adjusted sensorgram produced by the difference between Fc2 and Fc1. Kinetic parameters were obtained using the Biacore X100 Evaluation Software. Global fittings of the Biacore sensorgrams were performed using the 1:1 binding model ($A + B \rightarrow AB$) and the two-state reaction model ($A + B \rightarrow AB \rightarrow AB^*$), according to which 1:1 binding is followed by a conformational change that stabilizes the complex. Although data were collected at 0 nM-1600 nM C3 concentrations, best fits were produced using the two-state model and the data in the 25 nM-800 nM or the 0 nM-800 nM range. The quality of the fits was assessed by visual inspection of the fits and their residuals and using the $R_{\max}$ and χ^2 values.

3.2.3 Normal Human Serum Kinetics

Normal human serum (NHS) was obtained from Complement Technology Inc. and was centrifuged for 2 min at 12000 rpm prior to its use. In the course of ELISA inhibition experiments, we observed that assay parameters were dependent on reagents used. Specifically, the assays were sensitive to the lot of serum used. Prior to performing ELISA, the optimal NHS incubation time was determined for each lot used by performing assay kinetics. A 96-well plate was coated with a 40 $\mu\text{g/mL}$ solution of lipopolysaccharides (LPS) from *E. coli* in phosphate buffer saline (PBS) and left overnight at 4 °C. The plate was washed with PBS–Tween (PBS-T), 5% milk was added, and the plate was incubated for an hour at room temperature. Then, 92 μL of diluent solution (gelatin veronal buffer (GVB), 5 mM MgEGTA) and 8 μL of NHS were added to each well.

In the 96-well plate, serum incubation times were varied between 15 and 120 min. After incubation, the wells were washed to remove unbound C3b. A 1:1000 dilution of horseradish peroxidase-conjugated antihuman C3 antibody (C3-HRP) in PBS-T was added to the wells and incubated for an hour at room temperature. The wells were washed to remove unbound antibodies. The addition of 1 mM ABTS-H₂O₂ caused a color change (green) that was measured spectrophotometrically at 415 nm after 10 min of incubation. A graph of absorbance versus incubation time was constructed to determine the proper incubation time, based on the highest time point in the linear segment of the curve.

3.2.4 Human C3b Enzyme-linked Immunosorbent Assay

Several concentrations of compstatin analogs were tested using human C3b ELISA to determine the concentration at which half of the C3 present is inhibited by the analog. Compstatin analogs were dissolved in physiological strength PBS, with some analogs dissolved in dimethyl sulfoxide (DMSO) to improve solubility. In Generation A, Peptides VIIA and IXA were dissolved in DMSO; Peptide IVA was initially dissolved in PBS, but subsequently lyophilized, and dissolved in DMSO to improve solubility. In Generation B, Peptides VIIIB-XIB were dissolved in DMSO. The DMSO concentration used was 6%. Initial concentrations of compstatin analogs were calculated using the Beer-Lambert Law with an extinction coefficient of 5500 (1/M·cm) for tryptophan (Trp)

and 5470 (1/M·cm) for methylated-tryptophan (MeW, (30)) present in the sequence, at 280 nm.

A 96-well plate was coated with a 40 µg/mL solution of LPS from *E. coli* in PBS and left overnight at 4 °C. The plate was washed with PBS-T, 5% milk was added, and the plate was incubated for an hour at room temperature. The plate was washed and placed on ice while the dilutions of the compstatin peptides were prepared. Serial dilutions of the compstatin analogs were prepared in an empty 96-well plate, using a diluent solution of GVB and 5 mM MgEGTA. The first column (eight wells) consisted of the highest concentration of the peptide, and serial (two-fold) dilutions were performed. A volume of 92 µL of each serial dilution was transferred to corresponding wells in the ELISA plate. NHS was heat inactivated (HI-NHS) in a 60 °C water bath for 30 min and used as the negative control for complement activation in each experiment. A volume of 8 µL NHS was added to each well, and the plate was incubated at room temperature for 45-120 min, depending on the serum sample. If the compstatin peptide was present at inhibitory concentrations, cleavage of C3 was prevented. After incubation, the wells were washed to remove unbound C3b. A 1:1000 dilution of C3-HRP in PBS-T was added to the wells and incubated for an hour at room temperature. C3-HRP binds to the C3b-LPS complex. The wells were washed again to remove unbound antibodies. The addition of 1 mM ABTS-H₂O₂ caused a color change (green) that is measured spectrophotometrically at 415 nm after 10 min of incubation. Percent C3b was plotted against peptide concentration and was fitted using a logistic dose response curve with

the software Prism (GraphPad, San Diego, CA, USA) to determine IC_{50} . A representation of the ELISA method can be found in Figure 3.5.

In Generation B, several runs were performed with 20% DMSO in the diluent solution (GVB, 5 mM MgEGTA), to avoid aggregation and precipitation of the peptide due to dilutions. A 96-well plate ELISA was run, as described above, with different percentages of DMSO in the diluent (0, 25, 50, 75, and 100%). Absorbance values were plotted against DMSO percentage. At high levels of DMSO (50% and higher) a precipitate is observed; it is likely that the proteins in NHS are been denatured by the high amounts of DMSO (since there is more DMSO present than NHS), causing the observed precipitate. On the other hand, at lower levels of DMSO (20% and lower), the NHS absorbance values are similar to those observed when normal diluent solution is used. Thus 20% DMSO was selected for additions to diluent in cases where the analogs were dissolved in DMSO.

3.3 Results and Discussion

3.3.1 Surface Plasmon Resonance

SPR studies were performed to experimentally assess the binding abilities of the analogs in Generation A, including the two positive controls. Figures 3.6-3.7 show the SPR sensorgrams for the two positive controls, the three computationally design analogs, and three rationally designed analogs. Acetylation was not used in the

computationally designed analogs and controls for comparison with the computational predictions because they are based on native amino acid libraries and parametrization. Table 3.3 summarizes the dissociation constants, K_D , for the binding Peptides IIA–IXA. Maximum RU ($R_{\max}$) and χ^2 values are also listed in Table 3.3 to assess the quality of the fits, in addition to visual inspection of Figures 3.6–3.7. As shown in Table 3.3, the K_D values are sensitive to the fitting model used. Although better fits were generated using the two-state model (Figures 3.6–3.7), we included in Table 3.3 the parameters from fits using the 1:1 model for comparison (sensorgrams not shown). In general, better fits were generated using C3 concentrations up to and including 800 nM. It is likely that at 1600 nM, C3 concentration aggregates may be forming, which change the protein conformation. This may cause rapid surface saturation, which can affect the binding kinetics. The fits were not very sensitive upon inclusion of a 0 nM concentration. The two-state model suggests conformational change of compstatin and/or C3 upon binding, which has been previously proposed using isothermal calorimetry data (16) and by a comparison of the crystal structure of the C3c:W4/A9 complex (21) and the solution structure of free parent compstatin (12).

The SPR data using the two-state model suggest the following binding order (from strongest to weakest): Peptide IIIA > Peptide IXA > Peptide VIIIA > Peptide VIIA > Peptide VA > Peptide IVA > Peptide VIA > Peptide IIA (Table 3.3). Small differences in binding order are seen in the SPR data using the 1:1 model: Peptide IIIA > Peptide IXA > Peptide VA > Peptide VIA > Peptide IVA > Peptide VIIIA > Peptide VIIA > Peptide IIA

(Table 3.3). Although the fits using the two-state model were better (see section 3.2.2 Surface Plasmon Resonance), the results from the 1:1 model produced comparable K_D values, but with different kinetics. The data extracted from both models are reported here to demonstrate the sensitivity of the obtained K_D values and their effect on binding order. The K_D values for the two positive controls, Peptides IIA and IIIA, are in agreement with recently reported data (28). Overall, the SPR data are indicative of strong binding for the newly designed peptides (Peptides IVA–IXA, Table 3.3).

The complexity of the binding kinetics is reflected in the variation of the curve shape and maximum binding response values in the plots of Figures 3.6–3.7. Peptide XA (Table 3.1) contains five Trp substitutions located every third residue, creating a symmetrical arrangement of Trp. However, binding to C3 was not observed in the SPR studies.

3.3.2 Human C3b Enzyme-linked Immunosorbent Assay

3.3.2.1 Generation A

IC_{50} values for inhibition of C3b formation at different analogs concentrations were determined by ELISA (Figure 3.8). Table 3.4 summarizes the determined IC_{50} values of Peptides IIA–IXA of Generation A. The inhibition order (from strongest to weakest) is Peptide IIIA > Peptide VIIIA > Peptide IIA > Peptide IVA > Peptide VIIA > Peptide IXA > Peptide VIA > Peptide VA. Direct correlations between the SPR binding

and ELISA inhibition results are not possible, because the SPR assays are binary with one component immobilized and the ELISA assays are multicomponent detecting inhibition in NHS. However, there are notable differences in some of the peptides used for SPR and ELISA experiments, perhaps owed to the N-terminal acetylation patterns.

In the ELISA studies, all peptides were acetylated, with the exception of Peptides IVA and VIA. Also, the studied peptides have different C-terminal attachments, given the pegylation and biotinylation additions in the SPR peptides (compare Tables 3.3 and 3.4). In the ELISA experiments, acetylation was introduced when the peptide synthesis, cyclization, and purification processes were convenient. Finally, in the ELISA studies, Peptides VII and IX were dissolved in DMSO because of poor solubility in aqueous solutions. Differences in solubilities of peptides that were used in ELISA studies compared with those used in SPR studies may reflect a loss of the hydrophilic N-terminal backbone amine upon acetylation. Solubility was not an issue in the SPR studies, presumably because of the hydrophilic PEG groups in the long PEG-biotin chains at the C-terminal. These data indicate that Trp at position 13 affects peptide solubility, perhaps because of increased hydrophobicity in the region. The relative IC_{50} values of the two positive controls, Peptides IIA and IIIA, are in agreement with the recently reported data, despite differences in their absolute values, owing possibly to differences in the assays and reagents used (28). Overall, the ELISA data are indicative of strong binding and activity for the newly designed peptides (Peptides IVA–IXA). The data suggest that

the retention of the polar character at residues 5-6/9-11 is necessary for both binding of compstatin to C3 and solubility of the peptide.

Molecular dynamics studies of analogs with Trp substitutions at the termini showed that the interactions of C3c:W4/A9 were reproduced. The analog with Trp1 substitution (Peptide VIIA) showed that its side chain interacts with His10 and Leu492, and that Asp6 interacts with Arg459 and had an affinity similar to W4/A9, which is observed in IC_{50} values (Table 3.4, (22)). The analog with Trp13 substitution (Peptide VIIA) showed a π -cation interaction was observed with Arg11, which affects the Cys2-Asn390 hydrogen bond, resulting in a reduced computed affinity in comparison with W4/A9 (22). The same Arg11 interaction was observed in the Trp1/Trp13 substitution (Peptide IXA), but in this analog it appears to facilitate interactions of Trp1 with residues Pro347, Ser388, Asn390, Leu454, and Leu492 (22).

3.3.2.1 Generation B

IC_{50} values for inhibition of C3b formation at different analogs concentrations were determined by ELISA (Figures 3.9-3.10). Table 3.5 summarizes the determined IC_{50} values of Peptides IB–XIB of Generation B. The inhibition order for analogs dissolved in PBS (from strongest to weakest) is Peptide XIB > Peptide IIIB > Peptide IB > Peptide IIB > Peptide XB > Peptide IVB > Peptide VB > Peptide VIB > Peptide VIIIB > Peptide VIIB (Figures 3.9 and 3.10). The analogs with substitutions at position 1 (Peptides IB-IIIB, no N-terminal extension) show similar levels of inhibition as the positive control W4/A9

(Peptide XB). On the other hand, the analogs with N-terminal extensions (Peptides IVB-VIIIB) show lower levels of inhibition than the positive control W4/A9 (Peptide XB).

Molecular dynamics studies have shown that an Arg1 substitution (Peptide IB) maintains the interactions observed in C3c:W4/A9 and a new salt bridge was observed with Asp349 that causes the computed affinity to be slightly better than W4/A9 (22). The Arg1 substitution aids in solubility while obtaining C3 inhibition values slightly better than W4/A9 (22). The N-terminal extension (Peptide IVB) maintains the Cys2-Asn390 interactions while forming a new hydrogen bond between Ser0 and Asn390 (22). The N-terminal extension involving residues Ser or Arg (Peptides IVB-VIIIB), aid in the solubility of the analogs.

The inhibition order for analogs dissolved in DMSO (from strongest to weakest) is Peptide XB > Peptide VIIIB > Peptide IXB > Peptide XIB (Figure 3.10A). These analogs were dissolved in DMSO due to poor solubility in PBS (as visual inspection reveal precipitation). Assays for Peptides VIIIB, XB and XIB were performed in 20% DMSO in the diluent solution to retain a high enough concentration of DMSO to avoid aggregation of the peptide and further precipitation. Different IC_{50} values are observed in several analogs that were dissolved in both PBS and DMSO. It is possible that analogs dissolved in PBS are forming non-visible aggregates because of the hydrophobic nature of the peptides, favoring peptide-peptide interaction and resulting in fewer peptides available for interaction with C3 (taking into account that it has been shown to be a 1:1 interaction). This phenomenon may cause lower IC_{50} values. However, when DMSO is

added in the diluent, aggregation may occur to a lesser extent, because of the presence of organic solvent in the solution. This may cause higher IC_{50} values compared to analogs dissolved in PBS. Table 3.5 also shows statistics for Peptides VIIIB and XB-XIB, for which several ELISA's were performed in PBS and/or DMSO.

Lipophilicity values were calculated for analogs in Generation B. Lipophilicity is a measure of the partition coefficient of a compound between octanol and water. Values are reported as LogP (logarithm base 10 of the partition coefficient), with values of 1 or higher suggesting the use of more organic solvent than aqueous solvent (10:1 ratio or higher), while values of -1 or lower suggesting use of less organic solvent than aqueous solvent (1:10 ratio or lower), and values of 0 suggesting the same ratio of aqueous and organic solvent. Thus, lipophilicity values provide insight into the preferred environment of the analogs from Generation B. Table 3.6 shows the lipophilicity values calculated using different algorithms. Differences in lipophilicity values are observed (reflecting the different algorithms used), however the overall trend is similar and the correlation coefficients are in the range of 0.95-0.99 (Table 3.7). Analog with N-terminal extensions and Arg1 have lower lipophilicity values (Peptides IB-IIIB and IVB-VIIB), demonstrating that polar amino acids at positions -1 to 0 aid in balancing the introduction of a hydrophobic amino acid at position 4, and reflecting their capability to dissolve in PBS. In contrast, Peptides IIIB and VIIIB-XIB (which including the control analogs) have higher lipophilicity values, reflecting dominance of hydrophobic character.

Peptide XIB shows the greatest inhibitory activity, however aggregation in PBS was observed, owed to its very hydrophobic nature. This analog has been a therapeutic candidate for age-related macular degeneration (AMD), a condition that causes loss of central vision because of damage to the macula (the central region of the retina). Several studies have linked proteins of the complement system to predisposition towards AMD (31-35). The analog was injected into the eye of patients and it was observed that it forms an agglomerate in the eye, causing slow release over time (36). This behavior demonstrates the hydrophobic nature of the analog and is in agreement with the observation of aggregates in aqueous solutions *in vitro* by visual inspection.

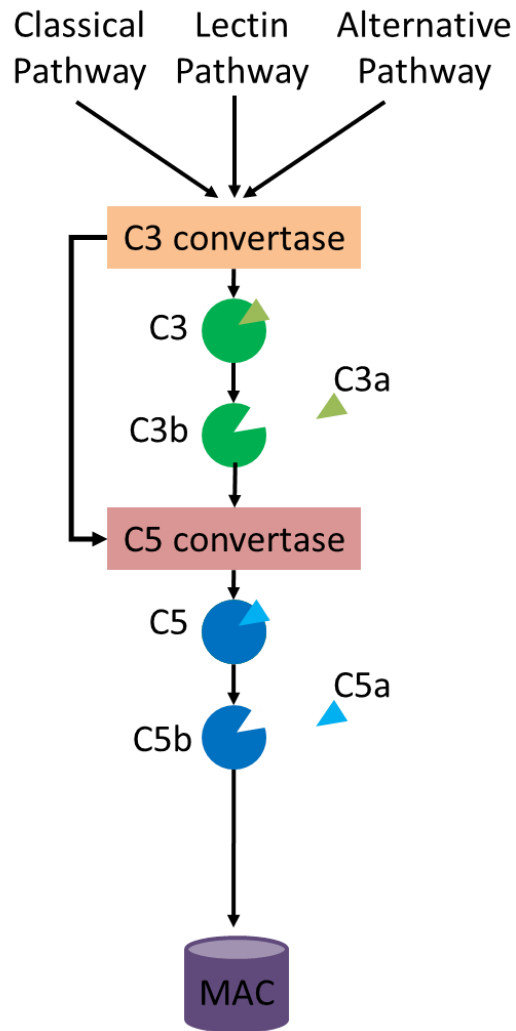
Comparing Generation B with Generation A analogs, substitutions at position 1 and N-terminal extensions (Peptides IB-VIIB) result in better inhibitory activities as well as better solubilities in PBS. Overall, this new Generation B of peptides shows significant improvement in solubility in aqueous buffer (independently of the presence of MeW4) while maintaining an inhibitory activity in the range of W4/A9, especially in analogs with Arg1 substitutions and N-terminal extensions containing polar amino acids (Table 3.5, Peptides IB-IIB and IVB-VIIB). Such polar substitutions produce peptides that can tolerate the presence of the hydrophobic MeW4, while taking advantage of MeW4's contributions to enhancing hydrophobic contacts with C3 and thus increasing activity (Table 3.5, Peptide IIB). Further studies are in process in a collaborative project with Dr. Lincoln Johnson at the University of California-Santa Barbara, to examine the effect of compstatin analogs presented here in the activation of complement system, using

human retinal pigmented epithelial (RPE) cell assays and complement activation by measuring the amount of C5b9 generated (better known as the membrane attack complex) ELISA's. Preliminary results show that several Generation B analogs are potent inhibitors of complement-associated deposit (drusen) formation in the RPE assays, by far surpassing the potency of parent compstatin and the W4/A9 analogs (27).

Further optimization of Generation B analogs is planned with the incorporation of tri-peptide extensions in the N-terminus, by critically selecting amino acids based on capability to form favorable inter-molecular interactions in MD simulations.

3.4 Figures and Tables

Figure 3.1: Diagram of the complement system and its three activation pathways in a simplistic model.



Membrane Attack Complex

Figure 3.2: Parent compstatin and W4/A9 analog. Ribbon representation of (A) parent compstatin free in solution, PDB Code 1A1P (12) and (B) W4/A9 analog, PDB Code 2QKI. (C) Molecular model of C3c:W4/A9 complex, PDB Code 2QKI.

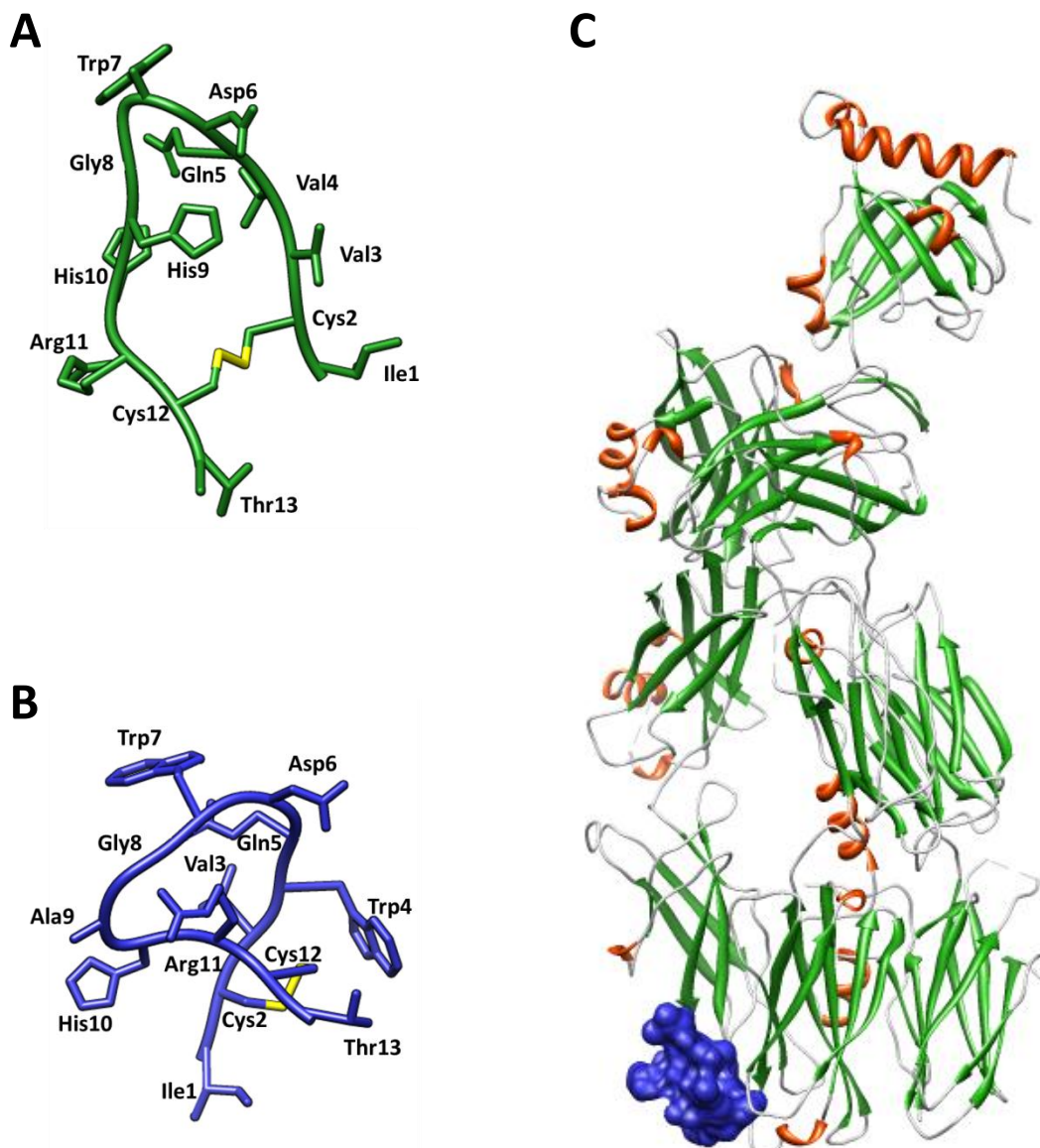


Figure 3.3: Diagram of the SPR procedure. (A) Streptavidin chip components. (B) Immobilization of analogs on the chip. (C) C3 interacts with the analogs on the chip. (D) Regeneration buffer caused disassociation between the analog and C3. (E) Resulting sensorgram for further analysis.

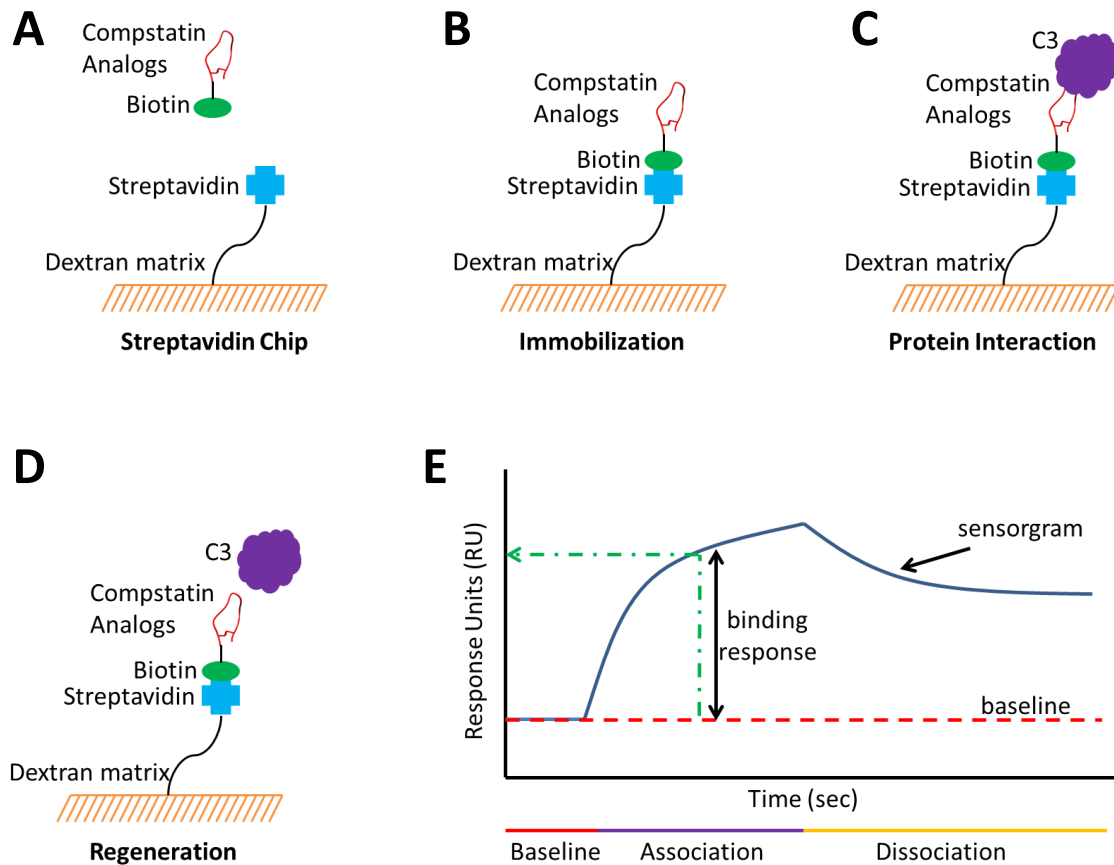


Figure 3.4: Example of SPR data quality. (A) Linear compstatin used as a control in flow cell one (Fc1) is shown in red. Compstatin analog W4/A9/W13 in flow cell two (Fc2) is shown in green at different concentrations of C3 (from top to bottom, 1600, 800, 400, 200, 100, 50, and 25 nM). (B) Adjusted sensorgram of analog W4/A9/W13 by taking the difference of Fc2-Fc1. The color code at the right corresponds to C3 concentrations.

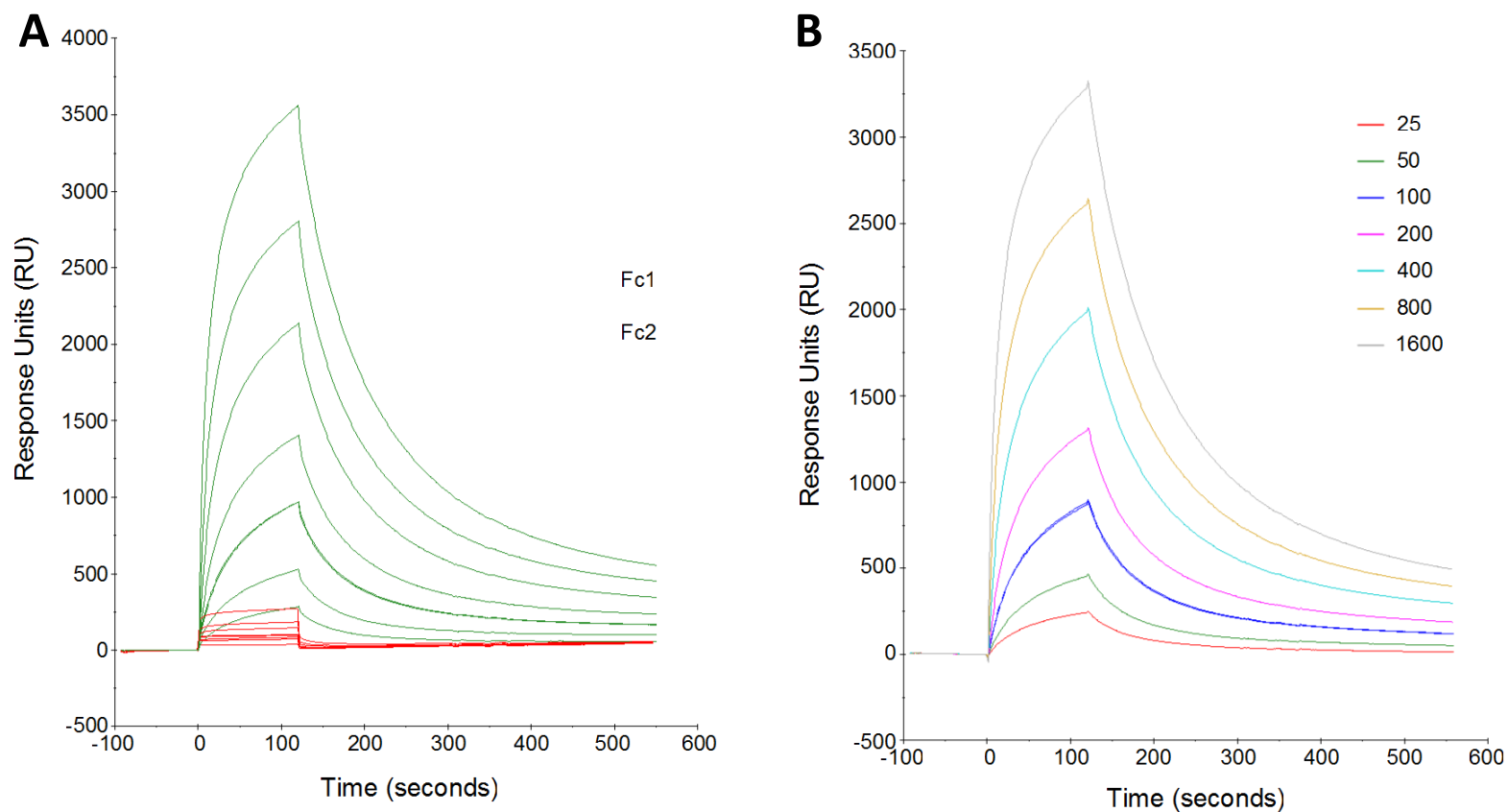


Figure 3.5: Diagram of the ELISA procedure to determine the concentration at which half of the C3 present is inhibited by the peptide.

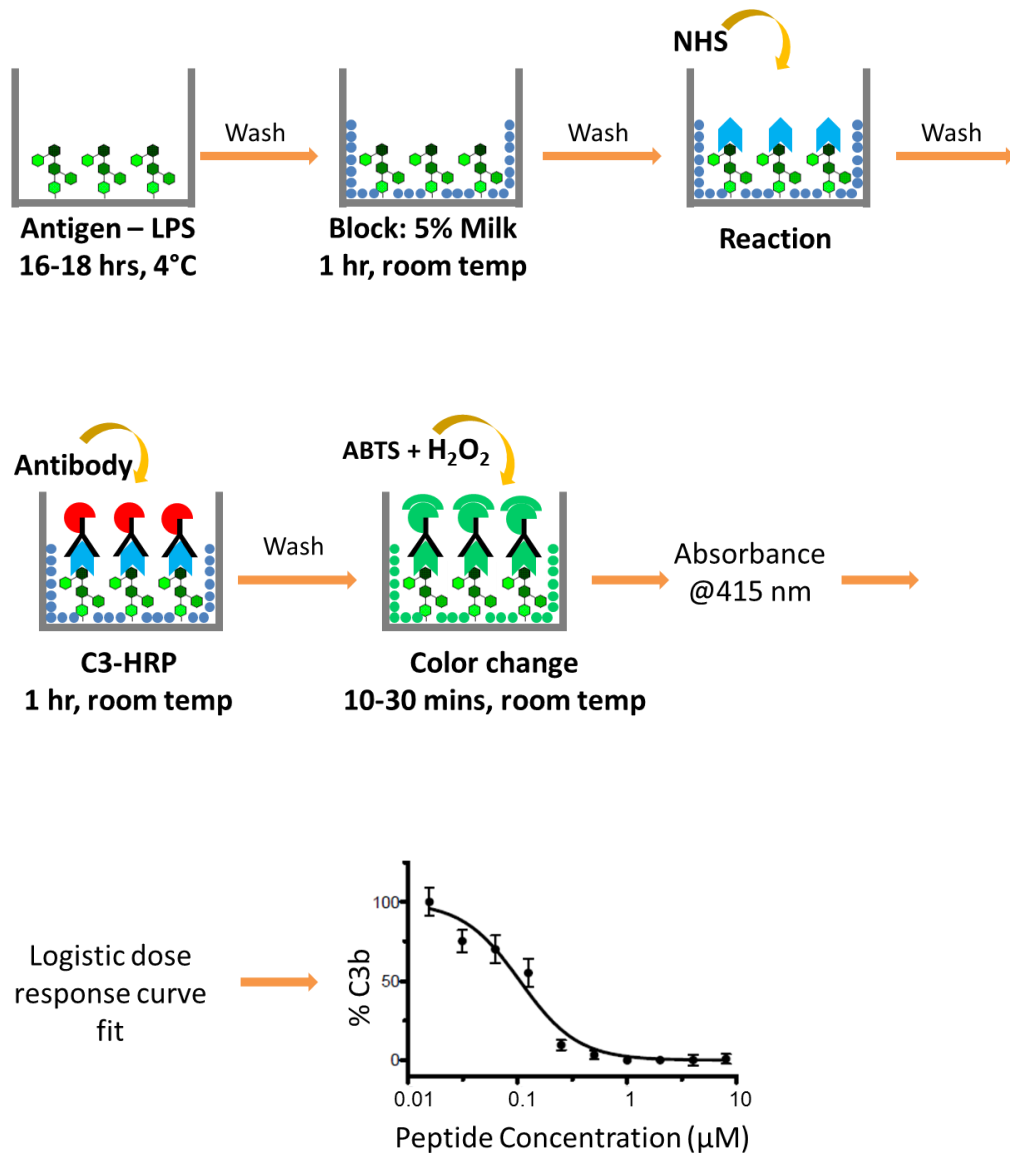


Figure 3.6: Sensorgrams of Peptides IIA–VA from Generation A. The C3 concentrations are 800, 400, 200, 100, 50, and 25 nM (from top to bottom). Solid color lines correspond to the experimental data and solid black lines correspond to the fits. All sensorgrams were fitted to two-state model. (A) Peptide IIA, (B) Peptide IIIA, (C) Peptide IVA, and (D) Peptide VA.

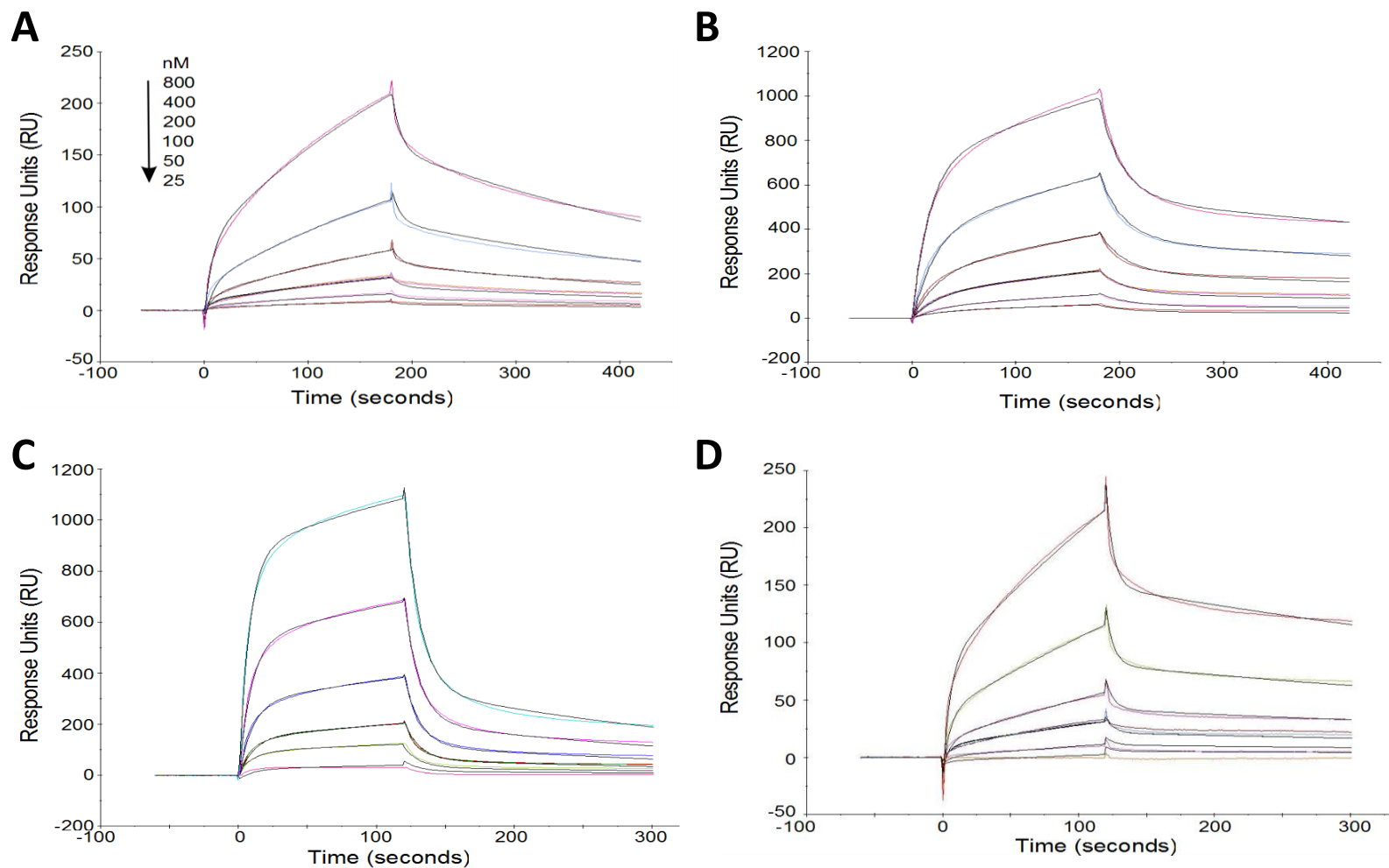


Figure 3.7: Sensorgrams of Peptides VIA–IXA from Generation A. The C3 concentrations are 800, 400, 200, 100, 50, and 25 nM (from top to bottom). Solid color lines correspond to the experimental data and solid black lines correspond to the fits. All sensorgrams were fitted to two-state model. (A) Peptide VIA, (B) Peptide VIIA, (C) Peptide VIIIA, and (D) Peptide IXA.

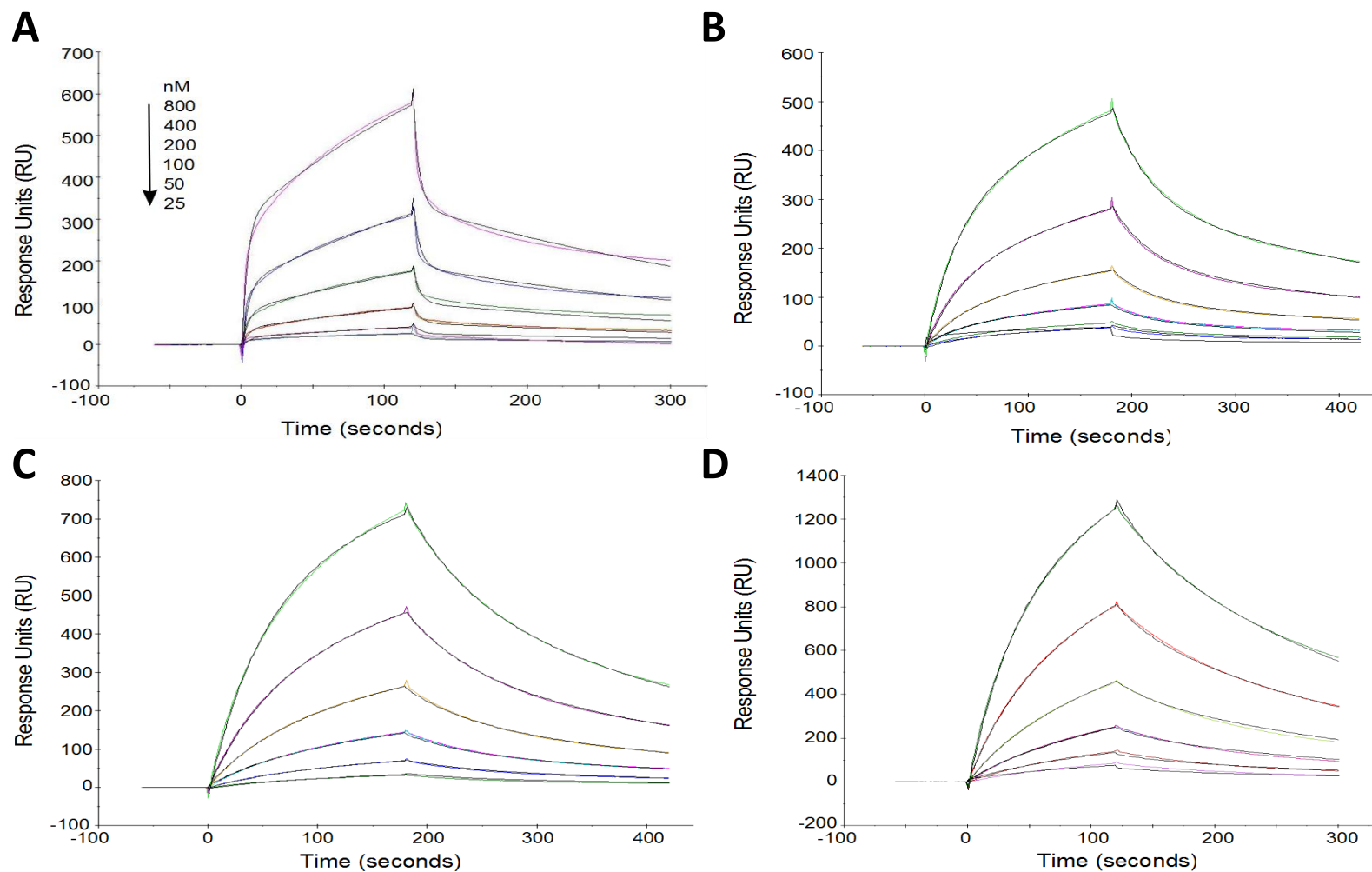


Figure 3.8: Human C3b ELISA data, from Generation A, used to extract the IC_{50} values of Table 3.4. Peptide concentration is plotted in the horizontal axis and percent C3b is plotted in the vertical axis. The plots represent the inhibition of cleavage of C3 to C3a and C3b by compstatin peptides, quantified as percent of C3b generated. Highest inhibition corresponds to 0% C3b generation.

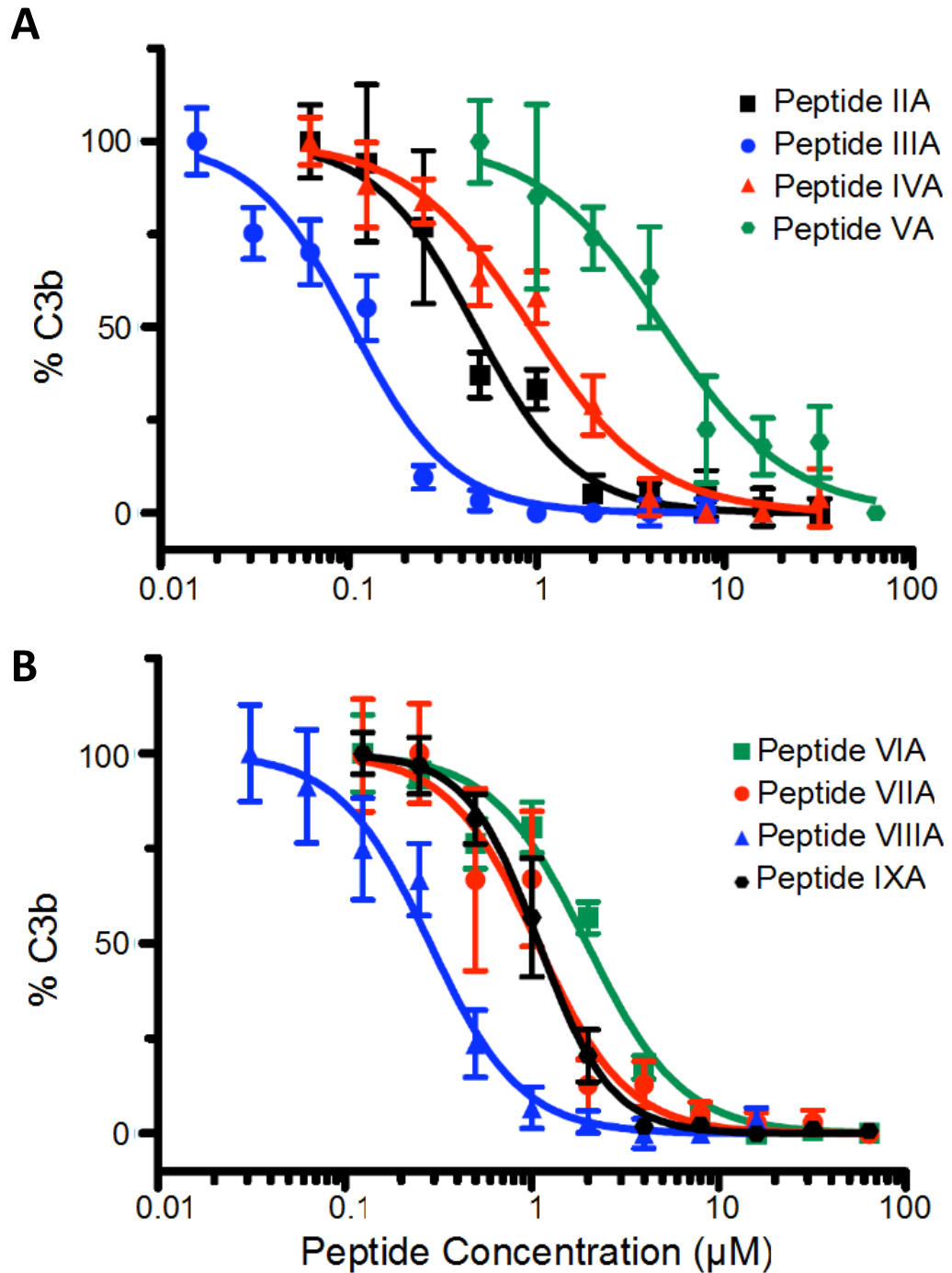


Figure 3.9: Human C3b ELISA data, from Generation B, used to extract the IC_{50} values of Table 3.5. Peptide concentration is plotted in the horizontal axis and percent C3b is plotted in the vertical axis. The plots represent the inhibition of cleavage of C3 to C3a and C3b by compstatin peptides, quantified as percent of C3b generated. Highest inhibition corresponds to 0% C3b generation. (A) Analogs with position 1 substitutions. (B) Analogs with N-terminal extensions.

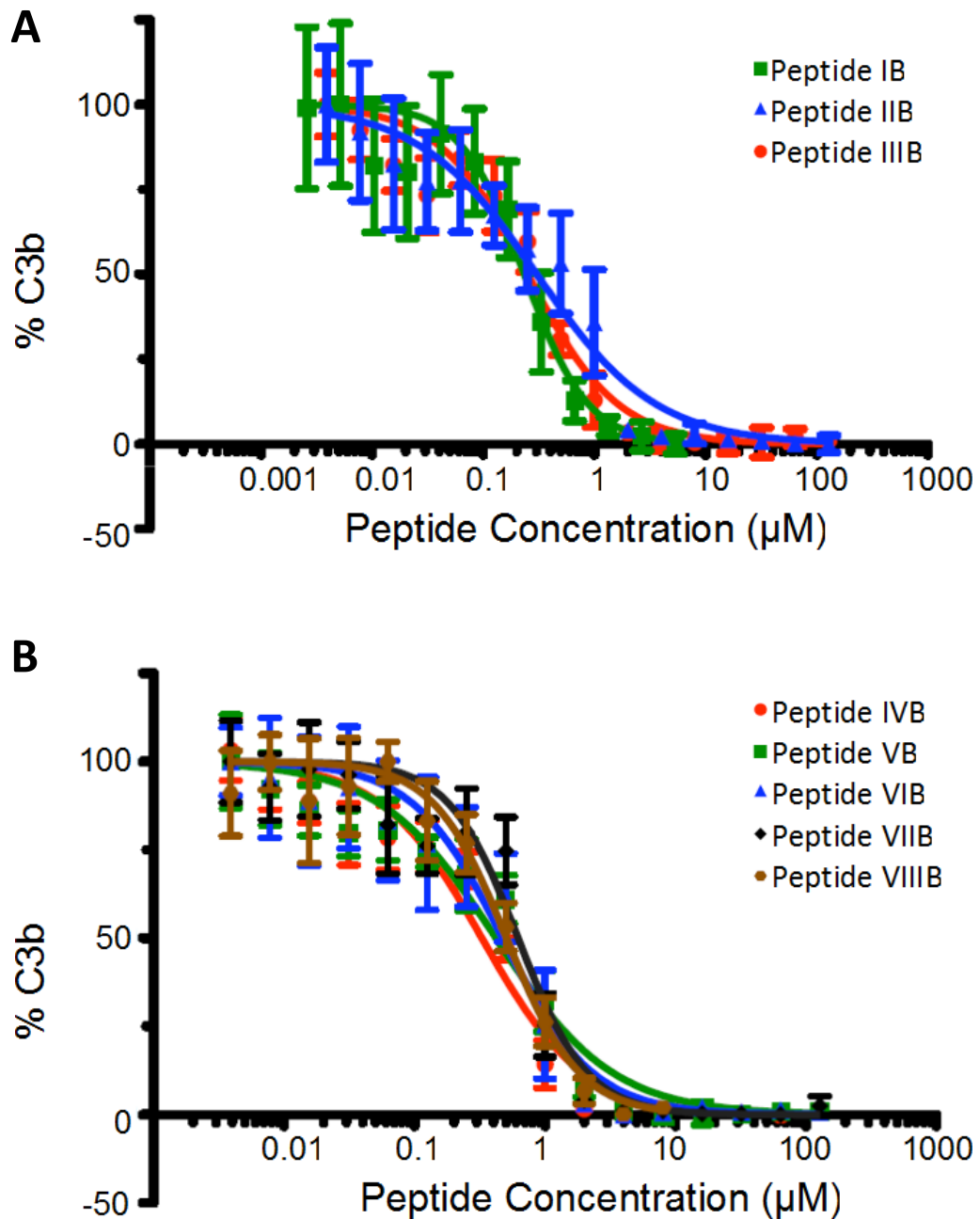


Figure 3.10: Human C3b ELISA data, from Generation B, used to extract the IC_{50} values of Table 3.5. Peptide concentration is plotted in the horizontal axis and percent C3b is plotted in the vertical axis. The plots represent the inhibition of cleavage of C3 to C3a and C3b by compstatin peptides, quantified as percent of C3b generated. Highest inhibition corresponds to 0% C3b generation. (A) Analogs dissolved in DMSO. (B) Control analogs.

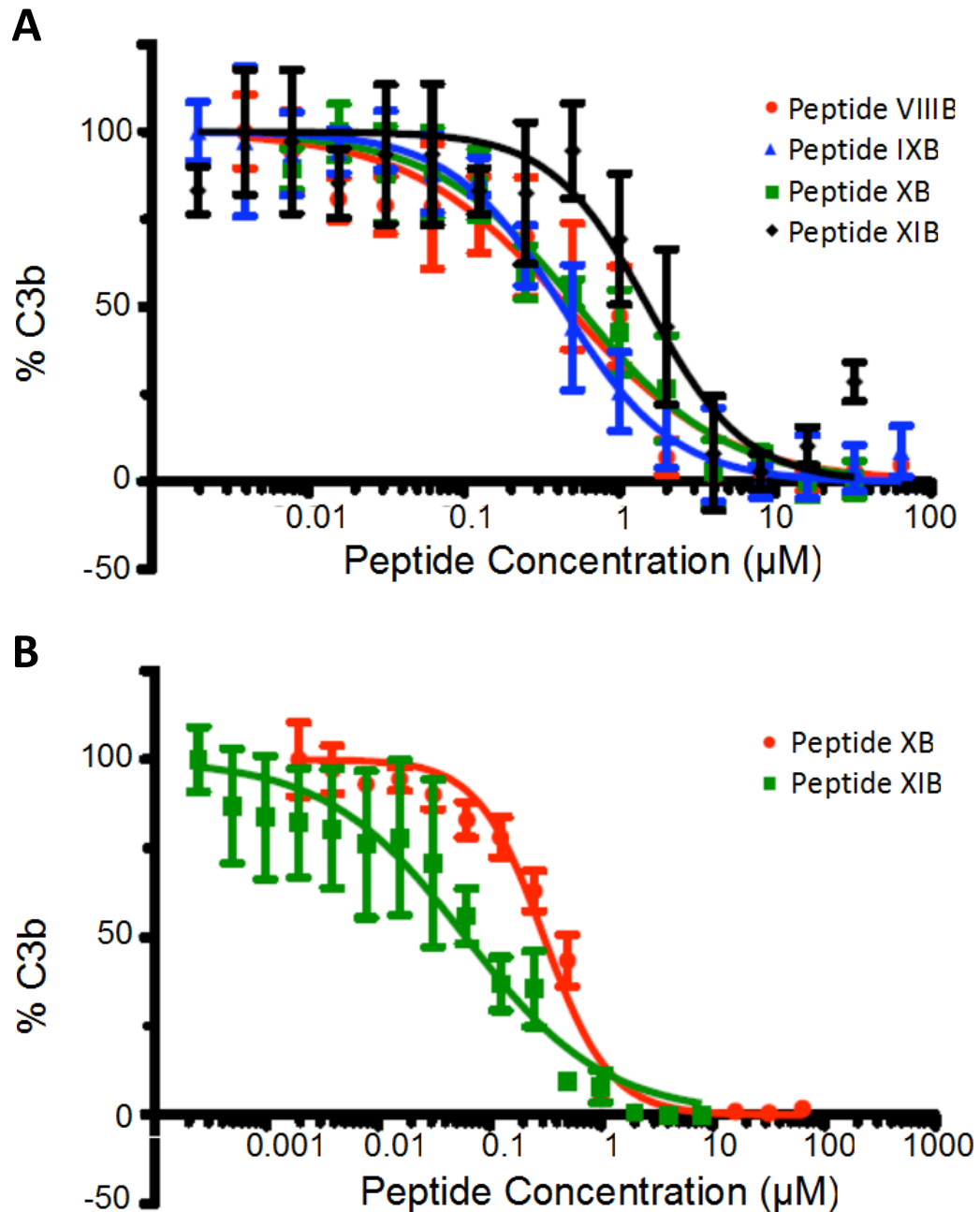


Table 3.1: Compstatin analogs from Generation A. Amino acid sequences and mass spectral analysis for compstatin peptides used in surface plasmon resonance experiments. Brackets denote cyclization. PEG denotes polyethylene glycol. Mass spectral analysis provided by Abgent Inc.

Analog #	Analog Name	Amino Acid Sequence	Mass Spectral Analysis
Peptide IA	Linear	I AVVQDWGHHRA T- (PEG) ₈ -K (biotin)-NH ₂	2267.06
Peptide IIA	Parent	I [CVVQDWGHHRC] T- (PEG) ₈ -K (biotin)-NH ₂	2327.6
Peptide IIIA	W4/A9	I [CV W QDWG A HRC] T- (PEG) ₈ -K (biotin)-NH ₂	2349.2
Peptide IVA	SQ027	W [CV W QDWG TN RC] W - (PEG) ₈ -K (biotin)-NH ₂	2514.9
Peptide VA	SQ059	D [CV W QDWG TN KC] W - (PEG) ₈ -K (biotin)-NH ₂	2414.93
Peptide VIA	SQ086	Q [CV W QDWG QN QC] W - (PEG) ₈ -K (biotin)-NH ₂	2454.4
Peptide VIIA	W4/A9/W13	Ac-I [CV W QDWG A HRC] W - (PEG) ₈ -K (biotin)-NH ₂	2476.0
Peptide VIIIA	W1/W4/A9	Ac- W [CV W QDWG A HRC] T- (PEG) ₈ -K (biotin)-NH ₂	2464.0
Peptide IXA	W1/W4/A9/W13	Ac- W [CV W QDWG A HRC] W - (PEG) ₈ -K (biotin)-NH ₂	2548.7
Peptide XA	W(i, i+3)/A9 ^a	Ac- W [CV W QDWG A WRC] W - (PEG) ₈ -K (biotin)-NH ₂	2598.5

^a This analog contains five Trp, each one at position (i, i+3), to create a symmetrical arrangement of Trp.

Table 3.2: Compstatin analogs from Generation B. Amino acid sequences and mass spectral analysis for compstatin peptides used in ELISA experiments. Sequence differences of various compstatin peptides from W4/A9 are shown in bold characters and methylated-tryptophans are shown by W. Brackets denote cyclization. Mass spectral analysis provided by Abgent Inc.

Analog #	Analog Name	Amino Acid Sequence	Mass Spectral Analysis
Peptide IB	R1/W4/A9	Ac- R [CVWQDWGAHRC] T-NH ₂	1655.4
Peptide IIB	R1/MeW4/A9	Ac- R [CV <u>W</u> QDWGAHRC] T-NH ₂	1670.3
Peptide IIIB	MeW1/W4/A9	Ac- <u>W</u> [CVWQDWGAHRC] T-NH ₂	1700.4
Peptide IVB	S-1/S0/W4/A9	Ac- SSI [CVWQDWGAHRC] T-NH ₂	1787.7
Peptide VB	S-1/S0/R1/MeW4/A9	Ac- SSR [CV <u>W</u> QDWGAHRC] T-NH ₂	1844.3
Peptide VIB	R-1/S0/R1/MeW4/A9	Ac- RSR [CV <u>W</u> QDWGAHRC] T-NH ₂	1913.4
Peptide VIIB	R-1/S0/MeW4/A9	Ac- RSI [CV <u>W</u> QDWGAHRC] T-NH ₂	1870.2
Peptide VIIIB	W-1/W0/R1/W4/A9	Ac- WWR [CVWQDWGAHRC] T-NH ₂	2029.0
Peptide IXB	MeW1/MeW4/A9	Ac- <u>W</u> [CV <u>W</u> QDWGAHRC] T-NH ₂	1714.7
Peptide XB	W4/A9	Ac- I [CVWQDWGAHRC] T-NH ₂	1612.9
Peptide XIB	MeW4/A9	Ac- I [CV <u>W</u> QDWGAHRC] T-NH ₂	1627.5

Table 3.3: Kinetic analysis of analogs from Generation A, calculated for the interaction of active compstatin peptides with human C3 at 0-800 mM.

Analog #	Analog Name	K _D (10 ⁻⁶ M)	R _{max}	K _{on1} (10 ⁵ /M·s)	K _{on2} (10 ⁻² /s)	K _{off1} (10 ⁻² /s)	K _{off2} (10 ⁻² /s)	χ ²	Model
IIA	Parent	1.26	937.7	0.172	1.40	13.04	0.282	4.69	2-State Reaction
		0.81	507.6	0.036		0.294		12	1:1 Binding
IIIA	W4/A9	0.19	1781	0.558	0.52	6.402	0.106	78.1	2-State Reaction
		0.23	1000	0.129		0.296		813	1:1 Binding
IVA	SQ027	0.76	2781.5	0.561	0.36	8.73	0.337	52.27	2-State Reaction
		0.51	1086.2	0.412		2.111		1522.7	1:1 Binding
VA	SQ059	0.74	1150	1.051	1.29	74.30	1.289	9.21	2-State Reaction
		0.45	484.9	0.050		0.225		28.3	1:1 Binding
VIA	SQ086	0.99	2154	2.091	1.03	84.76	0.326	36	2-State Reaction
		0.50	799	0.085		0.418		181	1:1 Binding
VIIA	W4/A9/W13	0.74	1397	0.100	0.55	2.184	0.282	23.8	2-State Reaction
		0.70	939.8	0.069		0.483		108	1:1 Binding
VIIIA	W1/W4/A9	0.50	1516	0.108	0.48	1.146	0.425	6.83	2-State Reaction
		0.52	1297	0.087		0.456		61.4	1:1 Binding
IXA	W1/W4/A9/W13	0.38	2442	0.154	1.38	1.534	0.866	41.7	2-State Reaction
		0.38	2187	0.126		0.477		111.3	1:1 Binding

Table 3.4: IC₅₀ results of analogs from Generation A. from human C3b ELISA experiments. Brackets denote cyclization. Sequence differences of various compstatin peptides from parent compstatin are shown in bold characters. Molecular masses were provided by Abgent Inc.

Peptide #	Peptide Name	Amino Acid Sequence ^a	Mass Spectral Analysis	Concentration Range (μM)	IC ₅₀ (μM)
IIA	Parent	Ac-I [CVVQDWGHHRC] T-NH ₂	1591.47	64-0.125	0.47
IIIA	W4/A9	Ac-I [CV W QDWG A HRC] T-NH ₂	1613.20	8-0.016	0.11
IVA	SQ027	W [CV W QDWG TN RC] W -NH ₂	1735.73	32-0.063	0.94
VA	SQ059	Ac- D [CV W QDWG TNKC] W -NH ₂	1678.27	64-0.500	4.73
VIA	SQ086	Q [CV W QDWG QNQC] W -NH ₂	1675.87	64-0.125	1.98
VIIA	W4/A9/W13	Ac-I [CV W QDWG A HRC] W -NH ₂	1697.87	64-0.125	1.08
VIIIA	W1/W4/A9	Ac- W [CV W QDWG A HRC] T-NH ₂	1685.60	16-0.031	0.29
IXA	W1/W4/A9/W13	Ac- W [CV W QDWG A HRC] W -NH ₂	1771.20	64-0.125	1.09

Table 3.5: IC₅₀ results of analogs from Generation B, from human C3b ELISA experiments.

Analog #	Analog Name	Concentration Range	IC ₅₀ (μM)	Mean ^a	Std. Dev. ^b	Min ^c	Max ^d	Sample (n) ^e
IB	R1/W4/A9	128-0.004	0.252	---	---	0.261	1.498	1
IIB	R1/MeW4/A9	128-0.004	0.303	---	---	0.263	1.254	1
IIIB	MeW1/W4/A9	5.33-0.0002	0.235	---	---	0.223	1.221	1
IVB	S-1/S0/W4/A9	128-0.004	0.328	---	---	0.252	1.848	1
VB	S-1/S0/R1/MeW4/A9	128-0.004	0.421	---	---	0.262	1.704	1
VIB	R-1/S0/R1/MeW4/A9	128-0.004	0.474	---	---	0.302	1.898	1
VIIB	R-1/S0/MeW4/A9	128-0.004	0.622	---	---	0.240	1.762	1
VIIIB	W-1/W0/R1/W4/A9	8-0.0002	0.507	---	---	0.212	1.415	1
	W-1/W0/R1/W4/A9 ^f	128-0.004	0.457	0.510	0.360	0.224	1.332	3
IXB	MeW1/MeW4/A9 ^g	32-0.001	0.542	0.358	0.260	0.249	0.612	2
XB	W4/A9	64-0.002	0.308	0.243	0.150	0.694	1.622	4
	W4/A9 ^g	64-0.002	0.426	---	---	0.229	0.572	1
XIB	MeW4/A9	8-0.0002	0.061	0.045	0.023	0.215	1.914	2
	MeW4/A9 ^g	64-0.002	1.521	1.289	0.330	0.260	0.645	3

^a Mean represent average IC₅₀ values for analogs where ELISA experiments on a 96-well plate were repeated.

^b Standard deviation represent the variation of IC₅₀ values from the mean value.

^c Minimum value represent the lowest absorbance taken for that particular analog.

^d Maximum value represent the highest absorbance taken for that particular analog.

^e Sample size identifies five replications of each concentration in one 96-well plate assay.

^f Assays performed in DMSO.

^g Assays performed in 20% DMSO.

Table 3.6: Lipophilicity values of analogs from Generation B and parent compstatin.

Analog #	Analog Name	JOELIB ^a	XLOGP3 ^b	ALOGPs ^b	miLogP ^c
	Parent	-2.06	-4.18	0.05	-5.40
IB	R1/W4/A9	-2.93	-5.71	-0.38	-5.66
IIB	R1/MeW4/A9	-3.13	-5.76	-0.23	-5.65
IIIB	MeW1/W4/A9	-0.25	-2.85	1.10	-0.67
IVB	S-1/S0/W4/A9	-3.98	-5.80	-0.21	-5.79
VB	S-1/S0/R1/MeW4/A9	-6.18	-8.37	-0.86	-6.11
VIB	R-1/S0/R1/MeW4/A9	-6.12	-8.52	-0.82	-6.13
VIIB	R-1/S0/MeW4/A9	-4.13	-6.00	-0.12	-5.80
VIIIB	W-1/W0/R1/W4/A9	-0.09	-2.79	0.90	-5.15
IXB	MeW1/MeW4/A9	-0.45	-2.90	1.20	-4.90
XB	W4/A9	-0.94	-3.19	0.57	-5.07
XIB	MeW4/A9	-1.14	-3.24	0.76	-5.05

^a Calculated values obtained from Chemmine (<http://chemmine.ucr.edu>).

^b Calculated values obtained from Virtual Computational Chemistry Laboratory (<http://www.vcclab.org/lab/alogps>).

^c Calculated values obtained from Molinspiration Chemoinformatics (<http://www.molinspiration.com>).

Table 3.7: Lipophilicity algorithms comparison. Correlation values were calculated using Excel.

	JOELIB^a	XLOGP3^b	ALOGPs^b	miLogP^c
JOELIB	1	0.99	0.95	0.97
XLOGP3	0.99	1	0.95	0.97
ALOGPs	0.95	0.95	1	0.97
miLogP	0.97	0.97	0.97	1

^a Calculated values obtained from Chemmine (<http://chemmine.ucr.edu>).

^b Calculated values obtained from Virtual Computational Chemistry Laboratory (<http://www.vcclab.org/lab/alogps>).

^c Calculated values obtained from Molinspiration Chemoinformatics (<http://www.molinspiration.com>).

3.5 References

1. Anderson DH, Radeke MJ, Gallo NB, Chapin EA, Johnson PT, Curletti CR, Hancox LS, Hu J, Ebright JN, Malek G, *et al.* The pivotal role of the complement system in aging and age-related macular degeneration: hypothesis re-visited. *Prog Retin Eye Res* 2010, **29**: 95-112.
2. Chen M, Daha MR, Kallenberg CG. The complement system in systemic autoimmune disease. *J Autoimmun* 2010, **34**: J276-286.
3. Sahu A, Lambris JD. Complement inhibitors: a resurgent concept in anti-inflammatory therapeutics. *Immunopharmacology* 2000, **49**: 133-148.
4. Morgan BP, Harris CL. Complement therapeutics; history and current progress. *Mol Immunol* 2003, **40**: 159-170.
5. Morikis D, Lambris JD. Structure, dynamics, activity and function of compstatin and design of more potent analogs. In: Morikis D, Lambris JD, editors. *Structural Biology of the Complement System*. Boca Raton, FL, USA: CRC Press/Taylor & Francis Group; 2005. p. 317-340.
6. Sahu A, Kay BK, Lambris JD. Inhibition of human complement by a C3 binding peptide isolated from a phage displayed random peptide library. *J Immunol* 1996, **157**: 884-891.
7. Morikis D, Lambris JD. Structural aspects and design of low molecular mass complement inhibitors. *Biochem Soc T* 2002, **30**: 1026-1036.
8. Morikis D, Soulika AM, Mallik B, Klepeis JL, Floudas CA, Lambris JD. Improvement of the anti-C3 activity of compstatin using rational and combinatorial approaches. *Biochem Soc T* 2004, **32**: 28-32.
9. Holland MC, Morikis D, Lambris JD. Synthetic small-molecule complement inhibitors. *Curr Opin Investig Drugs* 2004, **5**: 1164-1173.
10. Floudas CA, Klepeis JL, Lambris JD, Morikis D, editors. De novo protein design: an interplay of global optimization, mixed-integer optimization and experiments. *Sixth International Conference on Foundations of Computer-Aided Processes Design, Discovery Through Product and Process Design*; 2004; Princeton University, Princeton, NJ, USA: CACHE Corporation.
11. Ricklin D, Lambris JD. Compstatin: a complement inhibitor on its way to clinical application. *Adv Exp Med Biol* 2008, **632**: 273-292.

12. Morikis D, Assa-Munt N, Sahu A, Lambris JD. Solution structure of compstatin, a potent complement inhibitor. *Protein Sci* 1998, **7**: 619-627.
13. Morikis D, Roy M, Sahu A, Troganis A, Jennings PA, Tsokos GC, Lambris JD. The structural basis of compstatin activity examined by structure-function-based design of peptide analogs and NMR. *J Biol Chem* 2002, **277**: 14942-14953.
14. Sahu A, Soulika AM, Morikis D, Spruce L, Moore WT, Lambris JD. Binding Kinetics, Structure-Activity Relationship, and Biotransformation of the Complement Inhibitor Compstatin. *J Immunol* 2000, **165**: 2491-2499.
15. Klepeis JL, Floudas CA, Morikis D, Tsokos CG, Argyropoulos E, Spruce L, Lambris JD. Integrated computational and experimental approach for lead optimization and design of compstatin variants with improved activity. *J Am Chem Soc* 2003, **125**: 8422-8423.
16. Katragadda M, Morikis D, Lambris JD. Thermodynamic studies on the interaction of the third complement component and its inhibitor, compstatin. *J Biol Chem* 2004, **279**: 54987-54995.
17. Mallik B, Katragadda M, Spruce LA, Carafides C, Tsokos CG, Morikis D, Lambris JD. Design and NMR characterization of active analogues of compstatin containing non-natural amino acids. *J Med Chem* 2005, **48**: 274-286.
18. Katragadda M, Magotti P, Sfyroera G, Lambris JD. Hydrophobic effect and hydrogen bonds account for the improved activity of a complement inhibitor, compstatin. *J Med Chem* 2006, **49**: 4616-4622.
19. Mallik B, Lambris JD, Morikis D. Conformational inter-conversion of compstatin probed with molecular dynamics simulations. *Proteins* 2003, **53**: 130-141.
20. Mallik B, Morikis D. Development of a quasi-dynamic pharmacophore model for anti-complement peptide analogues. *J Am Chem Soc* 2005, **127**: 10967-10976.
21. Janssen BJ, Halff EF, Lambris JD, Gros P. Structure of compstatin in complex with complement component C3c reveals a new mechanism of complement inhibition. *J Biol Chem* 2007, **282**: 29241-29247.
22. Tamamis P, Lopez de Victoria A, Gorham RD, Jr., Bellows-Peterson ML, Pierou P, Floudas CA, Morikis D, Archontis G. Molecular dynamics in drug design: new generations of compstatin analogs. *Chem Biol Drug Des* 2012, **79**: 703-718.

23. Tamamis P, Morikis D, Floudas CA, Archontis G. Species specificity of the complement inhibitor compstatin investigated by all-atom molecular dynamics simulations. *Proteins* 2010, **78**: 2655-2667.
24. Tamamis P, Pierou P, Mytidou C, Floudas CA, Morikis D, Archontis G. Design of a modified mouse protein with ligand binding properties of its human analog by molecular dynamics simulations: The case of C3 inhibition by compstatin. *Proteins* 2011, **79**: 3166-3179.
25. Bellows ML, Fung HK, Taylor MS, Floudas CA, López de Victoria A, Morikis D. New compstatin variants through two de novo protein design frameworks. *Biophys J* 2010, **98**: 2337-2346.
26. López de Victoria A, Gorham RD, Jr., Bellows-Peterson ML, Ling J, Lo DD, Floudas CA, Morikis D. A new generation of potent complement inhibitors of the Compstatin family. *Chem Biol Drug Des* 2011, **77**: 431-440.
27. López de Victoria A, Forest DL, Kieslich CA, Banna CD, Tamamis P, Gorham RD, Jr., Bellows-Peterson ML, Floudas CA, Archontis G, Johnson LV, *et al.* New potent complement inhibitors of the compstatin family peptides prevent complement-associated deposit formation in human retinal pigmented epithelial cell cultures. *In Preparation* 2012.
28. Magotti P, Ricklin D, Qu H, Wu YQ, Kaznessis YN, Lambris JD. Structure-kinetic relationship analysis of the therapeutic complement inhibitor compstatin. *J Mol Recognit* 2009, **22**: 495-505.
29. Klepeis JL, Floudas CA, Morikis D, Tsokos CG, Lambris JD. Design of peptide analogues with improved activity using a novel de novo protein design approach. *Ind Eng Chem Res* 2004, **43**: 3817-3826.
30. Marmorstein RQ, Joachimiak A, Sprinzl M, Singler PB. The Structural Basis for the Interaction between L-Tryptophan and the Escherichia coli trp Aporepressor. *J Biol Chem* 1987, **262**: 4922-4927.
31. Johnson LV, Leitner WP, Staples MK, Anderson DH. Complement activation and inflammatory processes in Drusen formation and age related macular degeneration. *Exp Eye Res* 2001, **73**: 887-896.
32. Hageman GS, Anderson DH, Johnson LV, Hancox LS, Taiber AJ, Hardisty LI, Hageman JL, Stockman HA, Borchardt JD, Gehrs KM, *et al.* A common haplotype in the complement regulatory gene factor H (HF1/CFH) predisposes individuals to age-related macular degeneration. *Proc Natl Acad Sci U S A* 2005, **102**: 7227-7232.

33. Gold B, Merriam JE, Zernant J, Hancox LS, Taiber AJ, Gehrs K, Cramer K, Neel J, Bergeron J, Barile GR, *et al.* Variation in factor B (BF) and complement component 2 (C2) genes is associated with age-related macular degeneration. *Nat Genet* 2006, **38**: 458-462.
34. Maller JB, Fagerness JA, Reynolds RC, Neale BM, Daly MJ, Seddon JM. Variation in complement factor 3 is associated with risk of age-related macular degeneration. *Nat Genet* 2007, **39**: 1200-1201.
35. Fagerness JA, Maller JB, Neale BM, Reynolds RC, Daly MJ, Seddon JM. Variation near complement factor I is associated with risk of advanced AMD. *Eur J Hum Genet* 2009, **17**: 100-104.
36. Potentia Pharmaceuticals Inc., Francois C, Deschatelets P, Olson P, inventors. Sustained delivery of compstatin analogs from gels. USA 2009, Patent Application WO2009046198.

APPENDICES

A.1 HIV-1 Subtypes Studies Year 1999

A.1.1 Consensus Sequences

Consensus sequences for the V3 loop of several subtypes of HIV-1 were obtained from the Los Alamos National Laboratory Database, from the 1999 HIV Compendium (1). The Los Alamos consensus sequences have assigned frequencies of observation for each amino acid in every specific position. The database also provides additional sequences with lowest frequencies for all subtypes (1). For this study, the amino acid sequences between and including the first and last cysteines were used. Modifications were performed to some of the V3 loop consensus sequences that had more than the 35 amino acids of the templates. The consensus sequence of the subtype D has 34 amino acids and the alignment introduced 4 gaps (1). To obtain a 35 amino acids sequence for subtype D, four consensus sequences were generated, named D1 to D4, in which 3 gaps were eliminated and one gap was replaced with the amino acid of the next consensus sequence (the next amino acid with higher frequency, with the gap having the highest frequency), for all possible combinations (Table A.1). In the case of D1, the gap at position 8 (1) was replaced by Ile whereas the gaps at positions 24, 27, and 28 (1) were eliminated, to form a sequence of 35 amino acids that matches the amino acid number in the number template sequence. Subtype H and G have 35 amino acids, but the

consensus sequence alignments have introduced two gaps (1), which were eliminated. Subtype K has a V3 loop with a sequence of 36 amino acids and no gaps in the consensus alignment (1). In this case the amino acid with the lowest frequency was eliminated. Because three amino acids had the lowest frequency, one at a time was eliminated to form three subtype K sequences, named K1 to K3. The sequences used in this study are shown in Table A.1. There is a total of 21 sequences, including the two templates that belong to the B subtype.

The structures of the V3 loop subtypes with the 35-amino acid consensus sequences of Table A.1 were generated using the program WHAT IF (2) and templates from the structures with PDB codes 2B4C (3) and 2QAD (4). Amino acid replacements were performed using the side chain rotamer library of WHAT IF. Unfavorable van der Waals clashes, introduced by the mutations, were eliminated from the WHAT IF-derived structures using 1000 steps of steepest descent energy minimization with the program SPDBV (5).

A.1.2 Electrostatic Calculations

Poisson-Boltzmann electrostatic calculations and hierarchical clustering analysis were performed as described elsewhere (6-11). The program PDB2PQR (12) was used to prepare the V3 loop coordinates for electrostatic calculations by including van der Waals radii and partial charges for all atoms according to the PARSE forcefield (13).

Electrostatic potentials were calculated using the Adaptive Poisson Boltzmann Solver (APBS, (14)) and the linearized form of the Poisson-Boltzmann equation. A box with $129 \times 129 \times 129$ grid points was used. The box dimensions were $65 \text{ \AA} \times 65 \text{ \AA} \times 65 \text{ \AA}$ for 0 and 150 mM, for both structural templates. The molecular surface was calculated using a probe sphere with a radius of $1.4 \text{ \AA}$, representing a water molecule. The dielectric coefficients were set to 2 and 78.54 for the protein interior and solvent, respectively. The ion accessibility surface was calculated using a probe sphere with radius of $2.0 \text{ \AA}$, representing monovalent counterions. Calculations were repeated with ionic strengths corresponding to 0 mM salt concentration (representing Coulombic interactions unscreened by counterions) and 150 mM (representing physiological ionic strength in serum).

Electrostatic similarity distances (ESDs) were calculated according to

$$ESD_{a,b} = \frac{1}{N} \sum_{i,j,k} \frac{|\Phi_a(i,j,k) - \Phi_b(i,j,k)|}{\max\{|\Phi_a(i,j,k)|, |\Phi_b(i,j,k)|\}}$$

where Φ_a and Φ_b are the electrostatic potentials of proteins a and b at grid point (i, j, k) and N is the total number of grid points. This error-type relation compares the spatial distributions of electrostatic potentials of pairs of proteins. The normalization factor of the denominator assures small values in the vicinity of the 0-2 range, with 0 corresponding to identical spatial distributions of electrostatic potentials and 2 to totally different. Visualization of the spatial distributions of electrostatic potentials, as

isopotential contour surfaces, was accomplished using the program Chimera (15).

The ESD shown above was also applied to cluster subtype sequences based on charge distribution maps using APBS. Hierarchical clustering analyses were performed using the linkage function of MatLab (The MathWorks Inc., Natick, MA). The clustered data were plotted as dendrograms using MatLab.

A.1.3 Sequence Clustering

Alignment for all HIV-1 subtype sequences of Table A.1 was performed using ClustalW2 (16). The score matrix generated by ClustalW2 was used as the input distance file to create a clustering dendrogram using the linkage function of MatLab.

A.1.4 Results and Discussion

Consensus sequences for the V3 loop of several subtypes of HIV-1 were obtained from the Los Alamos National Laboratory database (1). The Los Alamos consensus sequences have assigned frequencies of observation for each amino acid in every specific position. The database also provides additional sequences with lowest frequencies for all subtypes (1). For this study, the amino acid sequences between and including the first and last cysteines were used. Modifications were performed to some

of the V3-loop consensus sequences that had more than the 35 amino acids as described above.

The sequences used in this study are shown in Table A.1. There are a total of 21 sequences, including the two templates that belong to the B subtype. Electrostatic potentials for all the V3 loop subtypes were calculated as described in above. Figure A.1 shows the dendrogram that clusters the electrostatic potential distributions of HIV-1 subtypes when 2B4C (3) was used as the template. V3 loop subtypes with similar spatial distribution of electrostatic potential cluster together. The V3 loop sequences studied have positive net charges, with the exception of subtype O, which has a neutral charge. Sequences of fourteen subtypes have a net charge +3 (A, G, AG, C, B, K3, D1, D2, F, AE, K1, AB, K3, 2QAD and 2B4C; Figure A.1). Sequences of the D subtypes differ in charge; D4 and D3 have charge +4; D1 and D2 have charge +3. Similarly for the K subtypes; K1 and K2 have charge +3, and K3 has charge +4. While subtypes with the same net charge cluster together in a module, smaller submodules exist for subtypes with the same charge. Comparable results were obtained using 2QAD (4) as the template (Figure A.2), and using ionic strength corresponding to 150 mM for the two structural templates (Figures A.3-A.4). Consistency in the clustering of spatial distributions of electrostatic potential independently of structural template or ionic strength value is observed in the data. The dominant factors in the clustering analyses are net charge and the spatial distribution of electrostatic potential. Correlations between the clusters and available epidemiological data on global prevalence and geographic distribution (year 2004, (17))

were made. The coreceptor assignment was made following the “11|24|25” rule (18), resulting in one subtype, subtype D, identified as X4 tropic. Subtype C is responsible for approximately 50% of the infected population and forms a module together with subtypes A, G and AG, together accounting for 75% of the infected population. The two crystal structures used as templates belong to subtype B and are found clustered with subtype B. The subtype B module accounts for an additional 10% of the infected population. Geographic distributions (17) are also presented in Figures A.1-A.2, though clear correlations were not possible.

The clustering of the distribution of charges in space for each subtype is shown in Figure A.5 for 2B4C as the template and Figure A.6 for 2QAD as the template. Some clusters within these dendrograms can be found in the electrostatic potential clustering (Figures A.1-A.4) (e.g., C, AG, A, and G). However, the subtypes are mostly mixed. In general, charge distribution does not depict subtle differences between the subtypes. This is because charges are localized in the structure and are independent from each other.

Alignment for all HIV-1 subtype sequences was performed using ClustalW2 (Table A.1, (16)). The core matrix generated by ClustalW2 was the input file for then sequence clustering dendrogram (Figure A.7). Subtypes G, C, AG, and A cluster together; and subtypes B, 2B4C, and 2QAD cluster together, accounting in total for 75% of the infected population. Subtype O (net charge 0) clusters on its own. Two of the subtypes

with charge +4 (D3, D4) are merged together with their related subtypes with net charge +3 (D1, D2) to form a D subtype cluster. Subtype K2 (net charge +4) clusters with its related subtype K1, although subtype K3 is not present in that cluster.

Clustering analyses were used to show that the spatial distributions of electrostatic potential of the V3 loops provide a refined analysis of similarities/dissimilarities among the consensus sequences of the HIV-1 subtypes, compared to clustering based on sequence alignments. Because the electrostatic potentials are generated by charge, which is included in the sequence information, sequence clustering also depicts the major similarities/dissimilarities among the subtype, albeit at a slightly coarser level.

A.2: V3 Loop Molecular Dynamics

An example of the configuration file used in NAMD (19) for the V3 loop molecular dynamics (MD) simulations.

```
#####
## JOB DESCRIPTION                                     ##
#####

# Minimization and Equilibration in a Water Box

#####
## ADJUSTABLE PARAMETERS                             ##
#####

structure          2b4c_v3a_wb10_ions.psf ;# topology file
coordinates         2b4c_v3a_wb10_ions.pdb ;# pdb with H2O & ions
```

```

set temperature      298
set outputname       2b4c_v3a_wb10_ions ;# name of files generated

firsttimestep        0

#####
## SIMULATION PARAMETERS                                     ##
#####

# Input
paraTypeCharmm        on
parameters             par_all127_ck.prm
temperature            $temperature

# Force-Field Parameters
exclude                scaled1-4
1-4scaling             1.0
cutoff                 12
switching              on
switchdist             10
pairlistdist           13.5

# Integrator Parameters
timestep               2.0 ;# 2fs/step
rigidBonds             all ;# needed for 2fs steps
nonbondedFreq          1 ;# nonbonded forces every step
fullElectFrequency     2 ;# PME only every other step
stepspercycle          10

# Constant Temperature Control
langevin               on ;# do langevin dynamics
langevinDamping         5 ;# damping coefficient (gamma) of 5/ps
langevinTemp            $temperature
langevinHydrogen        off ;# don't couple langevin bath to
hydrogens

# Periodic Boundary Conditions
cellBasisVector1       45    0    0
cellBasisVector2        0   47    0
cellBasisVector3        0    0   64
cellOrigin              93.16 -137.12 160.96 ;# center of the cell

wrapAll                on

```

```

# PME (for full-system periodic electrostatics)
PME                yes
PMEGridSpacing     1.0

# Constant Pressure Control (variable volume)
useGroupPressure   yes ;# needed for rigidBonds
useFlexibleCell    no
useConstantArea    no

langevinPiston     on
langevinPistonTarget 1.01325 ;# pressure in bar -> 1 atm
langevinPistonPeriod 100 ;# oscillation period around 100fs
langevinPistonDecay 50 ;# oscillation decay time of 50fs
langevinPistonTemp $temperature

# Output
outputName         $outputname

restartfreq        5000 ;# 5000steps = every 10ps
dcdfreq           5000 ;# how often it saved information
xstFreq           5000
outputEnergies     5000
outputPressure     5000

#####
## EXTRA PARAMETERS                                     ##
#####

#####
## EXECUTION SCRIPT                                     ##
#####

# Minimization
minimize           1000 ;# lower potential energy for 100 steps
reinitvels         $temperature

run 5000000 ;# 10 ns

```

A.2.1 Electrostatics Analyses

Electrostatic analysis using 100 snapshots from each V3 loop trajectory, one snapshot per nanosecond, was performed as described in section 2.2 Methods. Figures A.8-A.9 show the dendrogram that clusters spatial distributions of the snapshots electrostatic potentials using 0 mM and 150 mM ionic strength, respectively. Two superclusters are observed in both figures, with finer subclusters within. Scrambling within snapshots from both structures are found within the supercluster, while finer subclusters are composed mostly of snapshots from the same structure. However, there are several subclusters, mostly at 0 mM, where snapshots from both structures are found. In 150 mM, the snapshots from each structure cluster together, with the exception of a small set of snapshots of 2B4C within a 2QAD cluster. The scrambling of snapshots from both structures suggests that at those specific time points, the electrostatic potentials between both V3 loop structures are similar.

A.3: Surface Plasmon Resonance Protocol

The Surface Plasmon Resonance (SPR) assay is an interaction assay that tests the binding between compstatin (ligand) and C3 (analyte). SPR assays can also help determine the kinetics of binding between the species by fitting the results to known binding models.

Safety Notes: Always use gloves to protect yourself and the samples. Make sure to label everything to avoid mistakes.

A. Materials list and solution formulas are shown in Tables A.2

B. C3 Buffer Exchange (3-4hrs): Perform before running the Biacore Assay (normally during the morning).

1. **Note:** C3 comes in a buffer of 50mM phosphate, sodium azide, 145mM NaCl at pH 7.8.
2. C3 general information (concentration calculation example below, in #9)
 - a. MW: 185kDa
 - b. $\epsilon^{1\%} = 10.3$
 - i. Multiply the concentration (%) by 10 to convert to mg/mL
 - c. Store at -80°C
 - d. Do not refreeze
3. Add 300-500µL of protein sample to the Amicon-Ultra.
4. Centrifuge at 10000rcf for 15min at 4°C. The volume will drop to ~100 µL.
5. Discard flow-through
6. Add ~400µL running buffer
7. Repeat steps 3-6 3 times.
8. Recover the protein of the last run by adding 500µL of running buffer, 100µL at a time. Pipette up and down a few times and transfer to a new tube.
9. Measure absorbance of C3 at 280nm and calculate concentration.
 - a. Concentration will be lower than original.
 - b. Example (Beer-Lambert Law)
 - i. $Abs = \epsilon bc$
 - ii. absorbance 280nm = 0.009, using a cuvette = 1cm (Figure A.10), with a dilution factor (DF) of 100 (500 µL of sample/5 µL of protein = 100)
 $0.009 * 100 = 0.9 \rightarrow 0.9/10.3 = 0.087379\%$
 $0.087379\% * 10 \text{ (conversion factor)} = 0.87 \text{ mg/mL}$
 $0.87 \text{ mg/mL} * (1/185\text{kDa}) * 10^6 = 4.7\mu\text{M} * 103 = 4702.7\text{nM}$
10. Calculate the concentrations needed for the experiment using $C_1V_1=C_2V_2$
 - a. Example: If you want a final concentration of 25nM using 4.7µM as the starting concentration
 $4702.7\text{nM} * V_1 = 25\text{nM} * 200\mu\text{L} \rightarrow V_1 = 1.06 \mu\text{L}$
volume from original sample: 1.06 µL
volume from buffer: 198.94 µL
total volume: 200 µL

C. Compstatin Preparation (~30min)

1. Weight ~ 0.0007g (0.7mg) of Compstatin.
 - a. **Note:** Amount of peptide may have to be adjusted depending on its MW.

2. Dissolve in 0.5mL of ddH₂O. Make sure it is completely dissolved.
 $0.7\text{mg}/0.5\text{mL} = 1.4\text{mg/mL} \rightarrow 1.4\text{mg/mL} * (1/\text{MW}) * 10^6 = [\mu\text{M}]$
3. Create a 100nM concentration of Compstatin for the Biacore. It can be calculated using $C_1V_1=C_2V_2$ (see previous example).
4. Can be store as aliquots at -80°C.

Note: Turn the Biacore X100 on an hour before use to allow temperature to stabilize (25°C). If instrument is in “Standby”, it can be used immediately. Instructions have to be given to the machine by using the software before docking, adding solutions, etc.

D. Sensorship Docking (using the software)

- a. **Notes:** Equilibrate the sensorship to room temperature before use. Keep free of particles and dust. Store at 4°C before use and after use add a little bit of ddH₂O. Make sure to try an old sensorship before the actual experiment to make sure it is still good.
1. Tools
 - a. Stop Standby: Change the solution to buffer
 - b. Standby
2. Undocked chip → place new chip → Dock chip

E. Compstatin Immobilization

1. Using the Wizard, select immobilization and specify the parameters:
 - a. Prime the chip with conditioning solution.
 - i. 3 times, 1min per time, at 30μL/min
 - b. Select Fc2 channel for immobilization of Compstatin (sample).
 - i. Flow rate: 5μL/min
 - ii. 30secs injection with 200nM of Compstatin
 - iii. Target RU = 500
 - c. Select Fc1 channel for immobilization of the control.
 - i. Flow rate: 5μL/min
 - ii. 30secs injection with 200nM of Compstatin
 - iii. Target RU = 500
 - d. Change the channel mode to Fc1 and Fc2.
 - e. Stabilize the chip for 5min before the assay.

F. Test Binding Conditions

1. This step is necessary if the binding conditions are not known. For example, if the running buffer has to be changed, if testing another analyte, etc.
2. Use manual run to test the conditions.
3. Inject medium concentration of C3 for 120secs and observed RU.
 - a. Try to reach >100RU.

4. Regenerate the chip with the weakest solution first for 60secs. The minimum condition that returns the baseline to the start point is the best regeneration solution.
5. Repeat binding and regeneration with the optimal conditions to confirm reproducibility

G. Binding Assay (3-4hrs)

1. Choose Kinetics/Affinity Wizard
 - a. Flow cells: Fc1 and Fc2
 - b. Chip type: SA
 - c. One regeneration after each sample
 - d. Primer before run with buffer, 1 cycle
 - e. Sample Contact 180secs (3min) and Dissociation 240secs (4min)
 - a. Max association is 180secs.
 - b. Max dissociation is 600secs.
 - f. Regeneration for 30 secs with 10mM NaOH, and 30secs with 0.05% SDS
 - g. Input Concentrations (in nM, 25, 50, 100, 200, 400, 800, 1600)
 - a. The program required that the first concentration has to be 0nM.
 - b. All samples need to have the same name.
 - c. The program required that one concentration is run twice, for reproducibility purposes. It can be a concentration in the middle of the range.
 - a. 100nM is normally used
 - h. Flow rate is automatically set to 30 μ L/min (which is the maximum).
 - i. **Note:** This is saved as "template".
2. The program will give the order of the samples and the amount needed for each one (including buffer and regeneration solutions). Load each sample/solution in the tubes and place them in the rack.
3. Save the assay in a Folder as a result (name_date_hour) and start the assay.
4. The results are automatically shown in the Evaluation Software after the assay finishes (sensorgrams, Fc2-Fc1).
5. If the instrument is not going to be used during the next 4 days, shut it down (the system will guide you through the procedure). Otherwise put it on Standby (changing the chip to the maintenance chip and the buffer to ddH₂O).

H. Data Analysis (Evaluation Software)

1. Open the results. Make sure you already have the sensorgrams (Fc2-Fc1).
 - a. Adjust the sensorgram
 - i. Select the section after dissociation is finish by using right click. Right click again and select cut.

- ii. Tools (top right of the sensorgram window)
 1. y-adjustment: Select baseline
 2. x-adjustment: Select injection, select sample 1 start
2. Select kinetics (it will show a table).
 - a. If you want to use all concentrations for the fit, click next. If you want to eliminate certain concentrations, unselect them from the table (check mark at the left) and click next.
 - b. Select Kinetics
 - c. Select the Models to use.
 - i. Models used in Compstatin Study
 - a. 1:1 binding
 - b. Two state reaction
 - c. Heterogeneous ligand
 - d. Click fit.
 - i. First two should have a green checkmark.
 - e. Steps c and d can be repeated for each model, then click finish.
3. The results will be in the "Report" page (k_a , k_d , KD , R_{max} , X^2 , etc).
 - a. R_{max} : This number is calculated based on the tendency of the data. Is not the RU from the highest concentration.
 - b. χ^2 : This value will reflect how "good" the fit is. The lower the number, the better the fit.
 - i. Normally the X^2 should be 10% of the R_{max} .
 - ii. Better fits are achieved when χ^2 is lower than 10, but this is not easily achievable.
 - c. tc : Is a value related to mass transfer. The higher this value the better.
 - i. The value should be 10^8 - 10^9 or higher.
4. Revised the "Residual" tab. The closer the data points to the central line, the better the fit is.
5. Save the results (results_name_date).
 - a. Right click in the sensorgrams to copy and paste them in a drawing software (like Paint) and save them in your preferred picture format.

A.4 Human C3b Enzyme-linked Immunosorbent Assay

A.4.1 Kinetics Assay Protocol

The kinetics assay determines the incubation time for each normal human serum (NHS) lot used for the enzyme-linked immunosorbent assay (ELISA). Since normal human serum used is pooled (serum from several people is combined), the incubation time might be slightly different from lot to lot. Incubation time should be determined every time a new lot of NHS arrives, before performing any ELISA assays.

Safety Notes: Always use gloves to protect yourself and the samples. Make sure to label everything to avoid mistakes.

1. Determine the incubation time intervals to be tested and its distribution in a 96-well plate.
 - a. Example in Figure A.11
2. Materials list and solution formulas are shown in Tables A.3-A.4
3. **Notes on the Plate Reader:** It may have to be turned on at least 15min before its use.
 - a. Software: SoftMax Pro
 - i. Plate reader test its UV-Vis every time its turn on for the first time. It might take several minutes for this to end.
 - ii. Open SoftMax Pro, at the left hand side it should show that is connected with the plate reader (a plate reader icon) and that it has a temperature reading.
 1. If the plate reader icon has a “?” in red, the software is not connected. Try restarting the computer; if the problem doesn’t end restart the plate reader.
 - iii. Create a template that is setup to mix the plate for 10secs then read the whole plate at 415nm.
 1. Save it in Desktop, folder Plate Reader Users, folder Your_Name, file Template.
4. Coat wells with 100 μ L of 40 μ g LPS/mL PBS (4 μ g/well), pH 7.4 at room temperature and leave overnight at 4°C (16-18hrs).

- a. New LPS bottle: dissolve in 1ml of PBS and filter (1mg/mL). Add 440μL LPS in 10.56mL PBS.
5. Discard LPS and wash once with 200μL of PBS-Tween.
6. Add 300μL of 5% skim milk in each well (2q in 40mL PBS)
 - a. Incubate at room temperature for 1hr.
7. Discard skim milk solution and wash twice with 200μL of PBS-Tween
8. Diluent Solution: 17.1mL GVB + 900μL 0.1m MgEGTA (final concentration 5mM).
 - a. Add 92 μL of diluents in each well.
9. NHS → centrifuge 2min @ 12000rpm
 - a. Add 8μL of NHS to each well. (~770μL/plate)
10. Incubate plate at room temperature for pre-determined time.
11. Wash the wells four times with 200μL of PBS-Tween.
12. Add 100μL of 1:1000 diluted anti-Human C3-HRP conjugated (3μL in 3mL PBS-Tween per two rows) to each well and incubate at room temperature for 1hr.
13. Wash the wells four times with 200μL of PBS-Tween.
14. Add 100μL of ABTS to each well.
 - a. ABTS 10mM stock → 8.23mg in 1.5mL H₂O (make daily)
 - b. ABTS solution → 1.2mL stock 6mL citric acid (0.2M stock), 4.8mL H₂O
 - c. ~3mL ABTS solution per two rows.
 - d. Remember to add 1μL 30% H₂O₂ per mL of ABTS solution before use.
15. Wait **10min** after last pipetting and then **read absorbance at 415nm**.
 - a. Read at 20min and 30min.
16. Calculate and plot the mean absorbance for each incubation time. Select a time for incubation for the Abs in the linear portion of the curve (before it reaches a plateau).

A.4.2 Compstatin Concentration Calculations

This protocol details the steps to determine compstatin concentration in a sample (water or buffer), using absorbance. It should be noted that if the UV-Vis instrument being used does not have a reference button, absorbance of the buffer can be taken and subtracted from absorbance of the sample (which contains compstatin and water or buffer).

Safety Notes: Always use gloves to protect yourself and the samples. Make sure to label everything to avoid mistakes.

1. **Notes on the Plate Reader:** It may have to be turned on at least 15min before its use.
 - a. Use the cuvette bottom in the plate reader.
2. 1cm cuvette = b (Figure A.10)
3. $\lambda=280\text{nm}$
4. Buffer (solution use for peptide dilution) for reference absorbance (500 μL)
5. 495 μL buffer + 5 μL peptide (sample)
6. Dilution Factor (DF)=100
 - a. Since 5 μL of sample are diluted to make 500 μL of total volume
 - b. Any other dilution can be used, but be certain to apply the proper dilution factor when calculating concentration
7. Quantity of tryptophan = Trp
 - a. Trp $\epsilon = 5500\text{M}^{-1}\text{cm}^{-1}$
 - b. Trp-Methylated $\epsilon = 5470\text{M}^{-1}\text{cm}^{-1}$ (Mormostein et al 1987)
8. Beer-Lambert Law
 - a. Absorbance $\text{Abs} = \epsilon bc \rightarrow (\text{Abs} \times \text{DF}) \div [\epsilon (\text{Trp}) \times b] = c$
 - b. Example
 - i. absorbance 280nm = 0.138, using a cuvette = 1cm, with a dilution factor (DF) of 100 (500 μL of sample/5 μL of protein = 100), 2 tryptophans present = $5500\text{M}^{-1}\text{cm}^{-1} \times 2 = 11000\text{M}^{-1}\text{cm}^{-1}$
 $(0.138 \times 100) \div (11000\text{M}^{-1}\text{cm}^{-1} \times 1\text{cm})$
 $0.001255\text{M} = 1254.55\mu\text{M}$

A.4.3 ELISA Protocol

The Human C3b enzyme-linked immunosorbent assay (ELISA) protocol quantifies the amount of C3b generated to determine how much compstatin is needed to inhibit the cleavage of C3 by half. This protocol assumes that incubation time was previously determined using the Kinetics Assay Protocol.

Safety Notes: Always use gloves to protect yourself and the samples. Make sure to label everything to avoid mistakes.

1. Materials list and solution formulas are shown in Tables A.3-A.4
2. **Notes on the Plate Reader:** It may have to be turned on at least 15min before its use.
 - a. Software: SoftMax Pro
 - i. Plate reader tests its UV-Vis every time it is turned on for the first time. It might take several minutes for this to end.
 - ii. Open SoftMax Pro, at the left hand side it should show that is connected with the plate reader (a plate reader icon) and that it has a temperature reading.
 1. If the plate reader icon has a “?” in red, the software is not connected. Try restarting the computer; if the problem doesn’t end restart the plate reader.
 - iii. Template in Desktop, folder Plate Reader Users, folder Your_Name, file Template.
 1. The template should be setup to mix the plate for 10sec then read the whole plate at 415nm.
3. Coat wells with 100μL of 40μg LPS/mL PBS (4μg/well), pH 7.4 at room temperature and leave overnight at 4°C (16-18hrs).
 - a. New LPS bottle: dissolve in 1ml of PBS and filter (1mg/mL). Add 440μL LPS in 10.56mL PBS.
4. Discard LPS and wash once with 200μL of PBS-Tween.
5. Add 300μL of 5% skimmed milk in each well (2q in 40mL PBS)
 - a. Incubate at room temperature for 1hr.
6. Discard skimmed milk solution and wash twice with 200μL of PBS-Tween.
7. Human Serum → centrifuge 2min @ 12000rpm
 - a. Hi-NHS → 30min at 60°C H₂O bath → around 80μL per plate
8. Prepare **dilution plate** (Figure A.12) in a clean 96-well plate, while incubating.
 - a. *Diluent*: Solution of GVB and MgEGTA
 - i. 17.1mL GVB + 900μL 0.1m MgEGTA (final concentration of 5mM)
 - b. Add 125μL of diluent to columns 1-2 (Figure A.12)
 - c. Prepare 1.5mL of 128mM compstatin solution (or desire starting concentration) in a tube and place 250μL in wells 3A-7A (Figure A.12)
 - d. Add 125μL of diluent to all other wells (Figure A.12 rows B-H and wells 8A-12A)
 - e. Serial dilutions (mix the solutions every time before transferring to next wells, Figure A.12)
 - i. Transfer 125μL from wells 3A-7A to wells 3B-7B and so on.
 - ii. When wells 3H-7H are reached, transfer 125μL to wells 8A-12A.
 - iii. Transfer 125μL from wells 8A-12A to wells 8B-12B and so on.
 - iv. When wells 8H-12H are reached, discard 125μL.
9. Place the ELISA 96-well plate on ice if the dilution plate is not ready.

10. Add 92 μ L of the compstatin dilutions in respective ELISA plate wells (Figure A.12)
11. Add 8 μ L of NHS and Hi-NHS to respective ELISA plate wells.
12. Incubate plate at room temperature for predetermined time (based on kinetics).
13. Wash the wells four times with 200 μ L of PBS-Tween.
14. Add 100 μ L of 1:1000 diluted Anti-Human C3 HRP conjugated (10 μ L in 10mL PBS-Tween) to each well and incubate at room temperature for 1hr
15. Wash the wells four times with 200 μ L of PBS-Tween.
16. Add 100 μ L of ABTS to each well.
 - a. ABTS 10mM stock $\rightarrow$ 8.23mg in 1.5mL H₂O (make daily)
 - b. ABTS Solution $\rightarrow$ 1.1mL of stock, 5.5mL citric acid (0.2M stock), 4.4mL H₂O
 - c. Remember to add 1 μ L 30% H₂O₂ per mL of ABTS solution before use.
17. Wait **10min** after last pipetting and then **read Abs at 415nm**.
 - a. Read after 20min and 30min.
18. Plot and analyze the data (normally using reading at 10min only)

A.4.4 GraphPad Prism Protocol

GraphPad Prism is the software used to analyze all of the ELISA data. The software is typically available in the computer that controls the plate reader.

1. Open data from plate reader in Excel and copy. Open new Excel file, paste special, select transpose.
2. Open Prism
 - a. New Table & Graph (left) $\rightarrow$ Select XY
 - b. Sample data $\rightarrow$ Select Start with an empty data table
 - c. Choose a Graph $\rightarrow$ Select Point Only
 - d. Subcolumns for Replicates or Error Values $\rightarrow$ Select Y, Select enter **##** replicate values in side-by-side subcolumns, and **plot mean and error** and **SD**.
3. Copy the data (already transposed) from Excel and paste it in Prism.
4. Label the concentrations in the Prism Data Table.
5. In Analysis (top)
 - a. Select Analyze $\rightarrow$ XY analyses $\rightarrow$ Row means/totals with SD (will appear in next window).
 - b. Select Analyze $\rightarrow$ Transform, Normalize... $\rightarrow$ Normalize

- i. How is 0% defined? → Y: input lowest mean data value
- ii. How 100% is defined → Y: input highest mean data value

$$\frac{y - y_{min}}{y_{max} - y_{min}} * 100$$

- iii. Present results as → Percentages
 - 1. This will give % C3b
- iv. Other 0% and 100% values can be used, like Hi-NHS and NHS.
- c. Select Analyze → Transform, Normalize... → Transform
 - i. Select Transform X values using X=Log(X)
- d. Select Analyze → XY analyses → Nonlinear regression (curve fit)
 - i. Choose an equation → Dose-response-Inhibitor → log (inhibitor) vs normalized response – variable slope

$$y = \frac{100}{1 + 10^{[(\log IC_{50} - x) Hill Slope]}}$$

- ii. Fitting method → least squares
 - 1. This will give IC50
6. In Change (top)
- a. Format axes
 - i. X Axis
 - 1. Scale → linear
 - 2. Regularly spaced ticks → Minor ticks → 9, log
 - 3. Regularly spaced ticks → Number Format → Antilog
 - ii. Left Y Axis → Unchecked Automatically determine the range and interval
 - 1. Range, adjust range based on data
 - 2. Regularly spaced ticks → Minor ticks → 2
 - iii. If a tick mark is not present (but needed), just add it manually.
7. In each graph (Graph folder at the left), fix the title of each axis.
8. Double click on each sheet will allow renaming it.
9. Save project → yourname_inhibitor_date
10. Print project → bottom left, PDF → Save as PDF → yourname_inhibitor_date

A.5 VMD/APBS Tutorial

A. Abbreviations

- 1. VMD: Visual Molecular Dynamics
- 2. APBS → Adaptive Poisson-Boltzmann Solver

B. What you need before starting...

1. The PDB file with the molecule of interest
2. VMD installed
 - a. Download at <http://www.ks.uiuc.edu/Research/vmd>
 - b. Available free, but registration is required.
 - c. Make sure it's installed in the C:\drive directly
3. APBS installed
 - a. Download at <http://www.poissonboltzmann.org/apbs/downloads>
 - b. Available free, but registration is required.
 - c. Make sure it's installed in the C:\drive directly

C. Getting a PQR File

1. PQR files are PDB files where the occupancy and B-factor columns (last column) have been replaced by per-atom charge ("Q") and radius (R") information.
2. In the web browser go to <http://kryptonite.nbcr.net/pdb2pqr/>
 - a. Upload your PDB file
 - b. Forcefield → PARSE
 - c. Keep default output naming scheme
 - d. Available options:
 1. Unselect → "Optimized the hydrogen bonding networks" and "Create an APBS input file"
 2. Depends on your needs, you may need to select "Use PROPKA" to assign protonation states at pH 7"
 - a. Adjust to desired pH.
 - e. Submit your job and wait for your results.
3. From the results, open the PQR file
 - a. Observed that in the first lines it mentions the force field used and the total charge of the molecule.
 - b. Save the file (.pqr)
4. The PQR file can be opened anytime as a text, just as any PDB file (use a simple text editor (i.e. Notepad, Wordpad) rather than Microsoft Word)

D. Electrostatic Calculations

1. Load the PQR file into VMD (File → New Molecule)
2. Select Extensions → Analysis → APBS Electrostatics
 - a. A new *APBS Tool* window will open
3. Select Edit (top) → Settings (in APBS Tool window)
 - a. Working Directory → Select where your PQR file is located
 - b. APBS Location → Select where your executable file is
4. Select Edit (bottom, in APBS Tool window)
 - a. A new *ELEC values* window will open
 - b. Make sure the Molecule selected is correct

- c. Boundary condition → Single ion molecule
- d. Mobile Ions. This select the ionic strength of solution
 - 1. If selected → Concentration 0.150 M and Radius 2.0 Å
 - 2. If not selected → no ionic strength present
- e. Dielectric constants
 - 1. Solute → 2.0
 - 2. Solvent → 78.54
- f. Keep the default for all other options.
- 5. Run APBS (locally)
- 6. Check the VMD Console window for information about the job while it runs.
- 7. A new window *APBSRun: Load APBS Map* will open when job is finished
 - a. Select → Load files into new molecule

Note: If after visualization the surfaces have a hole, the Mesh Lengths need to be adjusted. Determine which axis the hole is in and increase its mesh length. The mesh lengths have to be the same ones for each axis in both sections (Coarse Grid Options and Fine Grid Options). After estimating, re-run the electrostatic calculations until there is no hole present in the surfaces.

E. Electrostatic Visualization

- 1. This will give you a visual representation of the distribution of electrostatic potential within and surrounding the molecule.
- 2. Select Graphics → Representations
 - a. A new *Graphical Representations* window will open
- 3. Coloring Method → Color ID
 - a. Select → 0 blue (for positive)
- 4. Drawing Method → Isosurface
- 5. Isovalue → 1 (for blue)
- 6. Draw → Solid Surface
- 7. Select → Create New Representation
 - a. Repeat steps 2-6, change color to 1 red (negative) and isovalue to -1 (red).

Note: If the surfaces have a hole in them see the Note in Section D.

- 8. To save an image:
 - a. Select File → Render
 - b. Filename → Browse (where you want to save it and name it)
 - c. Select → Start Rendering
- 9. Your image can be rotated by writing “rotate by axis angle” in the VMD Console. This can be done for x,y,z in any angle.

Note: Do NOT use spaces in filenames. This often causes APBS and/or VMD to not run properly. Make sure to get in the habit of using underscores rather than spaces.

A.6 Figures and Tables

Figure A.1: Electrostatic potential clustering of the HIV-1 subtypes, from the 1999 consensus sequences and structural template 2B4C. Electrostatic potential were calculated using ionic strength corresponding to 0 mM salt concentration. Isopotential contours are presented in 4 different corresponding to rotations in vertical axis. Isopotential contours are plotted at $\pm 1 k_B T/e$ with blue and red for positive and negative electrostatic potential, respectively. The net charge, global prevalence, coreceptor selectivity and geographical distribution are indicated in the figure for each subtype. N/A denotes that information was not available. The purple, green and orange rectangles highlight HIV-1 subtypes that have similar electrostatic potential and same charge. The red rectangles highlight sequences that belong to subtype D and the pink rectangle denotes sequences that belong to subtype B (including the sequences of the two crystallographic structures). Green circles in the branches of the dendrogram denote intersection points between net charges or infected population. The # refers to the global prevalence of subtype D, which include D1-D4. The ^ refers to the global prevalence of subtype K, which include K1-K3.

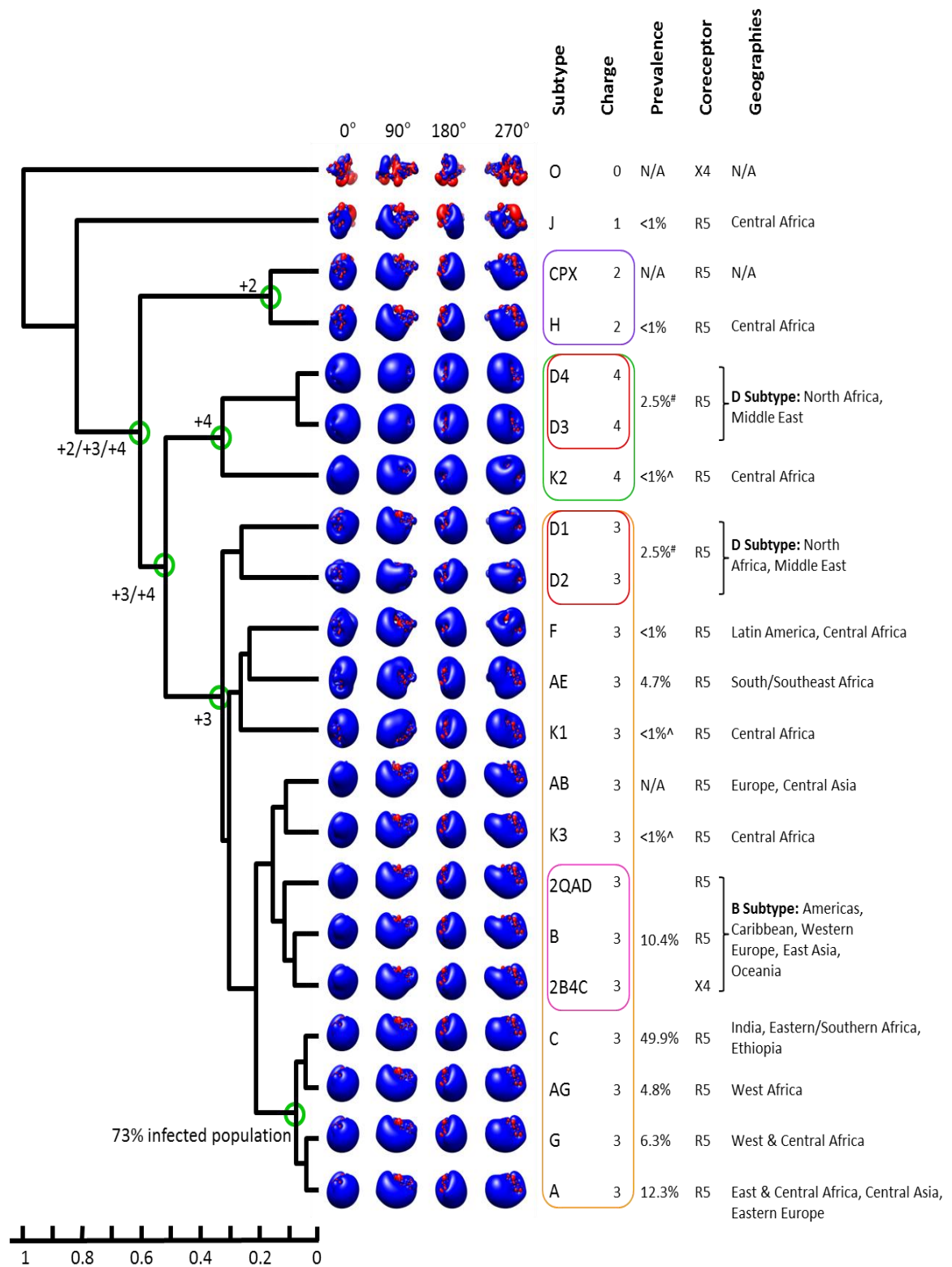


Figure A.2: Electrostatic potential clustering of the HIV-1 subtypes, from the 1999 consensus sequences and structural template 2QAD. Electrostatic potential were calculated using ionic strength corresponding to 0 mM salt concentration. Isopotential contours are presented in 4 different corresponding to rotations in vertical axis. Isopotential contours are plotted at $\pm 1 k_B T/e$ with blue and red for positive and negative electrostatic potential, respectively. The net charge, global prevalence, coreceptor selectivity and geographical distribution are indicated in the figure for each subtype. N/A denotes that information was not available. The purple, green, and orange rectangles highlight HIV-1 subtypes that have similar electrostatic potential and same charge. The red rectangles highlight sequences that belong to subtype D. The pink rectangle denotes sequences that belong to subtype B (including the sequences of the two crystallographic structures). Green circles in the branches of the dendrogram denote intersection points between charges or infected population. The # refers to the global prevalence of subtype D, which include D1-D4. The ^ refers to the global prevalence of subtype K, which include K1-K3.

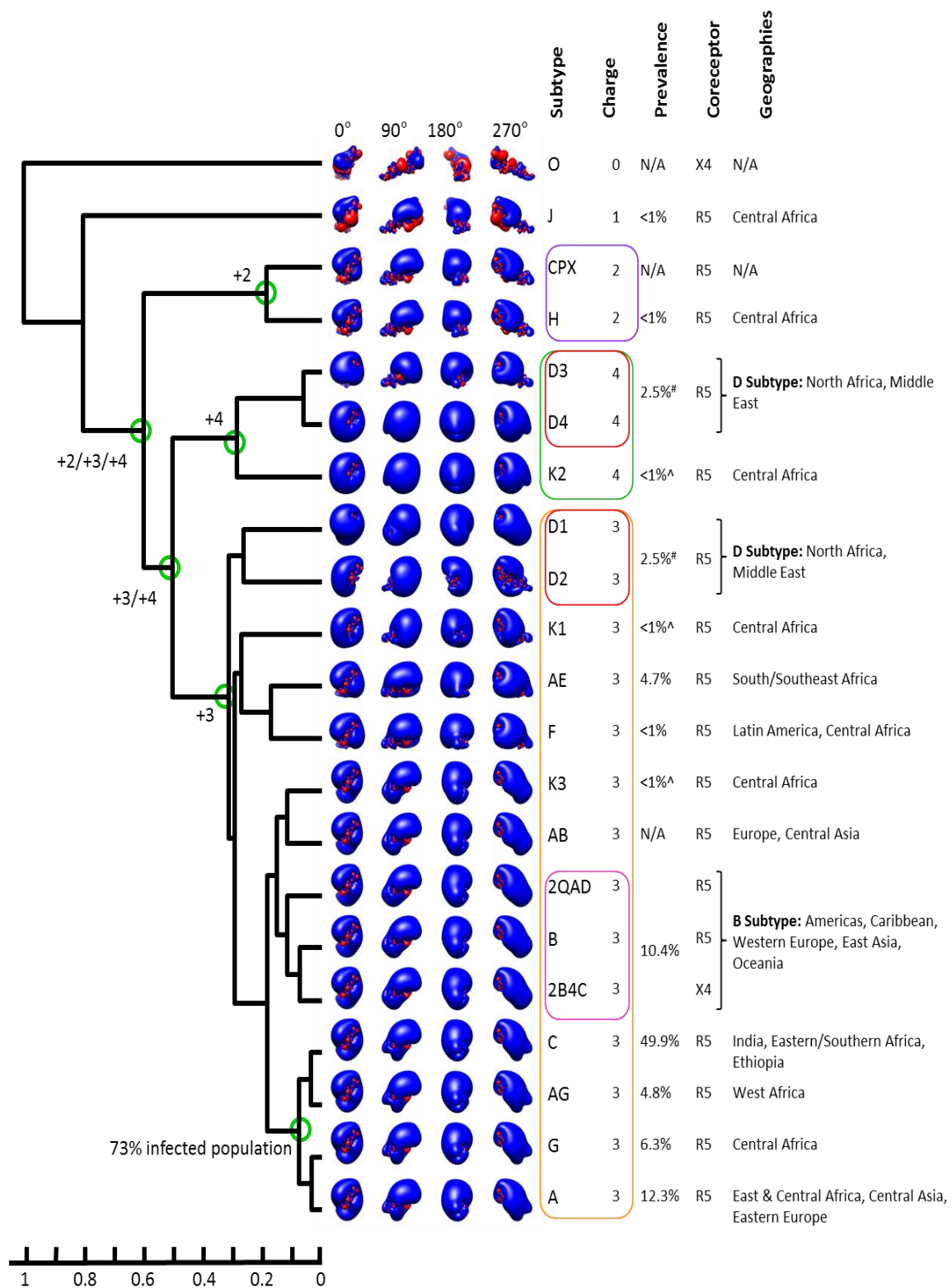


Figure A.3: Electrostatic potential clustering of the HIV-1 subtypes, from the 1999 consensus sequences and structural template 2B4C. Electrostatic potential were calculated using ionic strength corresponding to 150 mM salt concentration. Isopotential contours are presented in 4 different corresponding to rotations in vertical axis. Isopotential contours are plotted at $\pm 1 k_B T/e$ with blue and red for positive and negative electrostatic potential, respectively. The purple and orange rectangles highlight HIV-1 subtypes that have similar electrostatic potential and same charge. The red rectangles highlight sequences that belong to subtype D, while the pink rectangle denotes sequences that belong to subtype B (including the sequences of the two crystallographic structures).

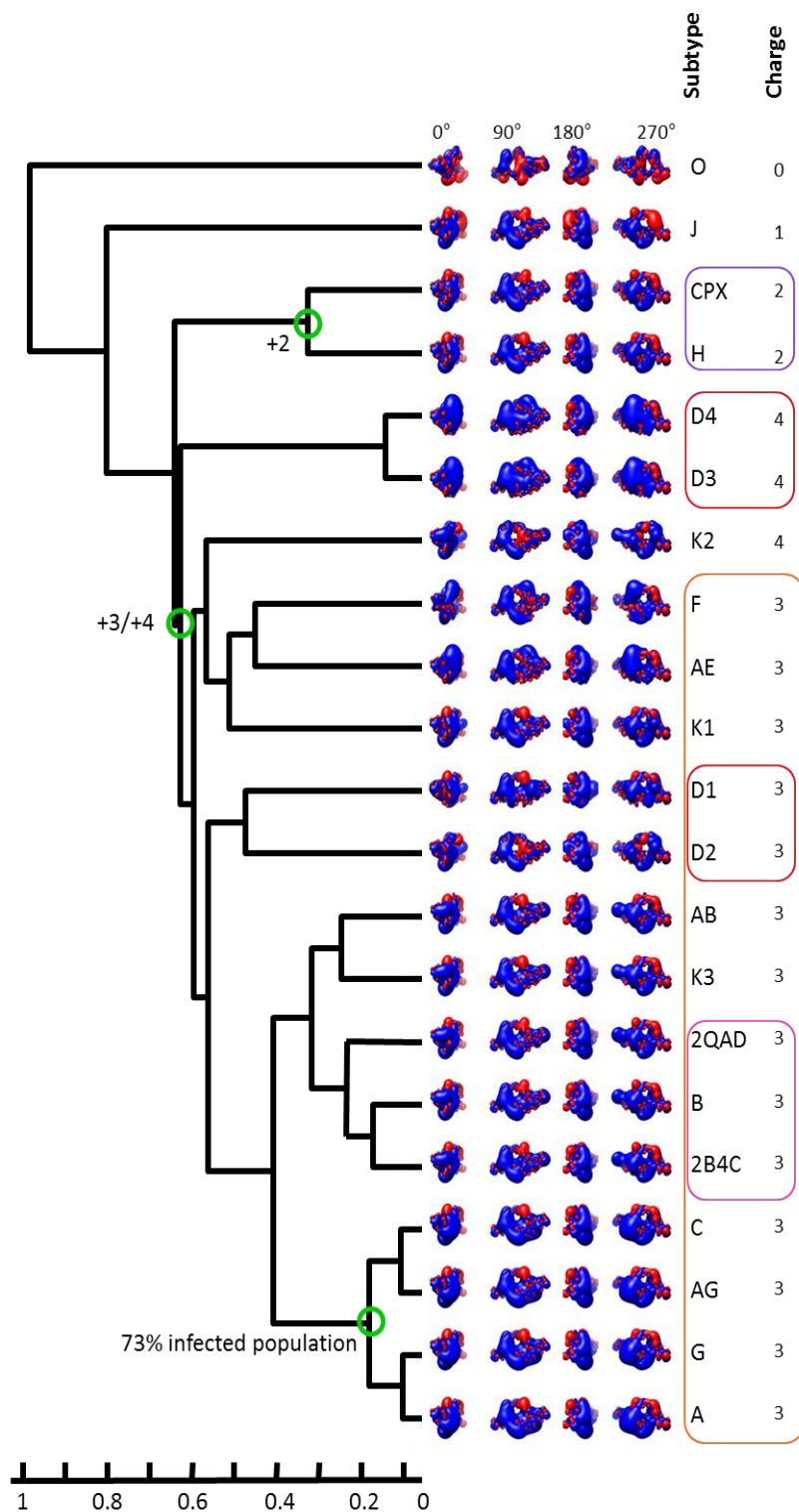


Figure A.4: Electrostatic potential clustering of the HIV-1 subtypes, from the 1999 consensus sequences and structural template 2QAD. Electrostatic potentials were calculated using ionic strength corresponding to 150 mM salt concentration. Isopotential contours are presented in 4 different orientations, corresponding to rotations about the vertical axis. Isopotential contours are plotted at $\pm 1 k_B T/e$, with blue and red corresponding to positive and negative electrostatic potentials, respectively. The purple, green, and orange rectangles highlight clusters with HIV-1 subtypes that have similar electrostatic potential and same charge. The red rectangles highlight sequences that belong to subtype D. The pink rectangle denotes sequences that belong to subtype B (including the sequences of the two crystallographic structures). Green circles in the branches of the dendrogram denote intersection points between net charges or infected population.

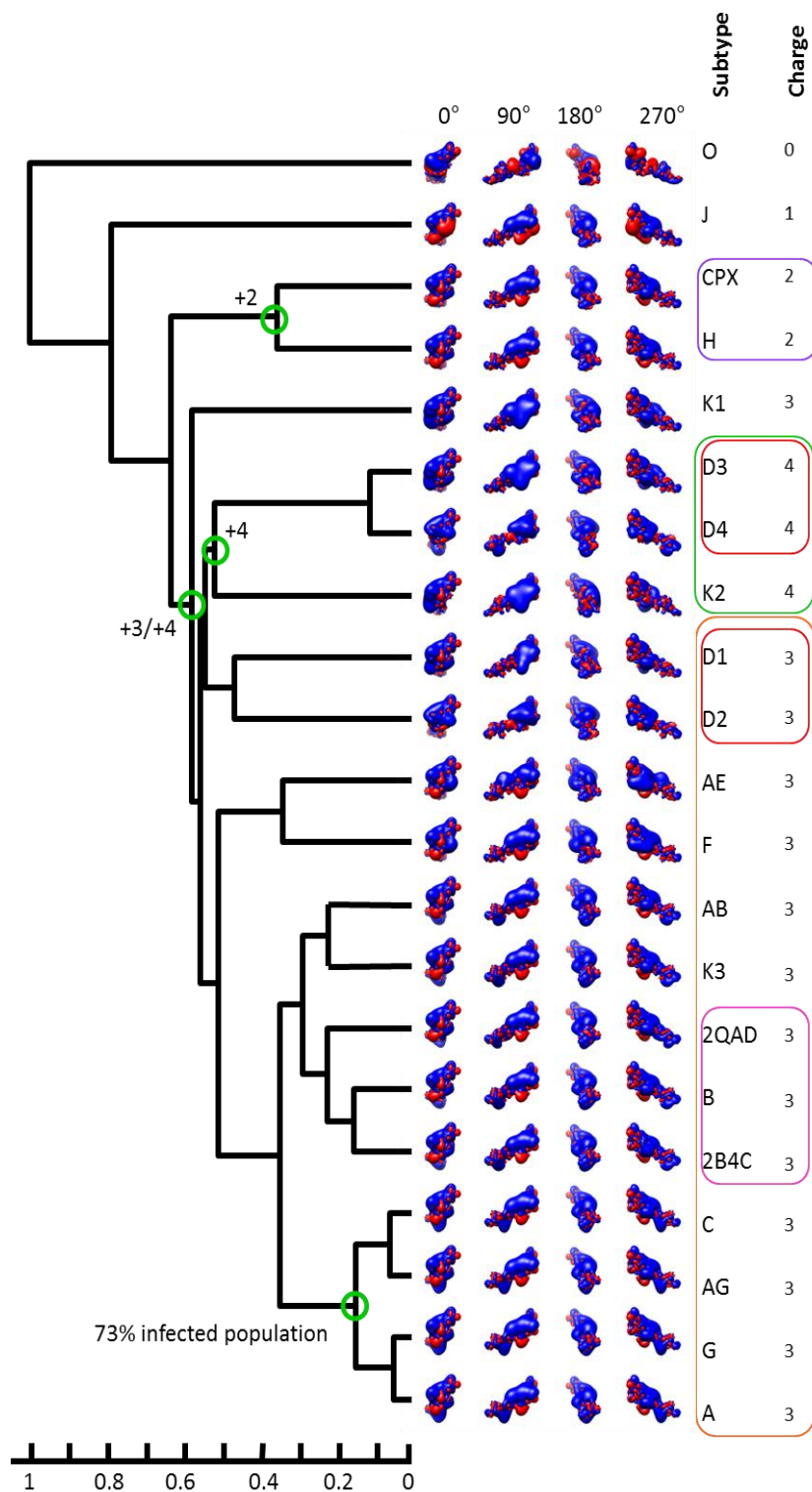


Figure A.5: Charge distribution clustering of the HIV-1 subtypes, from the 1999 consensus sequences and structural template 2B4C. The net charge, global prevalence, coreceptor use and geographical distribution are indicated in the figure for each subtype. N/A denotes that information was not available. Green circles in the branches of the dendrogram denote intersection points between net charges or infected population. The * refers to the global prevalence of subtype B, which include the two structural templates (2B4C and 2QAD). The # refers to the global prevalence of subtype D, which include D1-D4. The ^ refers to the global prevalence of subtype K, which include K1-K3.

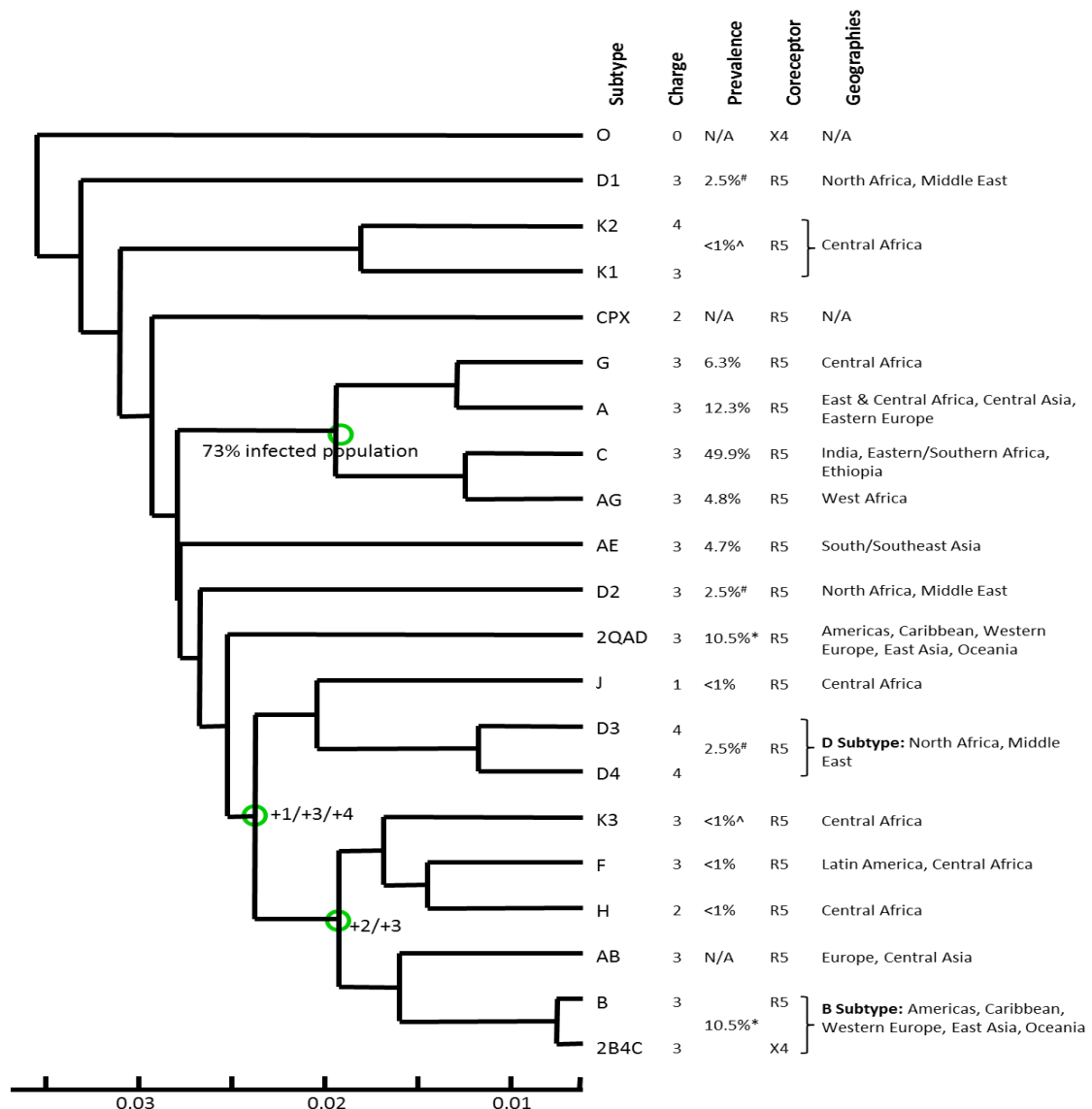


Figure A.6: Charge distribution clustering of the HIV-1 subtypes, from the 1999 consensus sequences and structural template 2QAD. The net charge, global prevalence, coreceptor selectivity and geographical distribution are indicated in the figure for each subtype. N/A denotes that information was not available. Green circles in the branches of the dendrogram denote intersection points between net charges or infected population. The * refers to the global prevalence of subtype B, which include the two structural templates (2B4C and 2QAD). The # refers to the global prevalence of subtype D, which include D1-D4. The ^ refers to the global prevalence of subtype K, which include K1-K3.

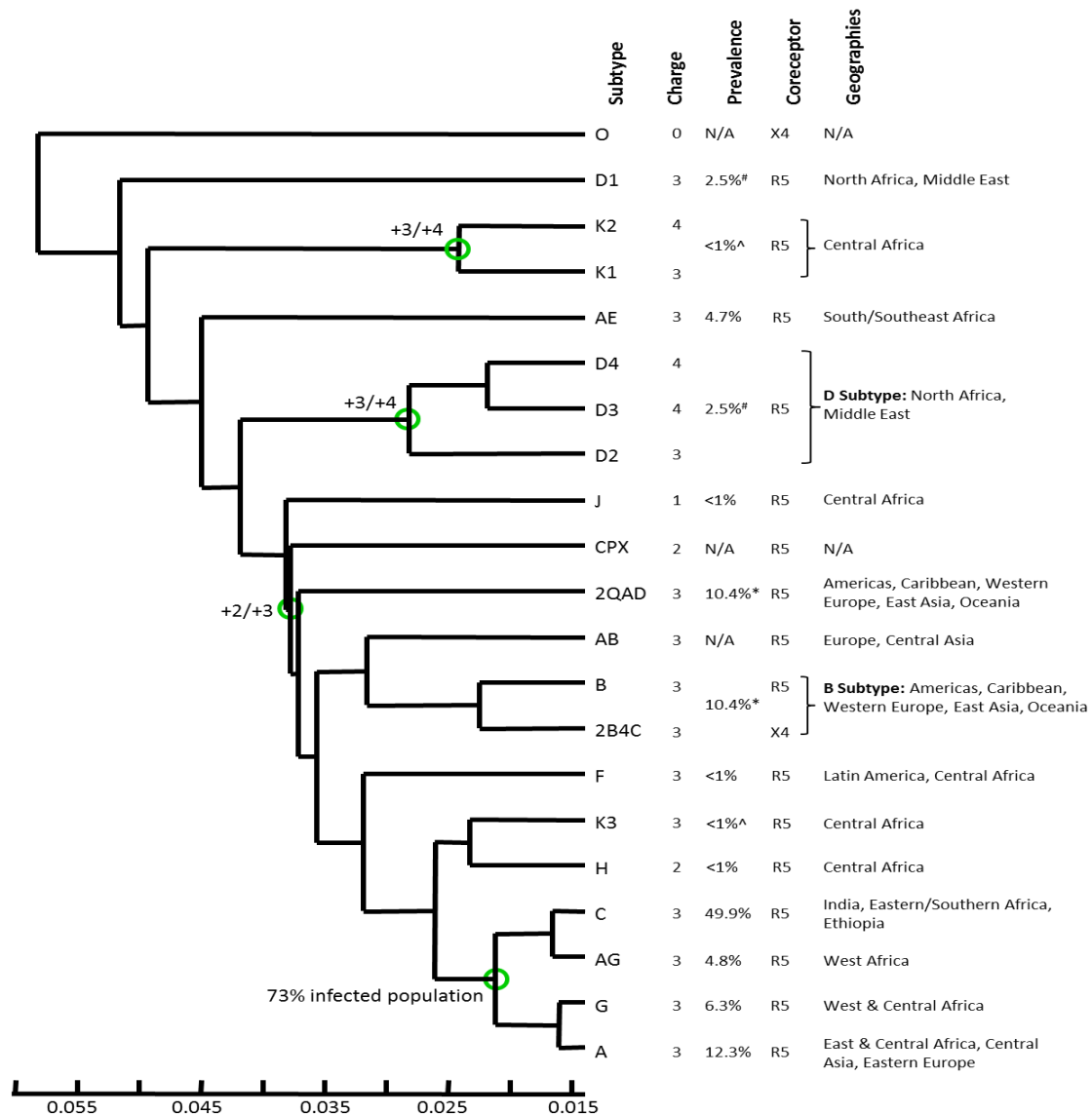


Figure A.7: Sequence clustering of the HIV-1 subtypes, from the 1999 consensus sequences. The global prevalence, coreceptor use and geographical distribution are indicated in the figure. N/A denotes that information was not available. The pink, blue, and orange rectangles highlight sequences that belong to subtype D, subtype B, and subtype K respectively. The ^ refers to the global prevalence of subtype K, which include K1-K3.

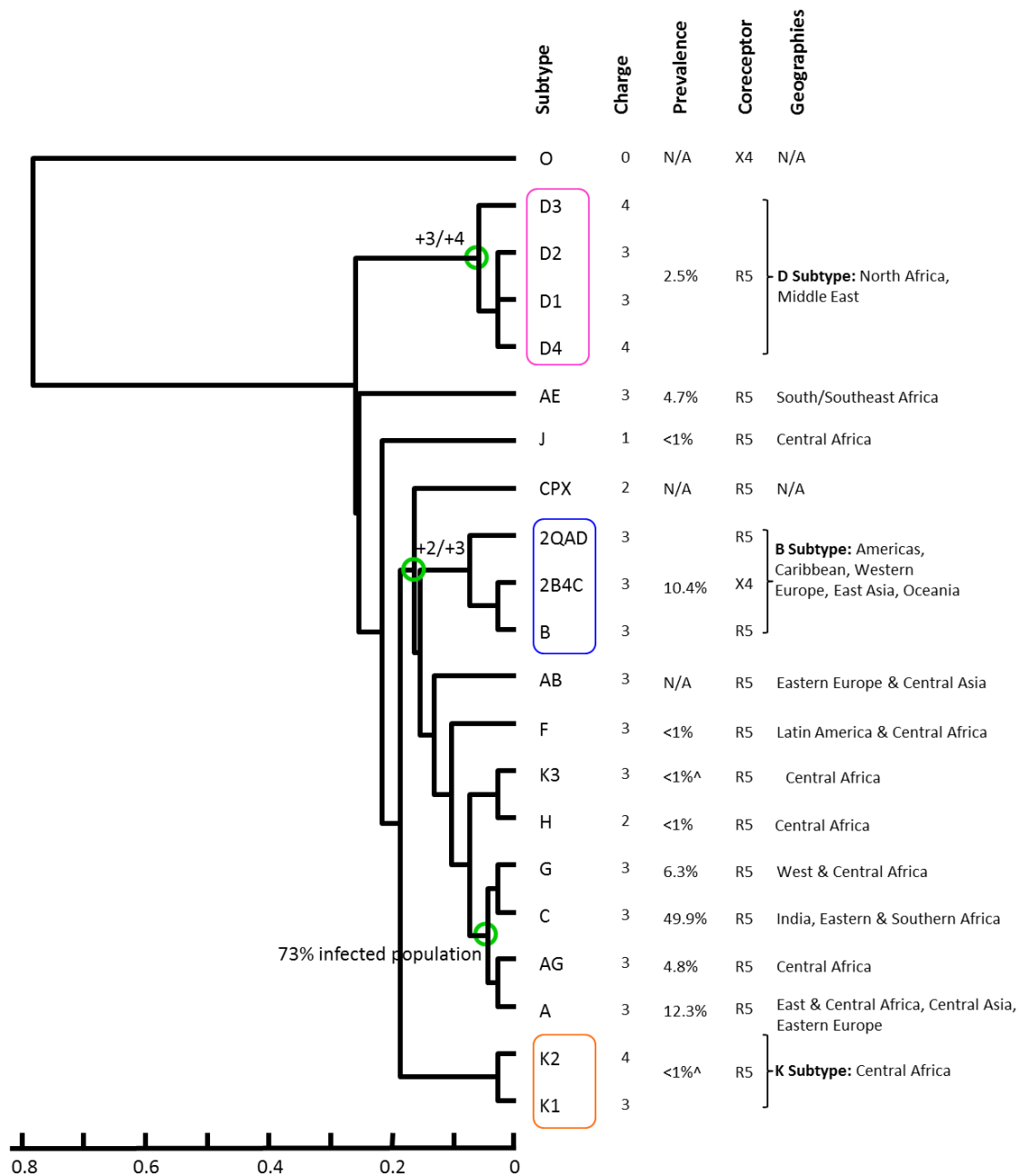
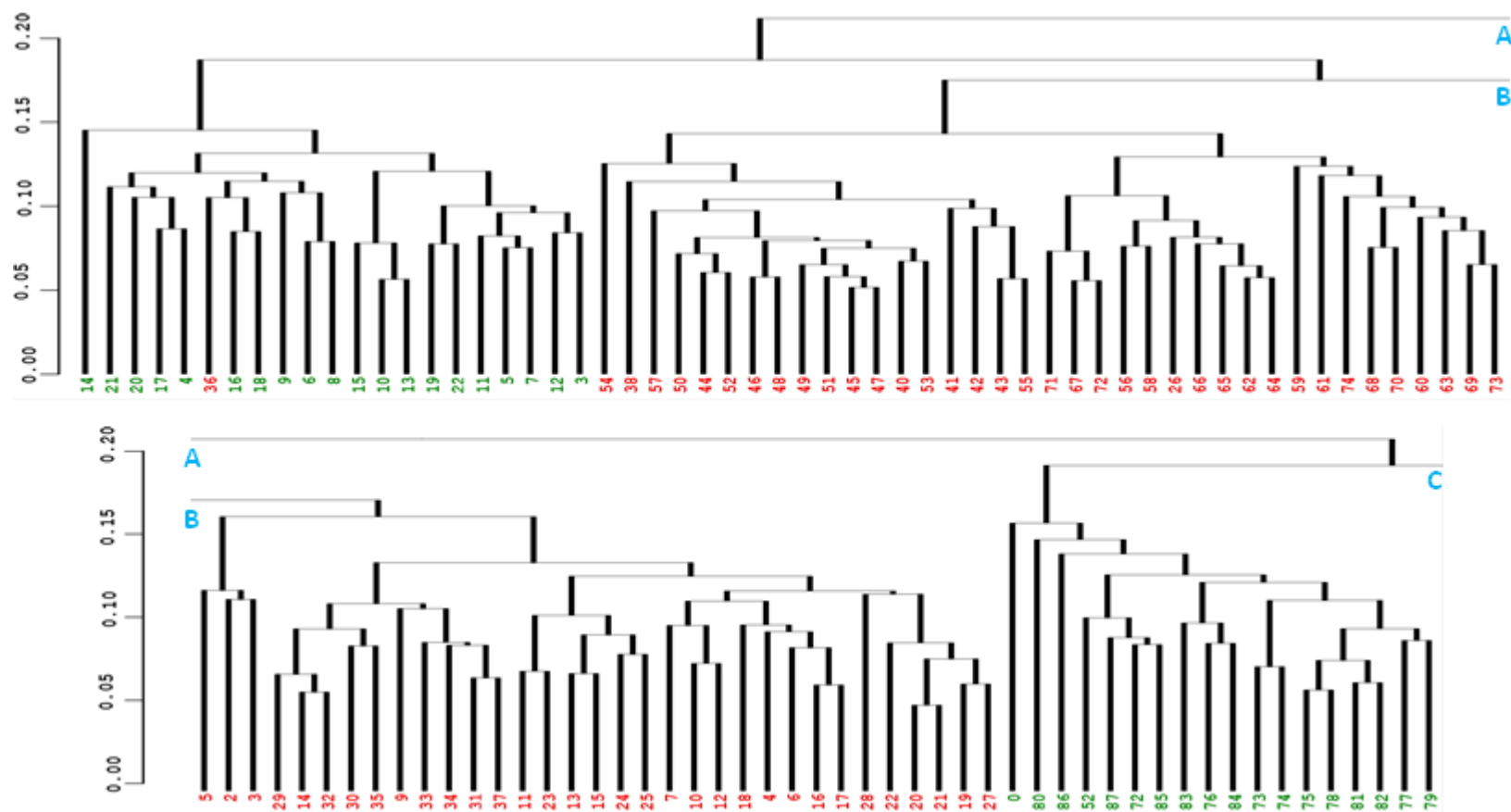


Figure A.8: Electrostatic clustering of snapshots of the MD trajectories of the V3 loop with ionic strength corresponding to 0 mM concentration. The vertical axis denotes the electrostatic similarity distances between snapshots. The horizontal axis denotes snapshot numbers corresponding to time in ns, with the red labels corresponding to the trajectory of 2B4C as starting structure and the green labels corresponding to the trajectory of 2QAD as starting structure. Blue label letters denote the connection between panels.



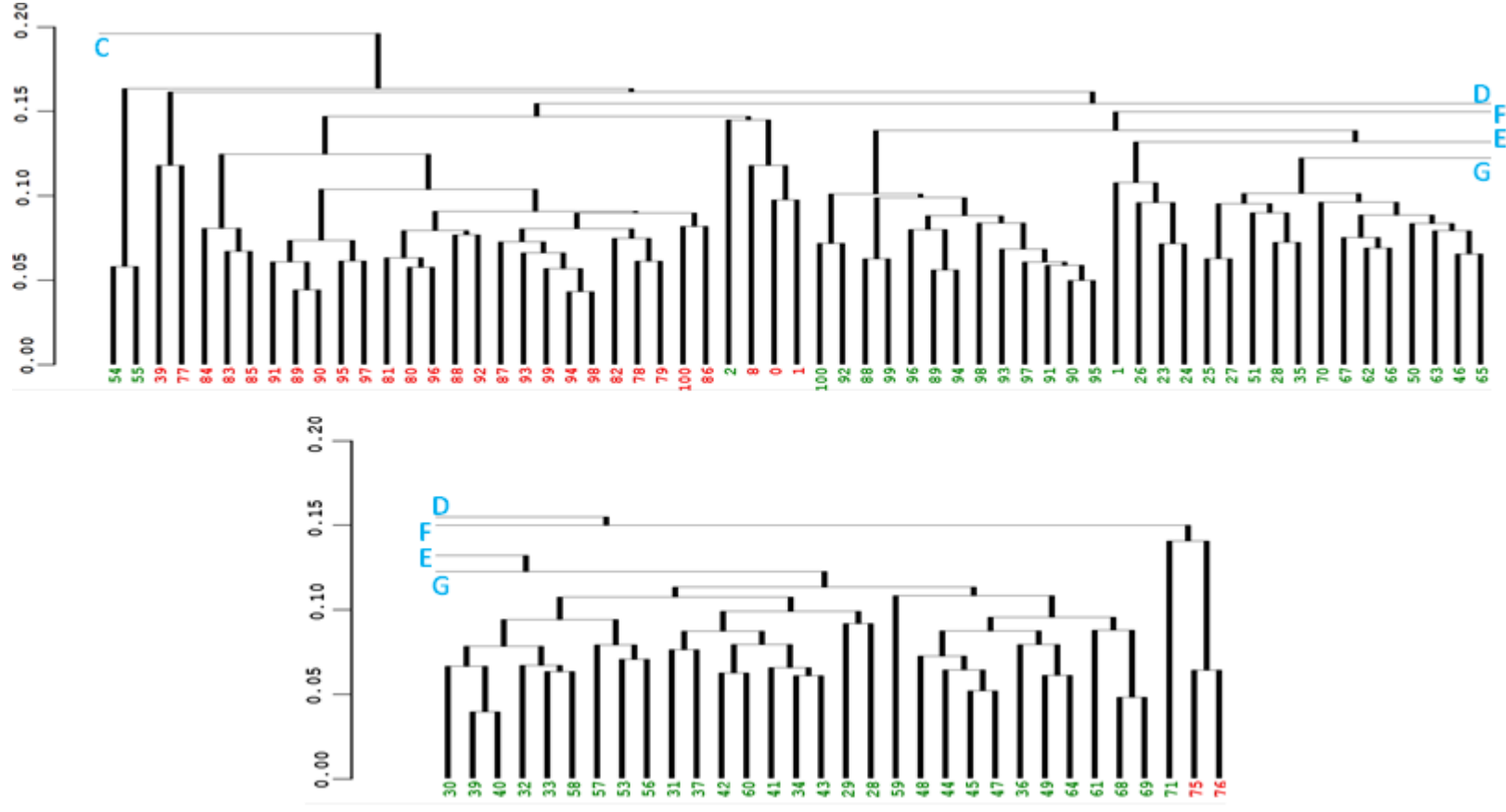
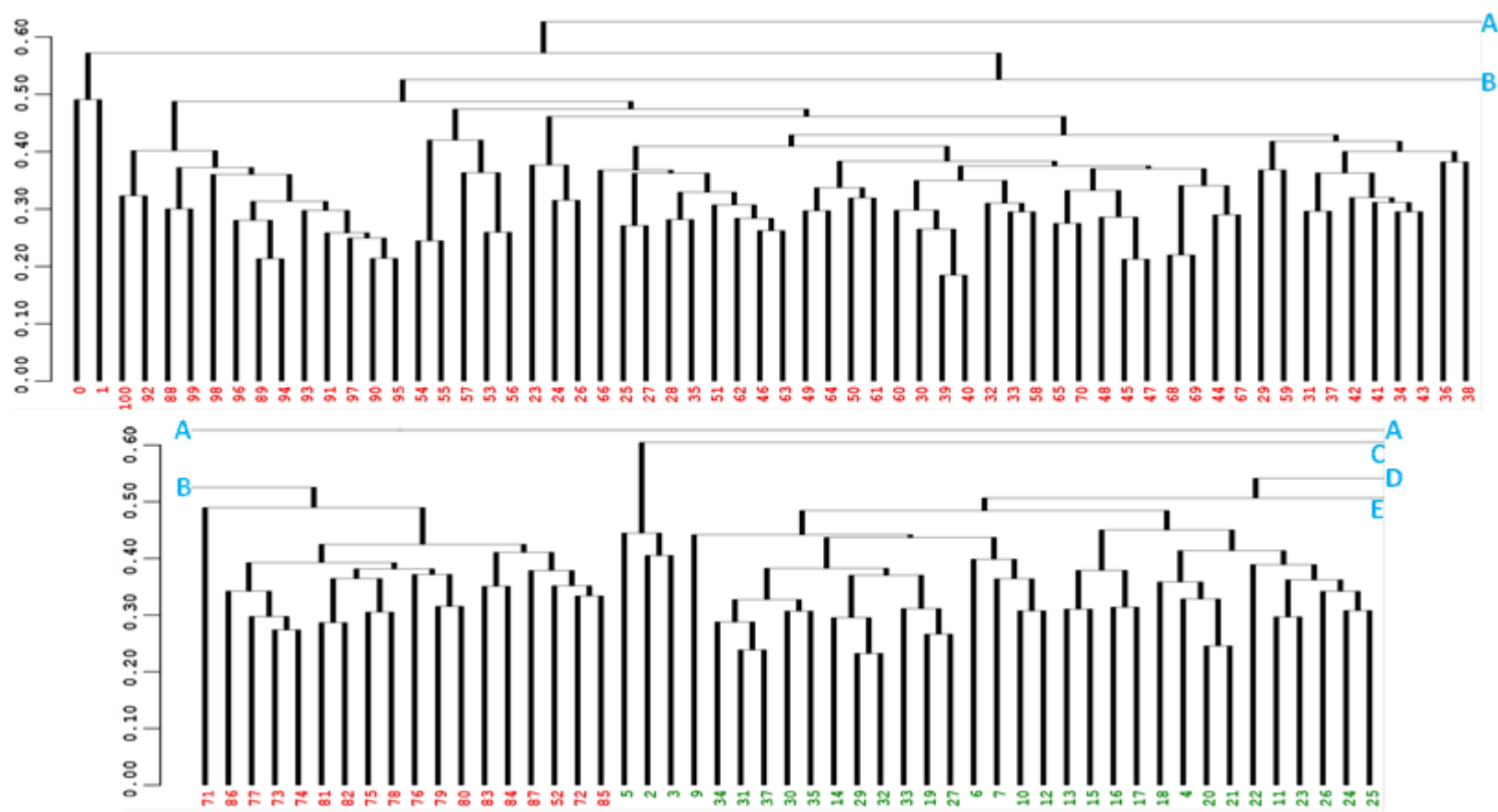


Figure A.9: Electrostatic clustering of snapshots of the MD trajectories of the V3 loop with ionic strength corresponding to 150 mM concentration. The vertical axis denotes the electrostatic similarity distances between snapshots. The horizontal axis denotes snapshot numbers corresponding to time in ns, with the red labels corresponding to the trajectory of 2B4C as starting structure and the green labels corresponding to the trajectory of 2QAD as starting structure. Blue label letters denote the connection between panels.



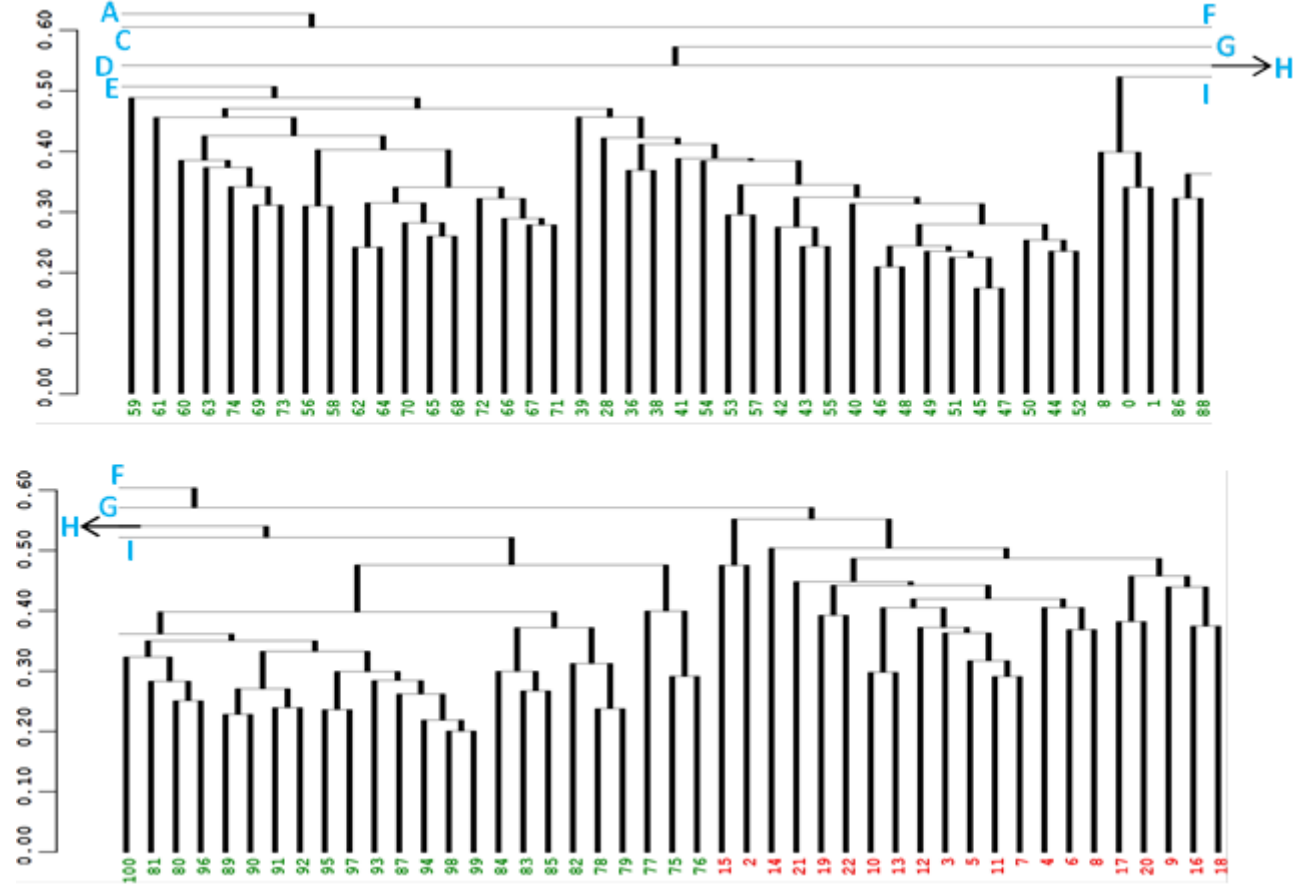


Figure A.10: Cuvette for SPR and ELISA. Micro self-masking 18B/Q/10 from Starna Cells, with a path length of 1cm (10mm).



Figure A.11: Kinetic assay example of times and distribution used in a 96-well plate.

	1	2	3	4	5	6	7	8	9	10	11	12	Incubation
A													15 mins
B													
C													30 mins
D													
E													60 mins
F													
G													120 mins
H													

Figure A.12: Dilution plate and ELISA plate examples. The plates are color coded, so samples in column 3 and colored beige in the dilution plate will go to column 3 in the ELISA plate.

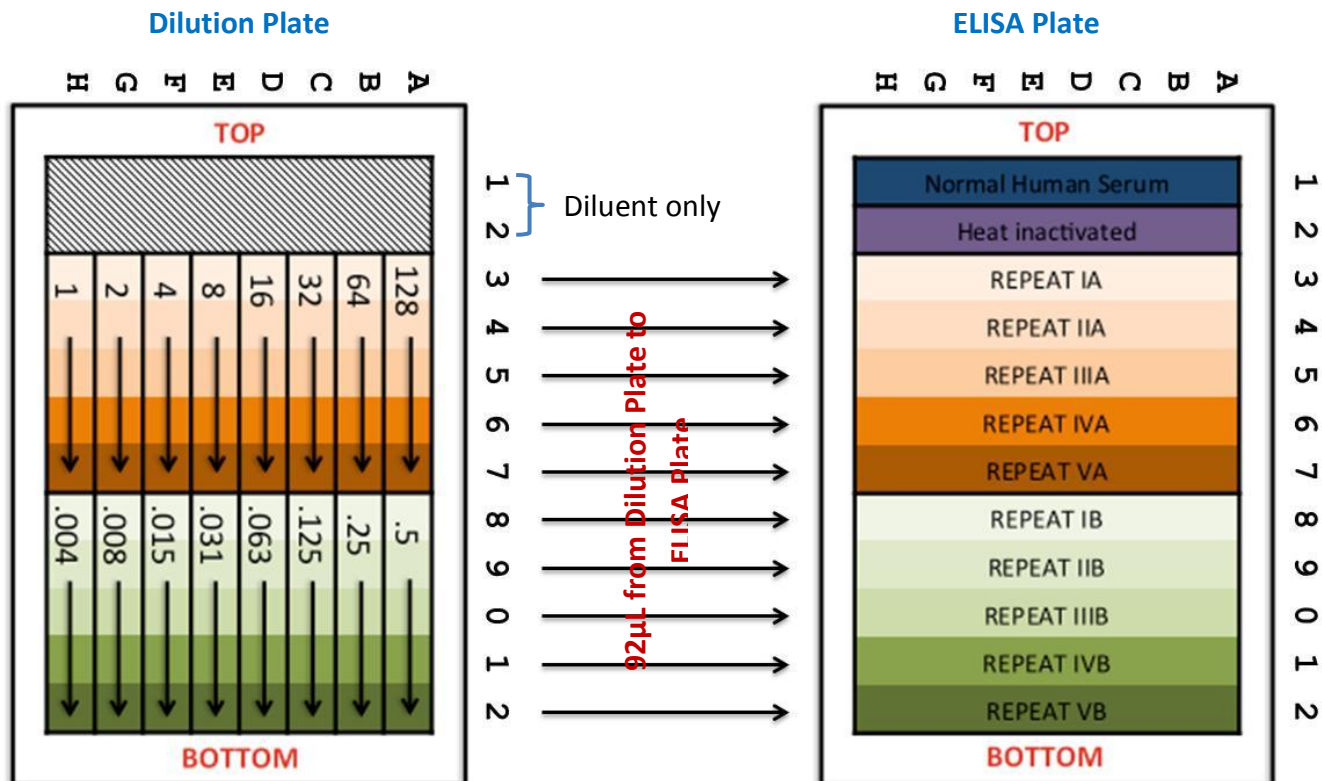


Table A.1: Alignment of HIV-1 V3 loop consensus sequences from 1999. Created using ClustalW2 (16). V3 loop subtype consensus sequences were obtained from Los Alamos National Laboratory (<http://www.hiv.lanl.gov>), except sequences 2B4C and 2QAD, which are the crystal structures. Subtypes AB, AE, AG, and CPX are circulating recombinant forms (CRFs). Gaps are introduced by the alignment.

Name	Sequences Alignment	Modifications from Original ^a
	01234567890123456789012345678901234567	
2B4C	CTRPNQN-TRKSIHIGPGRAFYTGTGEIIGD-IRQ-AHC	
2QAD	CTRPNNN-TRKSINIGPGRALYTGTGEIIGD-IRQ-AHC	
A	CTRPNNN-TRKSVRIGPGQAFYATGDIIGD-IRQ-AHC	
AB	CTRPNNN-TRKGIHIGPGRAFYATGDIIGD-IRQ-AHC	
AE	CTRPSNN-TRTSITIGPGQVFYRTGDIIGD-IRK-AYC	
AG	CTRPNNN-TRKSVRIGPGQTFYATGDIIGD-IRQ-AHC	
B	CTRPNNN-TRKSIHIGPGRAFYTGTGEIIGD-IRQ-AHC	
C	CTRPNNN-TRKSIRIGPGQTFYATGDIIGD-IRQ-AHC	
CPX	CTRPNNN-TRKSVHIGPGHTWYATGEIIGD-IRQ-AHC	
D1	CTRPYNNITRQSTHIGPGQALYTT-KIIGD-IRQ-AHC	eliminated gaps: 2, 3, 4
D2	CTRPYNN-TRQSTHIGPGQALYTTTKIIGD-IRQ-AHC	eliminated gaps: 1, 3, 4
D3	CTRPYNN-TRQSTHIGPGQALYTTTRRIIGD-IRQ-AHC	eliminated gaps: 1, 2, 4
D4	CTRPYNN-TRQSTHIGPGQALYTTKKIIGD-IRQ-AHC	eliminated gaps: 1, 2, 3
F	CTRPNNN-TRKSIHLGPGQAFYATGDIIGD-IRK-AHC	
G	CTRPNNN-TRKSIRIGPGQAFYATGDIIGD-IRQ-AHC	eliminated gaps: 1, 2
H	CTRPNNN-TRKSIHIGPGQAFYATGDIIGD-IRQ-AHC	eliminated gaps: 1, 2
J	CTRPNNN-TRKGIHMGPGQVLYATGEIIGD-IRE-AHC	

K1	CTRPNNN-TRKSIHIGPG-AFYATGDIIGD-IRQKAHC	eliminated: 18K _{40%}
K2	CTRPNNN-TRKSIHIGPG-KFYATGDIIGD-IRQKAHC	eliminated: 19A _{40%}
K3	CTRPNNN-TRKSIHIGPGKAFYATGDIIGD-IRQ-AHC	eliminated: 33K _{40%}
O	CERPGNT--VQEIRIGP-MAWYSMGLEGGNNTSRAAYC	eliminated 4 gaps, 7N _{26%} and 29N _{26%}

^a Modifications were performed to some consensus sequences with more/less than the 35 amino acids, as described in A.1.1 Consensus Sequences. Percent subscripts denote amino acid frequency.

Table A.2: SPR materials.

Name	Location/Supplier
Biacore X100	Genomic Center, Keen Hall
Sensorship SA	GE Healthcare #BR100032
C3	Complement Technology #A113
PBS-T	Invitrogen #10010, add 0.05% Tween 20 Individual Solutions: 1.00mM KH ₂ PO ₄ , 155mM NaCl, 3mM Na ₂ HPO ₄ , 0.05% Tween 20
Regeneration Solution	10mM NaOH, 0.05% SDS
Conditioning Solution	1M NaCl, 50mM NaOH
Amicon-Ultra 0.5mL	Millipore # UFC510024, cutoff 100kDa (through Fisher)

Table A.3: Human C3b assays materials.

Name	Company/Supplier	Item Number	Quantity
Lipopolysaccharides (LPS)	Sigma-Aldrich	L6529-1MG	1mg
Normal Human Serum (NHS)	Complement Technology	NHS	1mL
Thermo Scientific Nunc Immuno 96 well plates	Fisher	12-565-136	60plates/case
Sodium Barbitol	Sigma-Aldrich	B0500-25G	25g
ABTS	Roche Diagnostics	10102946001	2g
Complement C3 anti-human HRP	MP Biomedicals (through Fisher)	0855237	2mL
NaCl	Acros Organics (through Fisher)	AC207790010	1kg
Hydrogen Peroxide 30%	Fisher	H325-100	100mL
Na ₂ HPO ₄	Fisher	BP332-500	500g
NaH ₂ PO ₄	Fisher	BP329-500	500g
Citric Acid	Fisher	BP339-500	500g
Sodium azide	Sigma-Aldrich	S2002-25G	25g
Gelatin from bovine skin, type B, powder	Sigma-Aldrich	G9391-100G	100g
EGTA 0.5M pH 8.0	Fisher	NC9118216	100mL
Tween-20	Fisher	BP337-100	100mL
MgCl ₂	Acros Organics (through Fisher)	AC22321-1000	100g
Millex-GV Filter (for LPS)	Millipore (through Fisher)	SLGV004SL	100/pack
1mL syringes (for LPS)	BD Medical (through Fisher)	148232F	100/pack

Table A.4: Solution formulas and calculations for Human C3b assays (kinetics and ELISA).

<i>PBS/PBS-Tween Reagents @ pH 7.4</i>	
<i>Formula</i>	<i>Stocks Preparation</i>
6.7mM Na ₂ HPO ₄	20mM Na ₂ HPO ₄ → 1.416g in 500mL H ₂ O
3.3mM NaH ₂ PO ₄	10mM NaH ₂ PO ₄ → 0.600g in 500mL H ₂ O
145mM NaCl	500mM NaCl → 14.71g in 500mL H ₂ O
0.005% Tween	10% Tween → 5mL 100% Tween in 45mL H ₂ O
<i>PBS, Final Volume: 500mL</i>	<i>PBS-Tween, Final Volume: 1L</i>
20mM Na ₂ HPO ₄ → 167.5mL	20mM Na ₂ HPO ₄ → 335mL
10mM NaH ₂ PO ₄ → 165mL	10mM NaH ₂ PO ₄ → 330mL
500mM NaCl → 145mL	500mM NaCl → 290mL
Adjust pH 7.4	10% Tween → 500μL
22.5mL H ₂ O	Adjust pH 7.4
41mL/plate	44.5mL H ₂ O
	221mL/plate

<i>GVB Reagents @ pH 7.4</i>	
<i>Formula</i>	<i>Stocks Preparation</i>
3.2mM barbituric acid	10mM barbituric acid → 0.6405g in 500mL H ₂ O
1.8mM sodium barbital	10mM sodium barbital → 1.031g in 500mL H ₂ O
145mM NaCl	500mM NaCl → 14.71g in 500mL H ₂ O
0.1% gelatin	1% gelatin → 1g in 100mL H ₂ O (may need heating)
0.02% azide	1% sodium azide → 1g in 100mL H ₂ O
<i>GVB, Final Volume: 500mL</i>	
10mM barbituric acid → 160mL	
10mM sodium barbital → 90mL	
500mM NaCl → 145mL	
1% gelatin → 50mL	
1% sodium azide → 10mL	
Adjust pH	
45mL H ₂ O	
17.1mL/plate	

<i>MgEGTA @ pH 7.3</i>	
<i>Formula</i>	<i>Stocks Preparation</i>
0.1M MgCl ₂	0.2M MgCl ₂ → 0.9521g in 50mL H ₂ O
0.1M EGTA	0.5M EGTA (liquid)

MgEGTA, Final Volume: 50mL

0.2M $\text{MgCl}_2 \rightarrow 25\text{mL}$

0.5M EGTA $\rightarrow 10\text{mL}$

Adjust pH

15mL H_2O

900 μL /plate

Citric Acid @ pH 4.2

0.2M stock, Final Volume 500mL

19.213g in 400mL H_2O

Adjust pH 4.2

100mL H_2O

5.5mL/plate

A.7 References

1. HIV Molecular Immunology Database 1999. In: Korber B, Brander C, Haynes BF, Moore JP, Koup R, Walker BD, et al., editors. Los Alamos, New Mexico, USA: Los Alamos National Laboratory, Theoretical Biology and Biophysics; 1999. p. 602-607.
2. Vriend G. What If: A molecular modeling and drug design program. *J Mol Graphics* 1990, **8**: 52-56.
3. Huang CC, Tang M, Zhang MY, Majeed S, Montabana E, Stanfield RL, Dimitrov DS, Korber B, Sodroski J, Wilson IA, et al. Structure of a V3-containing HIV-1 gp120 core. *Science* 2005, **310**: 1025-1028.
4. Huang CC, Lam SN, Acharya P, Tang M, Xiang SH, Hussan SS, Stanfield RL, Robinson J, Sodroski J, Wilson IA, et al. Structures of the CCR5 N terminus and of a tyrosine-sulfated antibody with HIV-1 gp120 and CD4. *Science* 2007, **317**: 1930-1934.
5. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PDB Viewer: An environment for comparative protein modeling. *Electrophoresis* 1997, **18**: 2714-2723.
6. Gorham RD, Jr., Kieslich CA, Morikis D. Complement inhibition by *Staphylococcus aureus*: electrostatics of C3d-EfbC and C3d-Ehp association. *Cell Molec Bioeng* 2011, **5**: 32-43.
7. Gorham RD, Jr., Kieslich CA, Nichols A, Sausman NU, Foronda M, Morikis D. An evaluation of Poisson-Boltzmann electrostatic free energy calculations through comparison with experimental mutagenesis. *Biopolymers* 2011, **95**: 746-754.
8. Kieslich CA, Gorham RD, Jr., Morikis D. Is the rigid-body assumption reasonable? Insights into the effects of dynamics on the electrostatic analysis of barnase-barstar. *J Non-Cryst Solids* 2010, **357**: 707-716.
9. Kieslich CA, Yang J, Gunopulos D, Morikis D. Automated computational protocol for alanine scans and clustering of electrostatic potentials: application to C3d-CR2 association. *Biotechnol Prog* 2010, **27**: 316-325.
10. Gorham RD, Jr., Kieslich CA, Morikis D. Electrostatic clustering and free energy calculations provide a foundation for protein design and optimization. *Ann Biomed Eng* 2011, **39**: 1252-1263.

11. Kieslich CA, Vazquez H, Goodman GN, López de Victoria A, Morikis D. The effect of electrostatics on factor H function and related pathologies. *J Mol Graph Model* 2011, **29**: 1047-1055.
12. Dolinsky TJ, Czodrowski P, Li H, Nielsen JE, Jensen JH, Klebe G, Baker NA. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res* 2007, **35**: W522-525.
13. Sitkoff D, Sharp KA, Honig B. Accurate calculation of hydration free energies using macroscopic solvent models. *J Phys Chem* 1994, **98**: 1978-1988.
14. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* 2001, **98**: 10037-10041.
15. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF chimera - A visualization system for exploratory research and analysis. *J Comput Chem* 2004, **25**: 1605-1612.
16. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* 2007, **23**: 2947-2948.
17. Hemelaar J, Gouws E, Ghys PD, Osmanov S. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS* 2006, **20**: W13-23.
18. Cardozo T, Kimura T, Philpott S, Weiser B, Burger H, Zolla-Pazner S. Structural basis for coreceptor selectivity by the HIV type 1 V3 loop. *AIDS Res Hum Retrov* 2007, **23**: 415-426.
19. Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K. NAMD2: Greater scalability for parallel molecular dynamics. *J Comp Phys* 1999, **151**: 283-312.

ALIANA LÓPEZ DE VICTORIA SUÁREZ

alianaldv@gmail.com

EDUCATION

- Bioengineering PhD, University of California-Riverside (UCR) 2007-2012
 - **Adviser:** Dr. Dimitrios Morikis
 - **Thesis Title:** Studies of Protein Interactions and Knowledge-Based Drug Design: (A) The Electrostatic Nature of Recognition Between HIV-1 gp120 V3 Loop and Coreceptors CCR5/CXCR4, (B) Complement System Inhibition by Compstatin Family Peptides
- Industrial Biotechnology BS, University of Puerto Rico-Mayagüez Campus (UPRM) 2001-2007
 - *Cum Laude*

SELECTED AWARDS

- Dissertation Year Fellowship Award for the 2011-2012 academic year. Graduate Division at UCR.
- Graduate Research Mentorship Program Fellowship, September 2010-March 2011. Graduate Division at UCR.
- U.S. Department of Education Graduate Assistance in the Areas of National Need (GAANN) Fellowship. September 2009 – June 2010.

SELECTED PUBLICATIONS

- López de Victoria A, Forest DL, Kieslich CA, Banna CD, Tamamis P, Gorham RD Jr, Bellows-Peterson ML, Floudas CA, Archontis G, Johnson LV, et al. New potent complement inhibitors of the compstatin family peptides prevent complement-associated deposit formation in human retinal pigmented epithelial cell cultures. In Preparation, 2012
- López de Victoria A, Tamamis P, Kieslich CA, Morikis D. Insights into the Structure and Correlated Motions of the HIV-1 gp120 V3 loop using Molecular Dynamics Simulations. Submitted, 2012.
- Kieslich CA, Tamamis P, Gorham RD Jr, López de Victoria A, Sausman NU, Archontis G, Morikis D (2012) Exploring Protein-Protein and Protein-Ligand Interactions in the Immune System using Molecular Dynamics and Continuum Electrostatics. Current Physical Chemistry. In Press, available online
- Tamamis P, López de Victoria A, Gorham RD Jr, Pierou P, Bellows-Peterson ML, Floudas CA, Morikis D, Archontis G (2012) Molecular Dynamics in Drug Design: New Generations of Compstatin Analogs. Chemical Biology and Drug Design. **79**: 703-718.
- López de Victoria A, Kieslich CA, Rizo AK, Krambovitis E, Morikis D (2012) Clustering of HIV-1 Subtypes Based on gp120 V3 Loop Electrostatic Properties. BMC Biophysics, **5**:3.
- Kieslich CA, Vázquez H, Goodman GN, López de Victoria A, Morikis D (2011) The Effect of Electrostatics on Factor H Function and Related Pathologies. Journal Molecular Graphics and Modelling, **29**:1047-1055.

- López de Victoria A, Gorham RD Jr, Bellows-Peterson ML, Ling J, Lo DD, Floudas CA, Dimitrios Morikis (2011) A New Generation of Potent Complement Inhibitors of the Compstatin Family. *Chemical Biology and Drug Design*, **77**:431-440.
- Bellows ML, Fung HK, Taylor MS, Floudas CA, López de Victoria A, Morikis D (2010) New Compstatin Variants Through Novel De Novo Protein Design Frameworks. *Biophysical Journal*, **98**:2337-2356.

UNDERGRADUATE RESEARCH MENTORSHIP

- Foronda M, López de Victoria A, Morikis D (2010) Electrostatic Exploration of Membrane Cofactor Protein (MCP) using Computational Alanine Scans, Poisson-Boltzmann Electrostatics and Experimental Mutagenesis Data. *UCR Undergraduate Research Journal*, **4**:59.
- Vázquez H, Goodman GN, López de Victoria A, Kieslich CA, Morikis D (2009) The Role of Electrostatics in the Function of Factor H and its Relation to Complement System-Mediated Disease. *UCR Undergraduate Research Journal*, **3**: 49-56.

PRESENTATIONS AND POSTERS

- American Chemical Society Annual Meeting, San Diego, CA. March 2012. *Talk*: Role of gp120 V3 Loop Charge in HIV-1 Coreceptor Selectivity.
- Biophysical Society Annual Meeting, San Diego, CA. February 2012. *Poster*: Computational and Experimental Analysis of the Interactions between C3c and Compstatin Family Peptides.
- UC Systemwide Bioengineering Symposium, Santa Barbara, CA. June 2011. *Talk*: Iterative, Computational and Experimental Methods in Drug Design: The Paradigm of Compstatin Family Peptides.
- Inland Empire Technology Week, California State University, San Bernardino, CA. October 2010. *Poster*: Design of Potent Complement Inhibitors from the Compstatin Family.
- UC Systemwide Bioengineering Symposium, Davis, CA. June 2010. *Poster*: Design of Potent Complement Inhibitors from the Compstatin Family.
- American Association for the Advancement of Science (AAAS) Annual Meeting, San Diego, CA. February 2010. *Poster*: Clustering Sequences and Electrostatic Potentials of HIV-1 Subtypes.
- UC Systemwide Bioengineering Symposium, Merced, CA. June, 2009. *Talk*: Clustering of Sequences and Electrostatic Potentials of HIV-1 Subtypes.
- Biomedical Engineering Society (BMES) Annual Meeting, St Louis, MO. October 2008. *Poster*: Interactions of the V3-loop of HIV-1 gp120 and the N-terminal Domain of CCR5.
- UC Systemwide Bioengineering Symposium, Riverside, CA. June 2008. *Poster*: Interactions of the V3-loop of HIV-1 gp120 and the N-terminal Domain of CCR5.