UNIVERSITY OF CALIFORNIA, MERCED

Purification and Study of CC Chemokine-Based Strategies to Combat Chronic Inflammation and HIV

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Quantitative and Systems Biology

by Anna Faith Nguyen, B.S., University of California, Davis

Committee in charge:

Professor Michael Cleary, Chair of Advisory Committee

Professor Andy LiWang

Professor Patricia J. LiWang, Supervisor

Professor Maria Zoghbi

2017

The dissertation of Anna Faith Nguyen, titled, “Purification and Study of CC Chemokine-Based Strategies to Combat Chronic Inflammation and HIV”, is approved, and is acceptable ion quality and form for publication on microfilm and electronically:

_____ Date _____

Professor Andy LiWang

_____ Date _____

Professor Maria Zoghbi

Supervisor _____ Date _____

Professor Patricia J. LiWang

Chair _____ Date _____

Professor Michael Cleary

University of California, Merced

2017

To my friends, family, and especially husband.

I love you all so much; thank you for your endless support and patience.

And Marko, thank you for not popping out too soon.

Table of Contents

I. List of Figures

Chapter 1 Figures:

Chapter 2 Figures:

Chapter 3 Figures:

Appendix B Figures:

Appendix C Figures:

II. List of Tables

III. List of abbreviations

TBI	Traumatic Brain Injury
HIV	Human Immunodeficiency Virus
AIDS	Acquired Immune Deficiency Syndrome
GAG	Glycosaminoglycan
PLC	Phospholipase C
PIP2	Phosphatidylinositol 4,5-bisphosphate
IP3	Inositol 1,4,5-triphosphate
DAG	Diacylglycerol
PKC	Protein Kinase C
CD4	Cluster of Differentiation, 4
CCL“X”	Chemokine (C-C Motif) Ligand “X”, where “X” represents a number
CCR“X”	CC Chemokine Receptor “X”, Where “X” represents a number
RA	Rheumatoid Arthritis
RANTES	Regulated on Activation, Normal T-cell Expressed and Secreted; also called CCL5
Gln/Q	Glutamine
Glu/E	Glutamate
Asn/N	Asparagine
Phe/F	Phenylalanine
Ile/I	Isoleucine
DMEM	Dulbecco’s Modified Eagle’s Medium
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)
LTR	Long Terminal Repeat
CPRG	Chlorophenol red-β-D-galactopyranoside
HBSS	Hank’s Buffered Salt Solution
FBS	Fetal Bovine Serum
HSQC	Heteronuclear Single Quantum Coherence
CHO	Chinese Hamster Ovary
IPTG	Isopropyl β-d-1-thiogalactopyranoside

TFA	Trifluoroacetic Acid
vCCI	Viral CC Chemokine Inhibitor
vMIP-II	Viral Macrophage Inflammatory Protein II
NMR	Nuclear Magnetic Resonance
HSQC	Heteronuclear single quantum coherence
ITC	Isothermal Titration Calorimetry
NaOP	Sodium phosphate
HHV-8	Human Herpesvirus-8, Kaposi's sarcoma-associated herpesvirus
MIP-1β	Macrophage Inflammatory Protein Beta, also known as CCL4
IPTG	Isopropyl β-D-1-thiogalactopyranoside
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
βME	β-mercaptoethanol
RMSD	Root mean squared deviation
RMSF	Root mean squared fluctuation (for specific residues)
Tris	Tris(hydroxymethyl)aminomethane
MD	Molecular Dynamics
NLS	Nuclear Localization Sequence
PTD	Protein Transduction Domain
NaCl	Sodium Chloride
EDTA	Ethylenediaminetetraacetic acid
NaOAc	Sodium Acetate
NaOP	Sodium Phosphate
PCR	Polymerase Chain Reaction

IV. Acknowledgements

I would like to particularly thank my advisor, Prof. Patti J. LiWang, for her support and the opportunity to work with her, as well as her boundless patience throughout my graduate career. I thank Prof. Michael Cleary for and serving as chair of my committee. I thank Prof. David Ojcius and Prof. Jinah Choi for previously serving as members of my committee, as well as Prof. Andy Liwang and Prof. Maria Zoghbi for serving as committee members for my dissertation defense. I thank all the members of both Prof. Andy and Patti LiWang groups for taking their time to help and teach me whenever I was in need of assistance. I thank my undergraduate researchers, Mike, Megan, Laura, Tiffany, and Ricardo for helping me with my experiments and wish all of you luck in your future endeavors. I especially thank Dr. Li Zhang and Dr. Yonggang Chang for guiding me and teaching me how to truly think using the scientific method.

All the work presented was supported by NIH R01AI112011, as well as funding provided by UC Merced SNS Summer fellowships. The work in Chapter 2 was partially supported by the National Science Foundation through the NSF-CREST: Center for Cellular and Bio-molecular Machines at UC Merced (NSF-HRD-1547848), with M.E.C.'s contribution also partially supported by the Art and Fafa Kamangar Chair in Biological Science at UC Merced.

The text of Chapters 2 and 3 of this dissertation are a reprint of the material as it appears in the *International Journal of Molecular Sciences* with minor edits. The co-authors listed in those publications directed and supervised research which form the basis for the dissertation.

V. Vita for Anna Faith Nguyen (Maiden Name: Ankirskaia)

Education

2012-2017: Ph.D. University of California, Merced, Quantitative Systems Biology with research done under the supervision of Professor Patricia J. LiWang

- The design, purification and study of CC chemokine-based strategies to combat chronic inflammation and HIV utilizing biochemical, molecular biological, and cellular biological methods

2008-2011: Bachelors of Science, University of California Davis, Biochemistry and Molecular Biology with research done under the supervision of Professor Barry Wilson

- The study of the protective effects of pyridostigmine bromide against organophosphates with acetylcholinesterase utilizing molecular biological methods

List of Publications

Nguyen, A. F. (2017) Purification and Study of CC Chemokine-Based Strategies to Combat Chronic Inflammation and HIV: A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Quantitative and Systems Biology. *Dissertation, University of California, Merced,* 1-103

Nguyen, A. F., Schill, M. S., Jian, M., and LiWang, P. J. (2017) The Effect of N-Terminal Cyclization on the Function of the HIV Entry Inhibitor 5P12-RANTES. *Int. J. Mol. Sci. 18*, 1575.

Nguyen, A. F., Kuo, N. W., Showalter, L. J., Ramos, R., Dupureur, C. M., Colvin, M. E., and LiWang, P. J. (2017) Biophysical and computational studies of the vCCI:VMIP-II complex. *Int. J. Mol. Sci. 18*, 1778.

Henderson, J. D., Glucksman, G., Leong, B., Tigyi, A., **Ankirskaia, A**., Siddique, I., Lam, H., Depeters, E., and Wilson, B. W. (2012) Pyridostigmine bromide protection against acetylcholinesterase inhibition by pesticides. *J. Biochem. Mol. Toxicol. 26*, 31–34.

Awards and Fellowships

2017 QSB End of Spring Research Fellowship
2016 P. LiWang Lab Summer Research Fellowship
2014 QSB 2014 Fall Retreat: Outstanding Presentation for Interdisciplinary Audience Award
2014 QSB Summer Research Fellowship
2013 QSB Summer Research Fellowship
2013 Supplemental QSB Summer Fellowship
2011 UC Davis 22nd Undergraduate Scholarship and Creative Activities Conference presenter

VI. Abstract

Purification and Study of CC Chemokine-Based Strategies to Combat Chronic Inflammation and HIV

Anna Faith Nguyen

Doctor of Philosophy

University of California, Merced

2017

Supervisor: Professor Patricia J. LiWang

The human immune system is among one of the most complex organ systems in the human body. While it is responsible for healing wounds and protecting from disease, it can also have deleterious effects when over- or under-active. For example, many people who experience traumatic brain injury (TBI) suffer many of their ill effects due to the increased pressure on their brain, caused by an influx of immune cells. In another example, Auto-Immune Deficiency Syndrome (AIDS) is a progression of the Human Immunodeficiency Virus (HIV) that results in the death of immune cells, and thus, leaves the person open to opportunistic infections. Proteins involved in both these immune processes are known as chemokines (chemoattractant cytokines), which are small, secreted proteins that recruit and activate leukocytes, and can signal their movement towards a site of inflammation or infection (known as chemotaxis). As a result, more detailed study of the chemokine system is necessary in order to discover novel strategies to combat immune system-related ailments. This dissertation focuses on the broad applications involving utilizing certain segments of the chemokine system as biotherapeutics in chronic inflammation and HIV infection. First, the purification and structural study of the poxvirus viral CC Chemokine Inhibitor (vCCI) that naturally inhibits chemotaxis helps shed light on the mechanism behind this protein's broadly-applicable method of binding CC Chemokines, revealing key residues that are likely vital for the mechanism behind the protein's ability to broadly inhibit inflammation. Later, CC Chemokines are also studied in the context of HIV-inhibition, focusing on the N-terminal properties of a natural chemokine analog, 5P12-RANTES, which has been shown to inhibit HIV entry into the human cell. This inhibitory chemokine analog is also used to design and create second-generation HIV entry inhibitors, 5P12-Linker-T1144 and 5P12-Linker-T2635, which inhibit HIV at two separate stages of HIV entry process. And finally, chemokine analogs are suggested as a possible system for targeting cells likely to contain latent HIV DNA, to later activate transcription and allow the HIV-infected cells to be recognized by the immune system by utilizing the protein 5P14-Linker-Tat. Overall, this dissertation reveals structural insights into the proteins vCCI, vMIP-II and 5P12-RANTES that help elucidate the properties of the chemokine system, as well as offers three newly-engineered proteins, 5P12-Linker-T1144, 5P12-Linker-T2635 and 5P14-Linker-Tat to be used to combat HIV infection with preliminary results showing signs of efficacy.

Chapter 1

Introduction

1.1 Introduction to CC Chemokines

The human immune system is among one of the most complex systems in the human body, with new information being discovered about it every day. While it is responsible for healing wounds and protecting the body from foreign invaders or the development of cancers[1], it can also have deleterious effects when rendered too weak or too active. Too weak of a response, such as in the case of late stage Auto-Immune Deficiency Syndrome (AIDS), and the person is no longer able to protect their body against opportunistic infections[2]; infections that the immune system would normally eliminate from the body can suddenly lead to death. Too strong of a response, such as in the case of Traumatic Brain Injury (TBI), and intracranial pressure can increase to the point of brain damage or death[3]. Therefore, there is great need for therapeutics that can modulate the immune response during such extreme cases, but due to the inherent complexity of the immune system, this still proves a difficult goal to achieve with various disorders. Narrowing the focus to a specific part of the immune response may help provide an avenue for such therapeutics.

One of the key functions of the immune system is the migration of leukocytes from the blood stream to sites of infection or injury, also known as leukocyte extravasation[4]. The migration of the cells, known as chemotaxis, is regulated by the cell's response to certain small peptides known as chemokines (chemotactic cytokines). Chemokines are released by various immune cells (mainly mononuclear monocytes) during times of initial infection or injury[5]. These cytokines can then interact with other cell types, such as endothelial cells, to signal for the secretion of more chemokines and the expression of adhesion proteins, which will allow leukocytes to attach to the endothelial cell surface[6]. Chemokines then bind relatively weakly and non-specifically to sugar-based glycosoaminoglycans (GAGs, a type of proteoglycan) naturally found on the surface of these endothelial cells, and this chemokine retention creates an increased, localized concentration of chemokines near the sites of infection or tissue damage[7].

The production of a chemokine gradient along the outer surface of endothelial cells increases the probability that a passing leukocyte will be able to interact with a chemokine attached to a GAG. The leukocytes express chemokine receptors on their surface, which form a much higher-affinity interaction for the chemokine compared to the interaction between the GAGs and chemokines. So upon interaction with a leukocyte, chemokines bind to the leukocyte's chemokine receptor protein, causing a conformational change to the receptor that results in activation of the heterotimetric G-protein (which interacts and associates with the chemokine receptor)[8]. Heterotimetric G-proteins are signaling proteins that interact with the cell membrane, which are made up of several subunits (Alpha, Beta, and Gamma) that normally associate together when

inactive but dissociate upon activation (such as due to a signal from a chemokine receptor) to act as downstream signal activators[9]. Once activated, the G-protein subunits associated with chemokine receptors dissociate and begin to propagate across the cell to cause downstream signaling effects, with the Alpha subunit associating with and activate phospholipase C (PLC)[10]. PLC hydrolizes phosphatidylinositol 4,5-bisphosphate (PIP2, a common membrane phospholipid) to produce both inositol 1,4,5-trisphosphate (IP3), which stimulates intracellular Ca^{2+} mobilization from the endoplasmic reticulum[11], and diacylglycerol (DAG), which, with calcium, is required to activate protein kinase C (PKC)[12]. PKC continues this signal transduction cascade by activating several GTPases involved in the activation of actin, while the calcium released further results in the activation of myosin[13]. All these reactions together result in the generation of lamellipodia and filopodia, the main structures that allow for cell migration[14], ultimately allowing for leukocytes to travel down the increasing chemokine gradient, towards the initial site of infection or tissue damage. (Figure 1.1). Once there, the leukocytes can begin the process of combating the foreign invaders or aiding in healing the tissue damage.

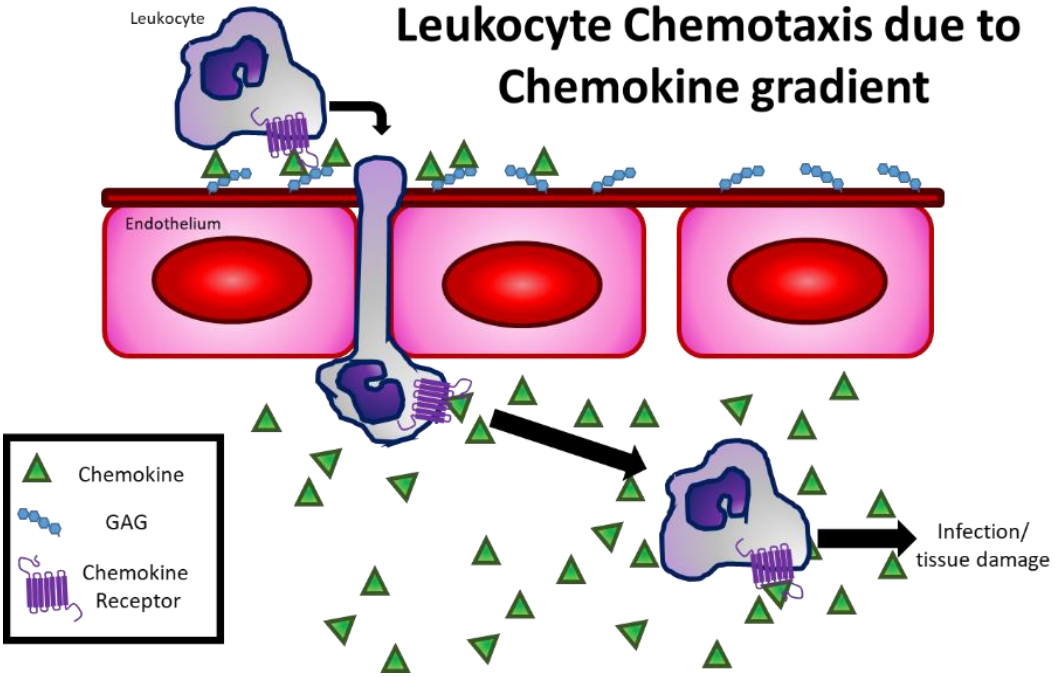

Figure 1.1: Leukocyte Chemotaxis Due to Chemokine Gradient. A seven-transmembrane receptor (purple) on a leukocyte surface will bind to a chemokine (green) and become activated. Once activated, a downstream signaling cascade will result in the cell moving up the increasing chemokine gradient, towards the site of infection or tissue damage.

Because chemokines are the instigators of this vital migration process, the chemokine system has been under heavy study ever since the discovery of the first chemokine, Interleukin-8 (now known as CXCL8) in the 1980s[15]. Chemokines are small, soluble polypeptide signaling molecules, usually 7 to 14 kDa in size, which are secreted by immune cells. Like antibodies, these secreted proteins contain disulfide bonds, the positioning of which is used to separate the chemokines into four classes: C, CC, CXC,

and CX_3C, where each “C” represents a conserved cysteine residue (capable of disulfide bonding to another cysteine) found near the N-terminus of the chemokines and the 'X' represents an intervening amino acid between the cysteine residues. Though originally the chemokines were given specific names related to their functions, more recently, an official naming system for the chemokines was employed, where chemokines were named based on the type of chemokine followed by “ligand” and a number, so CC chemokines would be CCL1, CCL2, etc, while CXC chemokines would be CXCL1, CXCL2, etc[16]. As a result, many chemokines are referred to by both their “common name” (for example, RANTES, which stands for Regulated on Activation, Normal T-cell Expressed and Secreted) and their official chemokine nomenclature name (CCL5, in the case of RANTES.)

CC and CXC are the largest and best-classified families of chemokines. The chemokines in both these families are generally involved in the processes of migrating leukocytes for the purpose of development, homeostasis, and inflammation. Just as the chemokines themselves are split into classes, so are their receptors. CC Chemokines bind only to CC Chemokine Receptors (CCR, numbered such as CCR1, CCR2, etc.), while the CXC chemokines bind only to CXC-Chemokine Receptors (CXCR1, CXCR2, etc.) While both families of receptors are specific for their class of chemokine, they show varying levels of specificity for the specific chemokines within that class that can activate them; for example, CCR1 can be activated by 11 different CC chemokines, while CCR6 has only been seen to be activated by CCL20 (Figure 1.2). This promiscuity within the chemokine system allows for various levels of redundancy and robustness of response[17]; many chemokines can be knocked out of the genomes of mice without resulting in lethal phenotypes, and instead, usually show differing responses to specific immune signals[18]. Further flexibility in the chemokine system is shown in that the same cell usually produces many different chemokines concomitantly in response to the same stimulus, a trait known as polyspeirism. Polyspeirism provides another level of robustness as certain fluctuations of the individual chemokine components will not greatly influence overall immune function[19], protecting the chemokine system from being easily modulated by infectious agents.

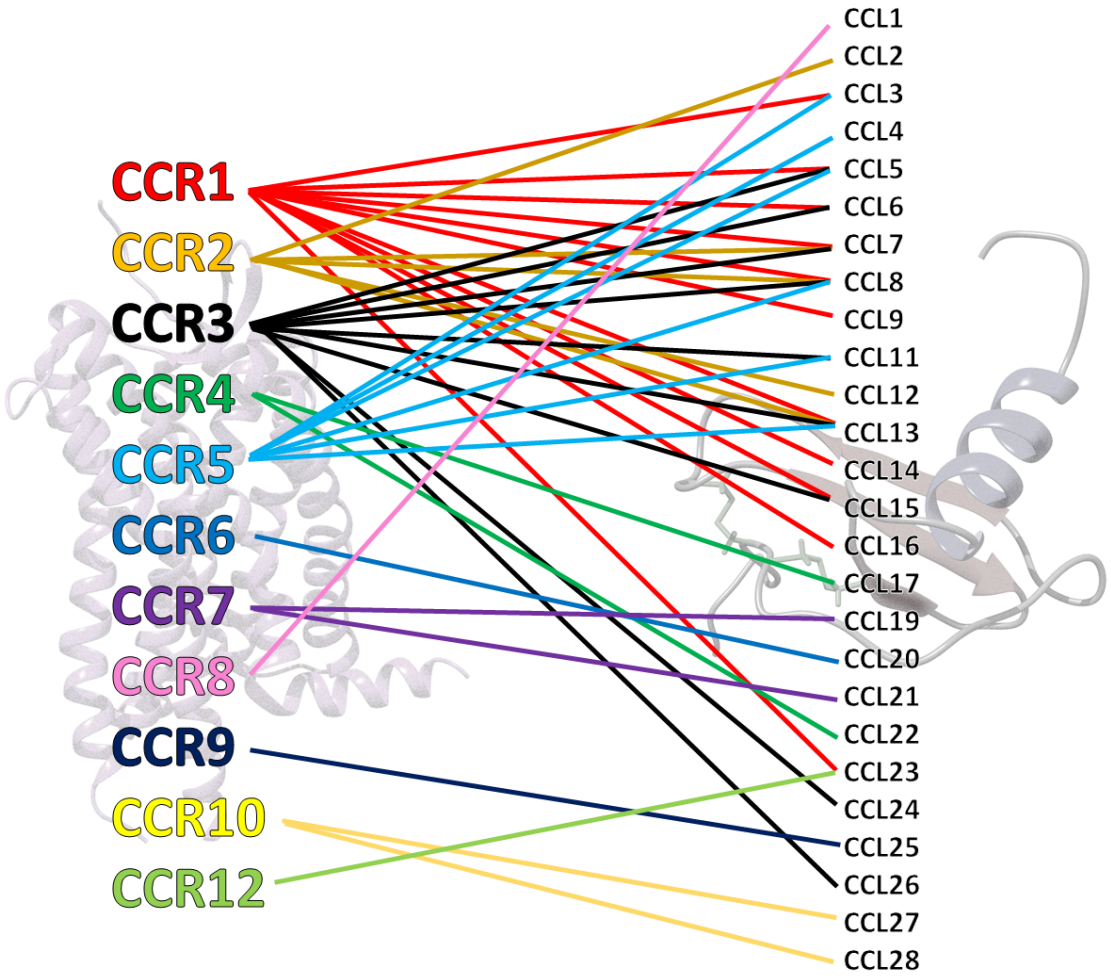

Figure 1.2: CC Chemokine Receptors and Their CC Chemokine Ligands. Chemokine receptor-ligand pairings show considerable redundancy and robustness, especially with CC Chemokines. Background structures of CCR5 (PDB code 5UIW) and Eotaxin-1 (CCL11, PDB code 1EOT) were prepared using UCSF Chimera.[20]

While the general function of chemokines is crucial for healthy growth and immune protection in the organism, malfunctions within the chemokine system have been shown to result in various diseases due to inappropriate levels of inflammation and the subsequent tissue damage[18]. The chemokine system has also been shown to be vital in allowing the entry of HIV into the host cell[21], further making chemokine receptors and their ligands of considerable therapeutic interest. Of the two main families, CC chemokines have been connected to a considerable number of immune-related diseases, and a CC chemokine receptor is known to be a vital entry point for initial HIV infection, making the study of the CC chemokine system of particular importance for those interested in both immune modulation and HIV intervention.

CC chemokines have a similar structure; they all contain a long, unstructured N terminus (which has been found to be vital for receptor activation and downstream signaling response), two adjacent cysteine residues near the N terminus that form

disulfides to other sections of the protein, three beta strands, and a C-terminal alpha helix. (Figure 1.3A). They also contain small loops in the 20-, 30- and 40-amino acid regions, which have been implicated in GAG and CC receptor binding[22,23]. The receptors themselves are G-coupled seven transmembrane (7-TM) receptors with seven alpha helices that span the cellular membrane and are known to have sulfated tyrosine residues at the external N-terminal loops[24]. These 7TM receptors are also known for having a short, intracellular “eight loop” that tends to run perpendicular to the other seven alpha helices (Figure 1.3B). Chemokine receptor binding has been shown to likely follow a two-step mechanism: first, there is an interaction between the chemokine core residues and the extracellular loops and sulfated N-terminal receptor domain that results in binding (showing the specificity of a chemokine to its receptor), and then the chemokine’s N-terminus interacts with the receptor's transmembrane domain and extracellular loops[23,25], triggering a conformational change in the receptor that results in a signal beginning through the G-coupled protein pathway[26] (which determines the downstream effects). Extensive study of the N-termini of chemokines have been performed in attempt to understand this step-wise binding mechanism, which has resulted in the recognition of various antagonistic forms of certain chemokines that are capable of binding receptors but prevent or alternate downstream signaling[27] – making them important therapeutic candidates.

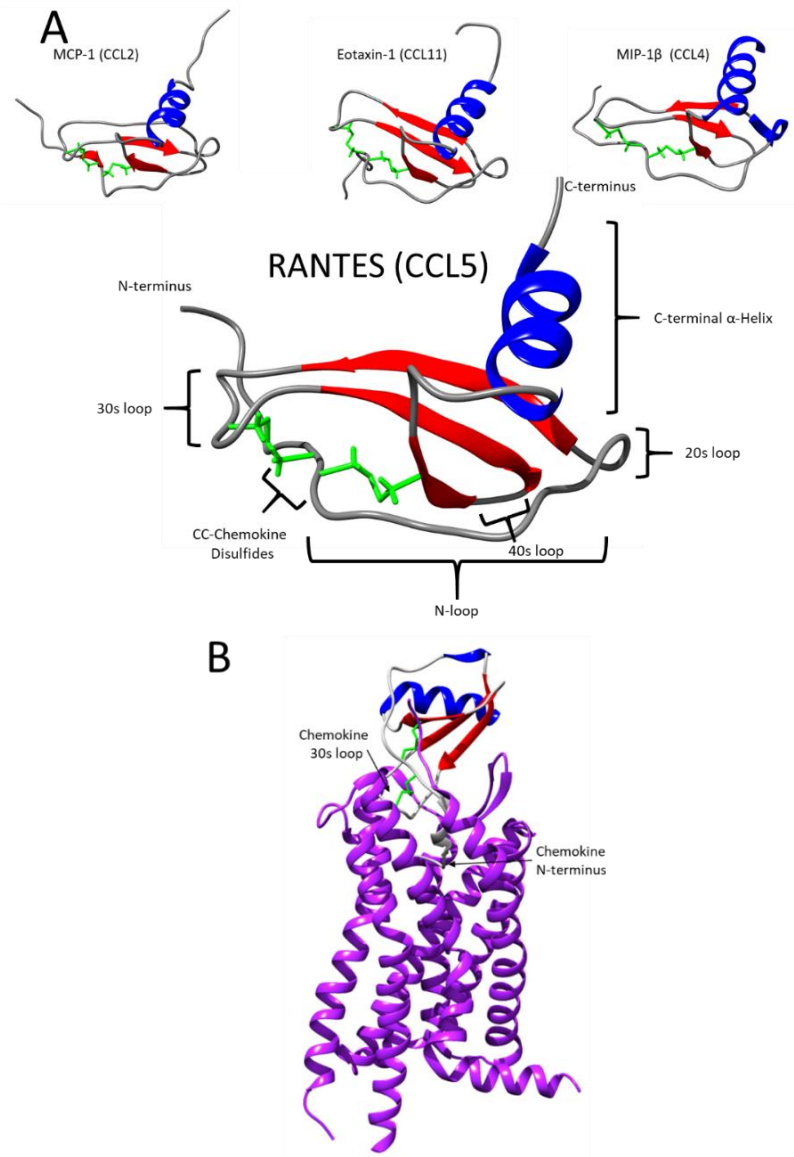

Figure 1.3: Structures of CC Chemokines and Receptors. (**A**) Four representative CC chemokines, MCP-1 (CCL2, PDB code 1DOM), Eotaxin-1 (CCL11, PDB code 1EOT), MIP-1β (CCL4, PDB code 1HUM) and RANTES (CCL5, PDB code 1RTO). Blue represents alpha helical regions, red represents beta strands, while the CC-disulfides are shown in green. (**B**) Crystal structure of CC Chemokine Receptor 5 (CCR5, in purple) in complex with high potency HIV entry inhibitor 5P7-RANTES (Grey, with blue alpha helices, red beta strands, and green disulfide bonds). Both receptor and ligand are from PDB code 5UIW, with rubredoxin removed from TM4/TM5 intracellular region, and the CCR5 N-terminal loop unresolved in the crystal structure[28]. Arrows point to main points of interaction between the chemokine and receptor, the chemokine N terminus and 30s loop regions. The chemokine is bound to the extracellular side of the receptor. All structured were prepared using UCSF Chimera.[20]

As a result, many potential therapeutic agents derived from the study of chemokines have been recognized for disorders from chronic inflammation to HIV prevention. Studying the functional properties and possible applications of the proteins involved in the CC Chemokine system could result in the discovery of further therapeutic uses for various immune-system related disorders, as well as helping elucidate the general understanding of the roles of these chemokines and receptors in the immune system. Two especially important areas of study that this dissertation will focus on are that of CC Chemokines in chronic inflammation and their involvement in HIV infection and prevention.

1.2 The Role of CC Chemokines in Chronic Inflammation

Inappropriate expression of multiple CC chemokines and chemokine receptors has been reported in a large number of inflammation-related disorders, including TBI, atherosclerosis, rheumatoid arthritis (RA), and asthma (Table 1.1). As a result, not only are the chemokines involved in these disorders of particular therapeutic interest, but so are any proteins that can down-regulate their expression or downstream signaling cascades. Of course, antibodies have been studied with this outcome in mind, and several anti-chemokine antibodies have been shown to reduce allergic inflammation and airway hyperactivity in mice[29]. But given the high specificity of antibodies, it would require a vast number of different antibodies to be able to utilize them as a viable therapy for such conditions, considering the amount of redundancy and robustness in the chemokine system (Figure 1.2). As a result, there is great need for an anti-chemokine therapeutic that is more broadly-acting than the very-specific antibodies.

Table 1.1: CC Chemokines Implicated in Inflammatory Conditions. Increases in certain CC chemokines and their receptors have been implicated in the progression of many inflammatory disorders.

Chemokine	Inflammatory Conditions
MCP-1 (CCL2)	Traumatic Brain Injury[18], atherosclerosis[18], multiple sclerosis[18], neuropathic pain[18], Rheumatoid Arthritis[18], insulin resistance[18]
MIP-1α (CCL3)	Rheumatoid Arthritis[18]
MIP-1β (CCL4)	Traumatic Brain Injury[18], Rheumatoid Arthritis[18]
RANTES (CCL5)	Atherosclerosis[18], rheumatoid arthritis[18], adipose inflammation[18], diabetes [18], asthma[30], Allergic rhinitis[30]
MCP-3 (CCL7)	Lesional Psoriasis[31], Rhinovirus-induced Asthma[32], Rheumatoid Arthritis[18], Asthma[30], Allergic rhinitis[30]
Eotaxin-1 (CCL11)	Asthma, Schizophrenia[33], Alzheimer's Disease[33], depression[33], Atopic Dermatitis[30], Allergic rhinitis[30]
MCP-4 (CCL13)	Asthma[30], Atopic Dermatitis[30], Allergic rhinitis[30]
TARC (CCL17)	Asthma[18], Atopic Dermatitis[30], Allergic rhinitis[30]
MDC (CCL22)	Asthma[30], Allergic rhinitis[30]
Eotaxin-2 (CCL24)	Asthma[30], Allergic rhinitis[30]
TECK (CCL25)	Crohn's Disease[18]
Eotaxin-3 (CCL26)	Asthma[30], Allergic rhinitis[30]

Certain viruses have evolved the ability to modulate chemokine signaling, as it results in fewer immune cells at the sites of infection, and thus, increased viral activity[34]. The viral proteins involved in this modulation come in three classes: chemokine receptor homologs (which mimic natural chemokine receptors, thus binding natural chemokines and preventing them from contacting their native cellular receptors), chemokine homologs (which mimic natural chemokines, thus occupying the cellular receptors and preventing the signal transduction of native chemokine interactions), and chemokine-binding proteins (CKBPs, which bind native chemokines, preventing interactions with native receptors or GAGs)[35]. Because CKBPs and viral chemokine analogs are soluble proteins capable of inhibiting a considerable variety chemokines, they have been of particular interest as therapeutic agents.

CKPBs come in three classes: Class I (proteins with a low affinity to all four classes of chemokines)[36], Class II (proteins with high affinity for a very broad range for chemokines only in the CC chemokine family)[37], and Class III (high affinity for chemokines from the three main classes of chemokines)[38]. Of these, Class II offers the tightest, broadest range of binding while still allowing for the specificity of targeting a particular class of chemokine, allowing for stronger binding than Class I, but more targeted binding than Class III, making them ideal candidates as a therapeutics for CC chemokine induced inflammatory conditions.

Of the CKPB Class II proteins, one of the best therapeutic candidates is the viral CC Chemokine Inhibitor protein (vCCI, also known as 35K) found in a family of poxviruses that includes rabbitpox virus, cowpox virus, vaccinia virus, variola virus, myxomavirus, and Shope fibromavirus[39]. (Figure 1.4). It has been shown to bind to over 80 CC chemokines across several species, about 20 of which have nanomolar affinity to this inhibitor.[40,41] This broad range yet tight binding has put vCCI has been under consideration as an anti CC chemokine therapeutic agent[18,42–45], making understanding its mechanism of binding of high importance. Structural studies of vCCI in complex with natural chemokines have revealed several close contacts that are critical for binding[41,46] while mutagenesis of vCCI by others *in vitro* and *in vivo* has confirmed the importance of several of the residues suggested by the structural studies[47,48]. Discovering the structural reason for this flexible yet tight binding of vCCI could be vital in creating new therapeutic methods of combating CC chemokine induced inflammatory conditions.

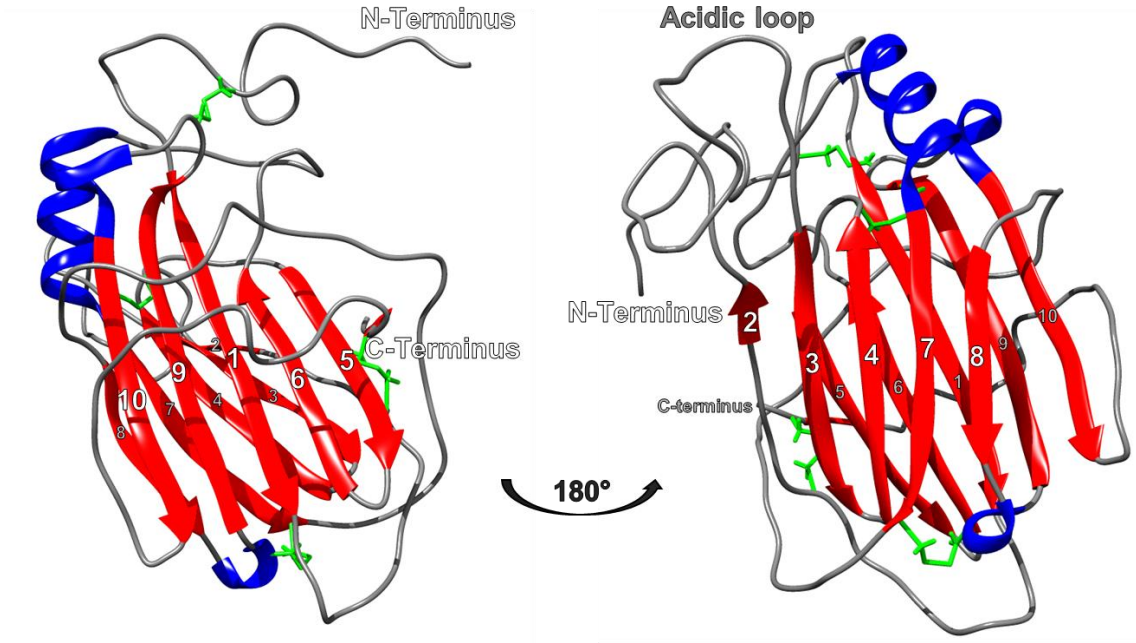

Figure 1.4: The Rabbitpox Viral CC Chemokine Inhibitor Protein. Rabbitpox vCCI protein (PDB code 2ffk) visualized at two angles 180° apart. Beta strands are in red, alpha helices in blue, disulfide bonds are shown in green, with the beta strands numbered and N and C termini labeled. The acidic loop between beta strands 2 and 3 labeled in the right image. Structures were prepared using UCSF Chimera.[20]

Mutagenesis studies have also been carried out on the chemokines that bind to vCCI, discovering several residues that are important for the binding of CC Chemokines to vCCI[41,44,49]. However, while extensive mutation to single chemokines such as MCP-1 has proven very useful[44,49], mutating every chemokine in search of which residues are the ones most vital for vCCI recognition would be a very large endeavor, so discovering another way to recognize which sites are most likely to be evolutionarily conserved in the binding of vCCI was desirable. Interestingly, Human Herpesvirus-8 (HHV-8) encodes a viral chemokine analog that has been shown to bind several CC Chemokine receptors (CCR1, CCR2, CCR5), as well as at least one CXC receptor (CXCR4)[50–52]. This HHV-8 protein is known as virally encoded Macrophage Inflammatory Protein-II, or vMIP-II, for its similarity to the natural chemokine MIP-1β (Macrophage Inflammatory Protein Beta, also known as CCL4). This range of receptor binding is much greater than a typical chemokine, implying that this viral protein contains certain motifs that are likely vital to the binding and activity for a very broad range of chemokines and thus, very likely to be evolutionarily conserved -- motifs that are likely also recognized by vCCI.

Thus, a hypothesis was formed that a significant amount of insight could be obtained by determining whether a tight complex could be formed by these proteins, vCCI and vMIP-II, and if so, which residues are similar between this new vCCI:vMIP-II complex and results found in previous investigations of vCCI with native chemokines and various chemokine mutants. Due to the broad range of anti CC Chemokine action of both these viral proteins, understanding their structural properties in more details will likely reveal novel protein-based areas of interest for manipulation or control of unwanted immune responses.

To test this hypothesis, Chapter 2 of this dissertation offers a technique for producing rabbitpox vCCI from *E. coli*, as well as experimental and a molecular dynamics analysis of the vCCI:vMIP-II complex and the vCCI:MIP-1β complex (vCCI with a natural chemokine), comparing these two complexes to explore the differences in binding between the virus-encoded chemokine analog and a natural, human chemokine[46].

1.3 The Role of CC Chemokines in HIV Infection

1.3.1 HIV Entry into the Human Cell

CC Chemokines also have a long history as tools against the Human Immunodeficiency Virus (HIV), the virus that causes Acquired Immune Deficiency Syndrome (AIDS), a condition that causes an individual's immune system to fail, allowing for opportunistic infections to prove fatal.[53] This virus has the ability to asymptomatically propagate infection for many years, causing many individuals to be unaware of their contagious state for extended periods of time.[54] Despite decades of effort in trying to find a vaccine or a cure, and considerable progress, success has yet to be atained[55], and despite educational attempts all over the world, there are still slightly under 2 million new infections every year, a number that has remained grimly static in recent years[56]. Even in the US, where there is very little stigma surrounding condom use, readily available sex education, and only about 2% of the world's total infections, there are still over 40,000 diagnoses every year[57].

While there is a recent focus on a cure for HIV, a different strategy to combat this virus is to prevent new infections entirely. While many preventative therapies use inhibitors that take effect only after the therapeutic is able to enter the cell (by targeting proteins involved in the formation of HIV proviral DNA or integration of that DNA into the human chromosome),[58], including the inhibitors used in the newly-popular HIV pre-exposure prophylaxis (PrEP)[59], a different type of HIV inhibitor known as entry or fusion inhibitors can inhibit the virus at an even earlier step of initial infection, during attempted viral entry into the human cell. By blocking entry into the cell, these inhibitors prevent any viral content from ever having the opportunity to interact with the inner machinery of the human cells[60].

For HIV inhibitors to be readily available in the countries with the highest rates of new infections, these inhibitors must be safe, stable, and cost-effective.[61] Many HIV entry inhibitors are based on antibodies or small molecules, which tend to be expensive to synthesize or have a higher possibility of the virus mutating immunity to them, as they are currently being used as treatments, and thus, would be difficult to make safely available in countries of the developing world.[62] So instead, some groups have focused on peptide or small protein-based HIV entry inhibitors, which could be synthesized in a more cost-effective manner compared to methods like chemical synthesis[63].

HIV is a positive-strand RNA lentivirus (a subtype of retrovirus), usually around 130 nm in diameter[64], whose surface is made up of a cellular membrane derived from the human host cell it had replicated inside, which was then studded with virally-encoded glycoproteins known as Env. Env proteins take the form of a “viral spike”, a trimer of heterodimers (six proteins total) roughly 9 nm tall and 12 nM in diameter[64,65], with the heterodimers made up of the proteins gp120 (a heavily, mostly mannose-glycosylated

subunit) and gp41 (the fusion peptide), two subunits that work together to allow for viral entry into the cell.[66] Understanding how these Env proteins work to allow fusion with and viral content release into the human cell is vital in the selection and construction of HIV entry inhibitors.

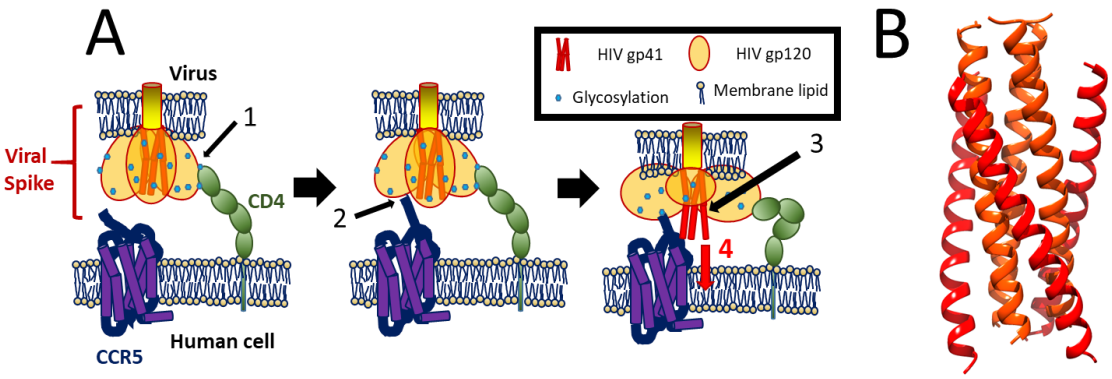

Figure 1.5: HIV Entry into the Cell. (A) HIV virus with a spike consisting of trimers of gp120 (orange spheres, glycosylated with blue) and gp41(red, obscured by gp120) spike on its surface approaches a $CD4^+$ cell and binds first to CD4(green) (black arrow, 1). The gp120 then binds to a co-receptor, CCR5 (purple), causing a conformational change (black arrow, 2). This change exposes gp41 (black arrow, 3) and allows for the fusion peptide gp41 to fuse the membranes together by forming a six-helix bundle from the three alpha helices, with the C- and N- terminal helices coming together (red arrow, 4). (B) The gp41 six-helix bundle formed to aid in the fusion of the viral and host membranes. C-peptides are in red, N-peptides are in orange (PDB code 1AIK). Structure was prepared using UCSF Chimera.[20]

The HIV entry mechanism occurs in several steps (Figure 1.5A). First, the virus uses gp120 to bind to its main receptor, the human protein CD4 (Cluster of Differentiation 4)[67], which is typically found on the surface of cells such as certain types of T-cells, macrophages, and dendritic cells, cells that are typically involved in immune response and accumulate at the site of infections[68]. The gp120:CD4 interaction induces a slight conformational change that allows for gp120 to then bind to its co-receptor, the human chemokine receptor CCR5[69] or CXCR4[70]. Once both the receptor and a co-receptor are bound, gp120 goes through a larger conformational change, revealing gp41, which previously was buried and protected inside the trimetric core[71]. For the final steps, gp41 goes through a conformational change and extends the fusion peptide found at its N-terminus towards the host cell to pierce the human cell membrane. The C-terminal helices and the N-terminal helices of gp41 then come together into a six helix bundle (Figure 1.5B) that pulls the two membranes together, allowing the virus to physically enter the cell[72]. As this process happens in a step-wise manner, several points are susceptible targets for HIV entry inhibitors. There have been inhibitors that block CD4 binding,[73–75] co-receptor binding[76,77], and membrane fusion,[78,79] as well as general lectins that bind to the glycoproteins on the HIV surface.[80]

Because the co-receptors for HIV are chemokine receptors, much research has been done on the effects of chemokines on HIV entry. There have been no reported cases of new infections through virus subtypes that use the CXCR4 receptor[70], which is

connected to later-stage HIV infection (progression into AIDS), so those focusing on preventing initial HIV infection generally have focused on the CCR5 co-receptor, and thus, CC Chemokines. Soon after the discovery of these co-receptors, scientists noticed that the native chemokines for CCR5 (MIP-1β/CCL4, MIP-1α/CCL3, and RANTES/CCL5) could slightly inhibit HIV infection.[81] This discovery led to subsequent study into the structural properties of these chemokines. It was discovered that the initial ten or so residues on these chemokines are important to their function; removal of these N-terminal residues from RANTES[82] and MIP-1β[83] resulted in a molecule that could bind to CCR5 and block HIV infection, but did not activate the signal transduction pathway that led to inflammatory response. This was intriguing, since unnecessary inflammatory response can be detrimental to the body (see section 1.2), as well as bring $CD4^+$ cells to the site of HIV infection, the cell types susceptible to HIV infection[18]. Preventing this inflammatory response made the CC chemokine RANTES an ideal scaffold for therapeutics that could prevent HIV infection (Figure 1.3A, lower image). Although it may seem worrisome to block an CC chemokine receptor, there are healthy individuals who are born with a mutation that prevents them from expressing functional CCR5, a mutation known as CCR5Δ32.[84] This suggests that there should be no negative effects to blocking this receptor, and clinical trials of small-molecule CCR5-blockers have shown this to be true[84].

One intriguing discovery was that the addition of a Methionine residue to the N-terminus of native RANTES (dubbed the "0" position, so that the numbering system for the other amino acids in the native proteins need not change) resulted in a molecule that could bind and block HIV infection better than the truncated RANTES variants, yet still did not cause a signaling cascade.[85] This result spurred the discovery of several other RANTES variants that focused on modifying N-terminal residues, including the additional 0 position, through mutagenesis studies.[86] The most potent variant discovered during this time was AOP-RANTES, which is RANTES with an aminopoxypentane group replacing the first three residues, with an IC_{50} of 1.2 nM and no signal cascade activation[87]. However, this modified peptide required chemical synthesis to make, and thus, was not as cost effective as a peptide made of natural amino acids. In an attempt to find a just-as-potent (but cheaper) inhibitor, the Harley group performed mutagenesis on the N-terminus of RANTES and then used phage display in order to try and find viable inhibitor candidates.[88] This resulted in the discovery of P1- and P2-RANTES, which in turn were used as scaffolds for creating new libraries of RANTES variants. At the end of this process, the potent inhibitor 5P12-RANTES was discovered (Figure 1.6.)[63]

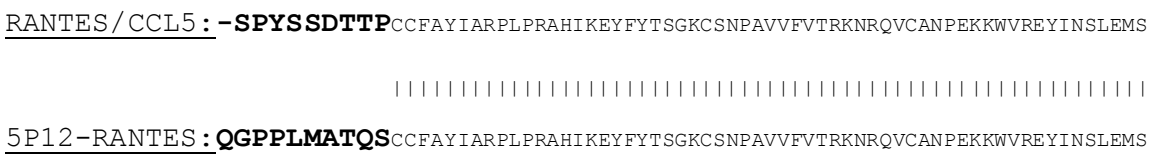

Figure 1.6. Sequence comparison of wild type RANTES/CCL5 and the variant 5P12-RANTES. Identical residues are linked with a line, while the differences between the N-termini of the proteins are highlighted in bold. Dash in wild type RANTES/CCL5 sequence indicates that it lacks a residue in the “0” position, which is filled in 5P12-RANTES.

This chemokine analog not only shows sub-nanomolar HIV inhibition with no signs CCR5 activation[63], it has also been shown that HIV has difficulty evolving around this inhibitor[89] while also displaying full inhibition in vaginally-challenged macaque models[76]. Because this entry inhibitor is such a promising therapeutic, understanding all of its properties is vital, both for academic and practical purposes. Subsequently, Chapter 3 of this dissertation will be focusing on the thoroughly characterizing the chemical properties N-terminus of this peptide inhibitor and their effects on HIV inhibition.

Due to HIV's propensity for mutation in order to lose sensitivity to theraprutic methods[90], it is difficult to bring these inhibitors into clinical trials in fear of patients mutating highly-resistant strains of HIV. Because 5P12-RANTES only binds to CCR5, viruses that use CXCR4 as its co-receptor will not be blocked by this inhibitor.[89] Even though these proteins are intended as a preventative microbicide, there is a risk that individuals already infected with HIV (but unaware of their infection) may attempt to use them. So, there is a worry that inhibitors that only block CCR5 may exert an evolutionary pressure to push the virus towards using CXCR4 instead, which has been seen in some individuals using a small molecule CCR5 inhibitor, maraviroc[91]. HIV usage of CXCR4 has been linked to many debilitating consequences of late-stage AIDS[92], and thus, a push towards mutations that would preferentially utilize this receptor is something that HIV inhibitor candidates would need to attempt to avoid.

In an attempt to get around some of this disadvantage, the P. LiWang group created a chimeric version of 5P12-RANTES by linking it to HIV C-peptide analogues, creating a protein called 5P12-Linker-C37.[93] C-peptides analogs are named so because they mimic the sequence of the C-terminal alpha helix found in HIV gp41.[94] By mimicking this region of gp41, these C-peptide analogs are able to form a coiled coil with the N-terminal alpha helices of gp41 once they are exposed after gp120 receptor and co-receptor binding, preventing them from forming the six-helix bundle necessary for membrane fusion[95] (Figure 1.5B). As a result, one of these C-peptide analogs, T20, had been allowed into clinical trials to help aid those who have developed resistance to other HIV inhibitors,[78] and has since been approved as a drug by the FDA[96]. Because these C-peptides interact with gp41, a highly-conserved part of the HIV virus, they are very broadly-acting with both CCR5 and CXCR4-utilizing clades of virus and can compensate for some of the disadvantages seen with 5P12-RANTES.[93]

However, it has been shown that there are several mutations that gp41 may evolve that C37 (The C-Peptide in 5P12-Linker-C37) is weak against[97], including several mutations that result in T20 resistance, as well. The ability to overcome such mutations would be vital for any new gp41 inhibitors. C37 also has poor pharmacological properties, making it less-than-ideal for use in patients, which is why is has generally been restricted to laboratory testing,[98] making it unlikely that 5P12-Linker-C37 would be truly considered for clinical use. However, a new generation of C-peptides have been created and optimized to be more resistant to the mutations that T20 is weak against (some of which C37 has also shown weakness to,)[99] making them promising candidates for new therapeutics. Two next-generation C-peptides, T-1144 and T-2635, both show good pharmacokinetic properties as well as an increased level of difficulty for HIV to be able to mutate resistance to these inhibitors.[100,101] These traits suggest that they are viable candidates for creating a next-generation set of chimeric inhibitors, which may be

more broadly-acting than the previous C37 variants, while still maintaining the same low-cost and broad inhibition. The expression, purification, and some preliminary testing of these new chimeric inhibitors are explored in Chapter 4 of this dissertation.

1.3.2 HIV Latency in Human Cells

Unfortunately, preventing HIV entry will not cure individuals already infected with the virus. Although the life expectancy of those infected with HIV has increased drastically since the discovery of AIDS[56], the treatment for the condition (known as combined antiretroviral therapy, or cART) is both costly and necessary to take throughout an entire lifetime, and it also results in certain negative long-term health effects that accumulate over an extended period of time.[102] Though there was a presumption that over time, patients would eventually be able to deplete the viral stores in their bodies, it had been shown that if those who take cART (even for years with non-detectable viral load, meaning that the therapy was successfully suppressing the virus), stop their treatment regimen, there is a rapid rebound in viral load.[103] This means that, even after years of treatment, the patients are somehow still harboring cells with viral DNA that is capable of replication, meaning that a cure through cART is not possible, and if a cure is to be found, the reason for this viral rebound must be thoroughly understood.

HIV is a retrovirus, meaning that the viral RNA genome is reverse-transcribed into a DNA template that is then integrated into the host DNA.[104] Once HIV enters the cell (See section 1.3.1 and Figure 1.5) and the viral contents are released from the membrane, the positive-stranded RNA template and the associated proteins, including reverse transcriptase and integrase, begin to form a DNA template to integrate into the host chromosome (Figure 1.7A). In the cytoplasm, reverse transcriptase then uses the RNA template to create a single-stranded negative DNA strand by utilizing a host tRNA as a primer. RNAse H then degrades a vast majority of the original RNA template until only a small section is left, which reverse-transcriptase uses with the negative-strand DNA template to form a positive-strand DNA primer. This primer finally attaches to the correct position on the negative-strand DNA, and reverse transcriptase extends both the segments for a final double-stranded DNA.[105] The new viral DNA, known as HIV provirus, then associates with several viral proteins, including integrase, and is transported to the nucleus. Integrase blunts the 3' ends of the HIV proviral DNA so that it can then help the proviral DNA attack the phosphodiester bonds in the target DNA (how exactly the location in the chromosome determined is still being elucidated,[106] however is it known to often preferentially select transcriptionally active regions[106]), and then integrate into the genome as host enzymes help repair the chromosome with the new HIV insert included.[107] In many cases, once the HIV DNA is integrated into the host genome, it then begins to produce infective HIV virions, but in rare cases, some integration events result in transcriptional silencing, instead.[108]

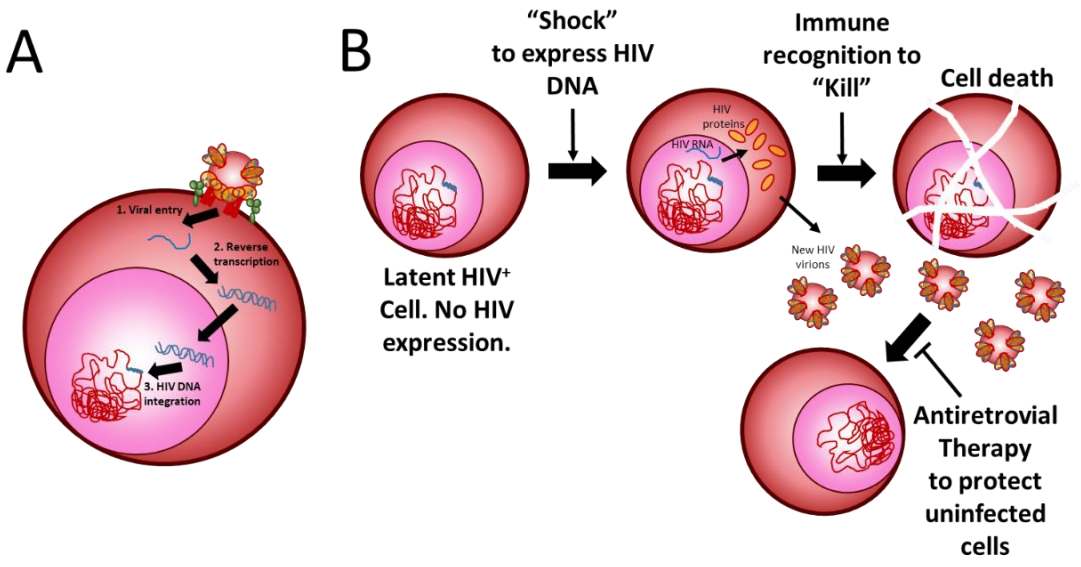

Figure 1.7. HIV Latency and Shock and Kill Method. (A) HIV virion infecting a $CD4^+$ cell and moving its RNA template to the cytoplasm (1). There, the reverse transcriptase turns the single-stranded RNA to a double-stranded DNA (2), and integrase is moved to the nucleus where it is able to integrate this DNA permanently into the human genome (3). (B) Latently infected HIV cells do not express HIV DNA despite it being in their genome, making them undetectable to the host immune system. If these cells can be "Shocked" into expressing the HIV template, the immune system will them be able to detect and "Kill" them, while cART will protect uninfected cells, which is the basis of the Shock and Kill theory of combating HIV.

These transcriptionally silent cells, also known as "latently infected cells", have long-been recognized as the main reason for viral rebound after cART[109], though more recently, it has become known that the "latent reservoir", the cells capable of harboring latent HIV DNA, is more diverse than first expected; while the main focus is on $CD4^+$ T-cells, latent HIV DNA has also been discovered in dendritic cells, macrophages (including the microglial cells), peripheral blood monocytes, astrocytes and hematopoietic stem cells.[110] If this latent reservoir could be removed from a patient, that there would be no more ability to create new HIV virions, and thus, the patient would be cured of the condition, so there is great interest in being able to somehow locate and kill the cells that are harboring latent HIV DNA. One of the leading ideas on how to go about this is called the "Shock and kill" technique[111], where the general idea is to "shock" the DNA into expressing its HIV DNA (by reversing the transcriptional silencing) and then aiding the immune system in removing the cells expressing the virus (Figure 1.7B).

There has been some success in cell[112] and animal models[113], but many strategies are very broadly acting, affecting all cell types, and many attempts still lead to toxic side effects[114,115]. As a result, there is a need for a more targeted "shock and kill" system, one that could at least narrow down the pool of cells affected by a treatment of transcriptional activation. One of the possible strategies is by using a targeted molecule to select only for certain cell types; certain chemokine receptors can be found on specific cell types[116], and though many chemokines bind to more than one receptor, in other cases, they can also be rather specific for one target receptor (Figure 1.2). The diversity

of chemokines and chemokine receptors make them an ideal system for targeting specific cell types in the immune system. In Chapter 5 of this dissertation, this idea will be explored further in the design and production of a possible HIV latency reactivation candidate.

Chapter 2

Biophysical and Computational Studies of the vCCI: vMIP-II complex

This work was published as part of Nguyen et. al (2017)[117] with slight modifications. This work was partially supported by the National Science Foundation through the NSF-CREST: Center for Cellular and Bio-molecular Machines at UC Merced (NSF-HRD-1547848). M.E.C. is partially supported by the Art and Fafa Kamangar Chair in Biological Science at UC Merced. Thank you to Ricardo Rodriguez for his assistance with protein purification and Laura Showalter with assistance in data analysis and figure preperation.

2.1 Introduction

Protein-protein interactions are critical for many aspects of biological and immunological function. Of particular interest are virally-encoded proteins that undermine the immune system, often by having the ability to promiscuously bind many targets and therefore help the virus evade immune surveillance. One such system targeted by viruses is the chemokine system, in which virally encoded proteins disrupt the chemokine receptor/ligand interaction[37]. Chemokines (chemotactic cytokines) are a class of small secreted proteins that mediate immune cell chemotaxis as part of the inflammatory response. There are about 18 human chemokine receptors that are activated upon binding to their cognate chemokine ligand[116]. About 50 chemokine ligands are known, spanning 4 sub-families. The two major subfamilies are CC chemokines (named for the adjacent Cys near the N-terminus of the protein) and CXC chemokines (named for having an intervening amino acid between the conserved N-terminal Cys). In general, CC chemokines (named numerically as CCL1; chemokine ligand 1, etc.) bind to and activate cognate receptors on the surfaces of monocytes, macrophages, and T-cells, and these receptors are numerically named as CC receptors (e.g. CCR1, CCR2). CXC chemokines tend to have cognate receptors on the surface of neutrophils[116], with receptors such as CXCR1. Chemokines can sometimes bind multiple receptors, and receptors often have more than one cognate ligand, although CC chemokines are restricted to CC chemokine receptors, and CXC chemokines have their own CXC receptors (Figure 1.2).

Because of the central nature of the chemokine system in activating and localizing immune cells, subversion of the process may be useful to a virus. Several types of chemokine-binding proteins (CKBPs) have been identified (reviewed in [118]), including those that bind chemokines from multiple subfamilies, such as M-T7 from myxoma[36], M3 from γ-herpesvirus-68[42,119–121], and the poxvirus-encoded SECRET domain[122]. These proteins have gained interest as inflammation inhibitors due to their ability to bind to pro-inflammatory proteins.

One of the most potent inhibitors of chemokine action is the poxvirus-encoded protein vCCI (viral CC chemokine inhibitor; also called 35K or p35). This approximately 240-amino acid protein binds 80 CC Chemokines across several species, about 20 of which have nanomolar affinity to this inhibitor[40,41]. The protein sequence across several poxviruses shows high identity, and the structures from cowpox[123], rabbitpox[46], and mousepox[48] reveal a beta sandwich with a binding face containing several key negatively charged amino acids, as well as a long acidic loop between beta strands 2 and 3. The P. LiWang group has previously carried out structural studies of vCCI in complex with MIP-1β (also called CCL4;[46]), which revealed details of the interaction between vCCI and this chemokine, including several close contacts that are critical for binding. Mutagenesis of vCCI by others *in vitro* and *in vivo* has confirmed the importance of several of the residues suggested by the structure, including residue E143 and amino acids of the acidic loop[47,48]. Mutagenesis studies on the chemokines themselves have also been carried out by the P. LiWang group and others and indicate that several evolutionarily conserved, positively charged residues on the chemokines are important for binding to vCCI[41,44,49]. In the group's work with a variety of eotaxin mutants (CCL11;[41]), it was shown that eotaxin's binding to vCCI was dependent on the presence of several basic residues in the chemokine, as well.

Viruses have also evolved the ability to interfere with the human chemokine system by producing chemokine homologs, small proteins that mimic the chemokine's ability to bind a chemokine receptor, thus blocking the native chemokine from interacting with native chemokines. Human Herpesvirus-8 (HHV-8) encodes several such chemokine analogs; of particular interest is the protein vMIP-II (virally encoded Macrophage Inflammatory Protein-II), which has about 40% identity with the human CC chemokine MIP-1β and has been shown to bind and antagonize several different CC chemokine receptors (CCR1, CCR2, CCR5, although it has been shown agonize CCR3 and CCR8), as well as at least one CXC receptor (CXCR4)[50–52]. This range of receptor binding is much greater than that of a typical chemokine. The P. LiWang group has previously studied vMIP-II to elucidate its ability to bind glycosaminoglycans and has shown that in solution, it is a soluble monomer with a fold similar to that of MIP-1β[124,125]. Due to its nanomolar affinity to and broad ability to bind to chemokine receptors[50,125] vMIP-II has engendered interest as an anti-inflammatory agent with some success seen in rat studies involving ischemic stroke[126], spinal cord injury[127], and kidney transplant rejection[128].

vCCI and vMIP-II are therefore complementary proteins, the former having evolved to bind a large variety of CC chemokines, and the latter having evolved to be a prototypical chemokine ligand with the ability to bind many receptors. While the P. Liwang group has studied these proteins in complex with their natural ligands[41,46,125,129], we recently developed the hypothesis that a significant amount of insight could be obtained by determining whether a tight complex could be formed by these two proteins. In other words, we set out to study the complex between an "ideal" chemokine binding protein (vCCI) and an "ideal" chemokine homolog (vMIP-II). Due to the broad action of these proteins, an understanding of these powerful viral tools may be helpful in designing new strategies to manipulate or control immune responses and could be applicable to fields ranging from autoimmunity to traumatic brain injury.

In this chapter, we present a technique for producing rabbitpox vCCI from *E. coli*, as well as experimental and a molecular dynamics analysis of the vCCI:vMIP-II complex

and the vCCI:MIP-1β complex, comparing these two complexes to explore the differences in binding between the virus-encoded chemokine analog and a natural, human chemokine[46]. Our results show that the affinity of vCCI to vMIP-II is higher than that between vCCI and natural chemokines[41,50] and suggest explanations for this high affinity as well as for previously-reported functional results.

2.2 Materials and Methods

2.2.1 Protein Purification

Please see section 5.2.1 for a more detailed explanation on the design and purification of proteins from *E. coli*.

Purifying vCCI from *E. coli*: The gene sequence encoding the rabbit pox vCCI was slightly modified by Polymerase Chain Reaction (PCR) to allow for proper removal of the protein by enterokinase, which is known to not cut efficiently near proline residues. DNA coding for Met-Pro in the first two amino acids was replaced with DNA coding for Ala-Met-Ala. The resulting gene was cloned into the pET32a vector downstream of the T7 promotor, utilizing the restriction sites *NcoI* and *HindIII* and allowing the final construct to contain a Thioredoxin tag (to allow for greater protein expression) and 6x-His tag (to allow for later purification via a Nickel affinity column) followed by an enterokinase cutting site (to later cleave off this tag; see Appendix A, Figure A-1.) The plasmid was transformed into *Escherichia coli* BL21(DE3) (Novagen) competent cells and expressed in either Luria broth (containing naturally abundant ^{14}N, so that all protein produced will contain ^{14}N) or minimal media with $^{15}NH_4Cl$ as the sole nitrogen source (so that all produced protein will contain ^{15}N; see section 2.2.2 for significance of the nitrogen isotope on NMR spectroscopy). Protein production was induced when the absorbance at 600 nm reached 0.70-0.75, meaning the cells had reached an optimal confluency[130], with the addition of 1mM (final concentration) isopropyl β-d-1-thiogalactopyranoside (IPTG) and incubated with shaking at 22°C for 20 h. In this expression system, the gene responsible for the RNA polymerase that transcribes the gene of interest (the vCCI construct in this case) is blocked off by transcriptional repressor, which can be removed from the gene upon lactose binding[131]. IPTG, a non-hydrolyzable lactose analog, can bind to the repressor for the RNA polymerase gene, causing the repressor to be removed and allowing the highly-active RNA polymerase to be expressed[132]. Once expressed, the RNA polymerase can bind to the promotor region of the gene of interest (in this case, the vCCI construct) and transcribe this DNA, leading to protein production. After expression of the vCCI protein was induced in this manner, the cells containing the expressed protein were harvested by centrifugation at 4200×g, 4°C for 12 min and supernatant was discarded.

(See Appendix A, Figure A-2 for a general overview of the following purification process.) The cell pellet was resuspended in 6 M guanidine hydrochloride, 200 mM NaCl, and 50 mM Tris (pH 8.0) and was lysed by three passages through a French press and then centrifuged at 27000 × *g* for 1 h to collect all proteins in the supernatant. Guanidine hydrochloride is a denaturant that causes all proteins in solution to lose their folded, tertiary structure, thus losing all their activity, as well as become soluble[133]. As many human proteins will be incapable of folding properly in *E. coli* upon expression (see section 5.2.1), unfolding all proteins in the cell helps prevent damage to the protein of interest by rendering natural *E. coli* protease enzymes inactive[134]. The supernatant was decanted and 15 mM beta-mercaptoethanol (βME) was added and allowed to stir at

room temperature for 2 hours to reduce. As both thioredoxin[135] and vCCI[136] contain cysteines, which are capable of non-specific disulfide bonding[137], this quick reduction was an attempt to help expose the 6xHis tag found at the center of the construct for easier access during the nickel chelating column.

The 6x-His tag can form a complex with nickel, allowing for selective binding of the protein of interest to a nickel chelating chromatography column and purification away from many other natural *E. coli* proteins[138] However, as nickel columns may react with reducing agents[138], particularly if free nickel is still available, unbound to the beads of the column, a 0.3 M imidazole wash was performed on the nickel chelating column (Quiagen) to remove any of unbound nickel. After several washes of the column with resuspension buffer (also containing 15 mM βME to maintain reduced conditions) to remove any imidazole and to equilibrate the column, the lysed protein solution was then loaded onto the nickel chelating column. The column containing the bound vCCI was washed with 10 column volumes of resuspension buffer containing 15 mM βME, and then with 10 column volumes of Wash Buffer (6 M Guanidinium chloride, 200 mM NaCl, 15 mM βME, 80 mM NaOP, pH 7.2) to remove nonspecific binding. Proteins were eluted from the column using 6 M guanidine hydrochloride, 200 mM NaCl, and 60 mM NaOAc, pH 4, as histidine has a pK_a of 6 and becomes positively charged when protonated[139], repelling away from the positively-charged nickel. Fractions containing the eluted protein were identified by absorbance at 280 nm (as aromatic amino acids such as tryptophan tend to absorb light at 280 nm[140]) and then pooled together. βME was then added to a final concentration of 25 mM and the fractions were allowed to stir for one hour at room temperature, followed by stirring for 12 hours at 4°C overnight to ensure complete reduction of all disulfides.

The protein was then refolded by dropwise addition to 20x volume of ice-cold refolding buffer (9.6 mM NaCl, 0.4 mM KCl, 2 mM $CaCl_2$, 2 mM $MgCl_2$, 550 mM L-Arginine hydrochloride, 400 mM sucrose, 3 mM reduced glutathione (GSH), 0.3 mM oxidized glutathione (GSSG), 50 mM Tris, pH 8), and then allowed to stir 24 h at 4°C for a slow refold, to refold all proteins in the solution. The solution was dialyzed 4 times into 4 liters 200 mM NaCl, 2 mM $CaCl_2$, 20 mM Tris, pH 7.4 buffer (ideal Enterokinase cutting conditions) at 4°C, to allow for cleavage of the fusion tag from vCCI.

To cleave the thioredoxin fusion tag from the purified protein, the samples were incubated for 12 hours at 4°C with 650 nM of the protease enterokinase, which cuts selectively after the protein sequence DDDDK[141]. The samples were then dialyzed 4 times into 4 L of 20 mM Bis-Tris, 50 mM NaCl pH 7.1 to prepare the sample for passage through an anion exchange column, which contains positively charged beads in order to bind negatively-charged proteins[142]. The pI of vCCI is 4.5, while the pI of the cleaved protein tag is 5.5, so by pH 7.1, both proteins would be negatively charged, with vCCI containing a more negative charge and thus being more difficult to remove from an anion exchange column at this pH. In order to prevent damage to the column, the protein solution was passed through a 0.2 µm nylon filter and then be purified on a HiTrap™ Q HP Column (GE Healthcare Life Sciences) on a High Pressure Liquid Chromatography (HPLC) system, using a gradient from 50 mM NaCl, 20 mM Bis-Tris pH 7.1 to 1 M NaCl, 20 mM Bis-Tris pH 7.1 to separate the cleaved tag (the earlier peak) from vCCI (the later peak). The fractions were analyzed on an SDS-PAGE gel to confirm purity and then fractions containing vCCI were concentrated using the Amicon concentrators (Millipore,

Billerica, Massachusetts) and buffer was changed to 100 mM NaCl, 20 mM NaOP pH 7.0 with 0.02% NaN_3 for NMR (see section 2.2.2) and ITC studies (see section 2.2.3).

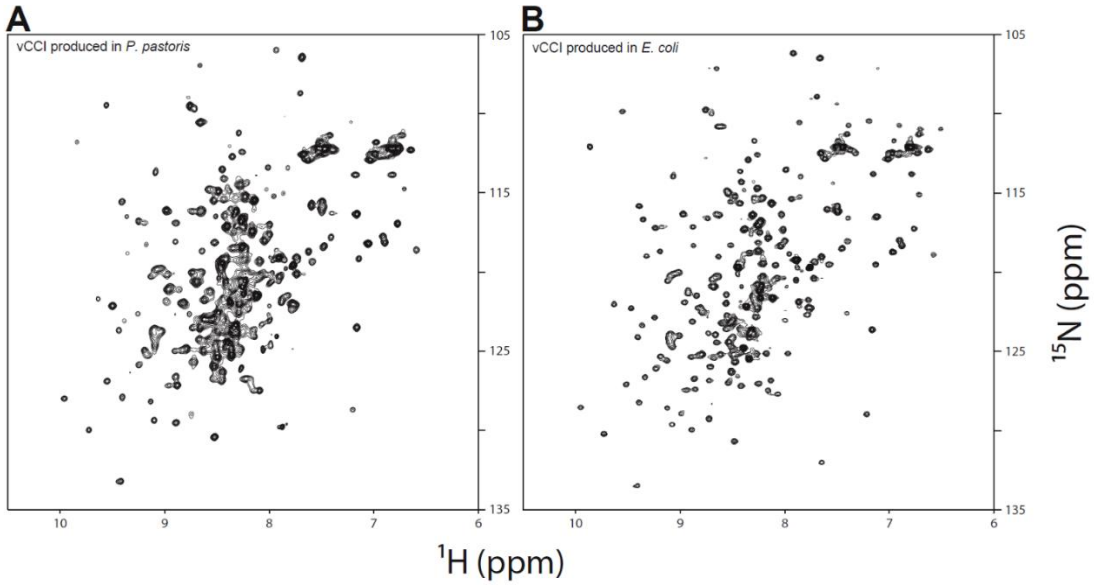

Figure 2.1: Comparison of unbound vCCI produced in yeast and *E. coli*. (**A**) Uniformly side-chain deuterated $^{2}H/^{13}C/^{15}N$-labeled vCCI produced in *p. pastoris* yeast. Spectrum was measured in 100 mM NaCl, 20 mM NaOP pH=7.0, at 37°C, sample taken from [129]. (**B**) ^{1}H-^{15}N HSQC spectra of ^{15}N-labeled vCCI produced in *E.coli*, measured under the same conditions as in (A).

This new method of purification from *E. coli* rather than through yeast resulted in cleaner peaks on the NMR with less signs of unfolded peaks near the center even with the yeast NMR sample being deuterated, suggesting a more uniformly folded protein product than had been available previously (Figure 2.1). See section 2.2.2 for a more detailed explanation of NMR spectroscopy.

Expression and purification of vMIP-II*:* The gene for vMIP-II was placed into a pET28a vector with a 6x-His tag and SUMO fusion tag and transformed into *Escherichia coli* BL21(DE3) (Novagen) competent cells. The cells were then grown in either minimal media with $^{15}NH_4Cl$ as the sole nitrogen source for ^{15}N-labeled samples or Luria Broth for ^{14}N-labeled samples. Protein production was induced by adding IPTG to a 1 mM concentration and incubated with shaking at 37°C for 5 h.

(See Appendix A, Figure A-3 for a general overview of the following purification process). The cell pellet was resuspended in 6 M guanidine hydrochloride, 200 mM NaCl, 10 mM benzamidine, 50 mM Tris (pH 8.0) and were lysed by French press and then centrifuged at 27000 × *g* for 1 h to remove protein-containing supernatant from cell components. The soluble portion was then loaded onto a nickel chelating column (Qiagen) equilibrated with the resuspension buffer, and then washed with 10 column volumes resuspension buffer and 10 column volumes of Wash Buffer (6 M Guanidinium chloride, 200 mM NaCl, 15 mM βME, 80 mM NaOP, pH 7.2). Proteins were eluted from the column using a pH gradient with 6 M guanidine hydrochloride, 200 mM NaCl, 60 mM NaOAc (pH 4). These pH 4 fractions were then pooled together and had 10 mM βME added to then stir for 2 hours at room temperature for reduction. As this protein contains only two possible disulfide bonds, less reduction time is necessary compared to a

protein such as the vCCI construct above. The proteins were then refolded by dropwise addition into 10x volume of refolding buffer (550 mM L-Arginine Hydrochloride, 200 mM NaCl, 1 mM EDTA, 1 mM reduced glutathione (GSH), 0.1 mM oxidized glutathione (GSSG), 50 mM Tris, pH 8), and then allowed to stir overnight at 4°C to continue refolding. The solution was dialyzed three times into 4L of 200 mM NaCl, 20 mM Tris pH 8 buffer at 4°C, ideal buffer for cleavage by the protease, ULP-1.

To cleave the SUMO fusion tag from the purified protein, the samples were incubated for 12 hours with 100 nM of the protease ULP-1. The protein solution was then dialysed into 80 mM NaCl, 20 mM Tris pH8 buffer and centrifuged to remove precipitated material. The solution was then added onto a second nickel chelating column (Qiagen), with the cleaved fusion tag (composed of the 6xHis tag and SUMO tags) binding to the column and the flow-through (containing the cleaved vMIP-II) being collected. The samples were then acidified with Trifluoroacetic acid (TFA) to a pH of 1-2 and loaded onto a C_4 reversed-phase chromatography column (Vydac, Hesperia, CA). While most chromatography utilizes hydrophilic resins (the "stationary phase", as this resin will immobilize proteins moving across it), reverse-phase chromatography instead utilizes hydrophobic resin, so proteins with a higher affinity for hydrophobic resin will bind rather than proteins with a high affinity for hydrophilic groups.[143] This is a useful technique, as it leaves proteins in a water or acetonitrile buffer which are ideal of freeze-drying (and thus being simple to afterwards add to any solution or buffer without use of dialysis or buffer exchange techniques). However, not all proteins are capable of passing over a reverse-phase column without unfolding, which is why vCCI was left in buffer during its purification profess, as previous attempts have proven the column renders the protein unusable. vMIP-II was then eluted off the column using an acetonitrile/water gradient with 0.1% trifluoroacetic acid (TFA) in both buffers. The fractions were analyzed on an SDS-PAGE gel to confirm purity and lyophilized on a Labconco freeze-dry system, so the protein could then be easily added to necessary buffers for techniques like NMR (see section 2.2.2) or ITC (see section 2.2.3).

Protease purifications*:* The proteases used in these purifications were produced and purified in our laboratory as briefly described: ULP-1 protease or Enterokinase protease were expressed in LB medium using a pET-28a vector and the cells were collected through centrifugation, resuspended in buffer and then French pressed. The ULP1 protease from the supernatant was purified using a nickel chelating column as previously described[144]. The enterokinase was found in the inclusion body and resuspended in 6M guanidinium buffer before being purified using a nickel chelating column. Enterokinase was then dialyzed into low-salt buffer to allow for refolding and tested for activity through self-cleavage of the fusion tag (manuscript currently in preparation by P. LiWang group).

Fluorescent Anisotropy Proteins Purifications*:* Proteins used in fluorescence anisotropy studies were purified as specified in [41] by Nai-Wei Kuo; methods including protein purification were kept consistent with previous publications in order to maintain consistency with the proteins used for quantification of K_d values with the previous publication. vCCI for these experiments were made using a gene encoding rabbitpox virus vCCI cloned into pPIC9K plasmid and then transformed into *Pichia pastoris* strain SMD1168 (Invitrogen) and purified as previously described[46] by Nai-Wei Kuo. (See Figure 2.1 for a comparison of the vCCI purification results.) Eotaxin-K63C for Fluorescent Anisotropy experiments (which utilize a competitive binding technique,

where fluorescently-labeled Eotaxin-K63C binds to vCCI and it then outcompeted off by another, non-labeled chemokine, and the change in anisotropy is measured over time as the concentration of competing chemokine increases) was also purified as previously described[46] by Nai-Wei Kuo to further maintain consistency with previously-published results.

2.2.2 NMR Spectroscopy

Nuclear magnetic resonance (NMR) is a natural phenomenon in which certain nuclei absorb and re-emit electromagnetic radiation when within a magnetic field[145]. When placed in a magnetic field, NMR-active nuclei absorb electromagnetic radiation at a frequency characteristic of the specific isotope; this property is known as the nuclei's "spin" and is characterized by its nuclear quantum spin number, *I*.[146] Nuclei with odd mass numbers have half-integral spin quantum numbers, nuclei with an even mass number and an even atomic number have spin quantum numbers equal to zero (and thus are invisible to NMR spectroscopy), and nuclei with an even mass number and an odd atomic number have integral spin quantum numbers (though NMR resonance lines for nuclei with quantum spin numbers greater than 1/2 are correspondingly broad and can thus be more difficult to study)[147]. Thus, nuclei with spin quantum numbers of 1/2 such as ^{1}H, ^{13}C, ^{15}N, ^{19}F, and ^{31}P are some of the most convenient biologically-relevant nuclei to study via NMR.

I=1/2 nuclei give rise to two energy levels in a magnetic field, which are characterized by the magnetic quantum number (m). This value is restricted to $-I$ and I and is quantized, meaning that in *I*=1/2 nuclei, there are two possibilities: 1/2 (also called "spin up" or α) and -1/2 (also called "spin down" or β).[148] Nuclei with a nonzero spin angular momentum (*I*) also possess a nuclear magnetic moment (μ), which is defined by:

$$\mu = \gamma \boldsymbol{I} \qquad \text{Equation 2.1}$$

where γ is the magnetogyric ratio, a ratio of magnetic moment (μ) and the angular momentum (*I*)[149]. The magnetogyric ratio is a constant for any given nucleus. Because of these nuclear properties, in the presence of a magnetic field (**B**), the spin states of a particular nucleus will have energy given by[147]:

$$E = -\mu \cdot \mathbf{B} \qquad \text{Equation 2.2}$$

The minimum energy is obtained when the projection of the nuclear spin angular moment onto the magnetic field is maximized. But in the presence of a static magnetic field, such as that used in NMR spectroscopy, the projections of angular momentum results in 2*I*+1 equally-spaced energy levels (so 2, in the case of *I*=1/2 nuclei). Because one state is at a lower energy level, the populations are not equally populated at equilibrium.[147] Transitions between these energy levels can be stimulated by applied electromagnetic radiation, and the sensitivity of NMR spectroscopy depends on the population differences between these states.[148]

In the absence of other magnetic fields, the bulk magnetization of a sample precesses at a rate known as the *Larmor frequency* (ω_0) around the main static magnetic field (B_0), which relate to each other through[150]:

$$\omega_0 = -\gamma \boldsymbol{B}_0 \qquad \text{Equation 2.3}$$

The magnitude of this precessional frequency is identical to the frequency of electromagnetic radiation required to excite transitions between energy levels. However, at thermal equilibrium, the bulk magnetization vector is collinear with the static field, and thus, no signal is produced in the NMR spectroscopy coil.[148] The key to producing an NMR signal is to disturb this equilibrium state of the nuclei being studied. In pulsed NMR experiments, a short burst of radiofrequency electromagnetic radiation (known as a "pulse") displaces the bulk magnetization from the equilibrium. After the pulse ends, the bulk magnetization will not, in general, be parallel to the static field; instead, the bulk magnetization will precess around the static field with an angular frequency that will generate a detectable signal in the coil, which can then be detected until the bulk magnetization relaxes back to their original position, aligning back with the static magnetic field.[151] (See Appendix A, Figure A-4.)

The intramolecular magnetic field around an atom within a molecule changes the resonance frequency based on the specific chemical environment the nucleus is found in; thus, being able to read this information through NMR spectroscopy can give access to details of the structural properties of a molecule and its individual functional groups. However, since proteins tend to be larger macromolecules, using ^{1}H NMR will often give too much information to easily parse through. To obtain a more readable spectrum, two dimensional NMR can be used, which can give information on two nuclei that have a relationship with each other[152].

Conventional one-dimensional NMR spectra are plots of intensity vs. frequency; in two-dimensional spectroscopy intensity is instead plotted as a function of two frequencies that are related to one another (usually through a chemical bond)[153]. The two dimensional technique used in the experiments within this dissertation is known as heteronuclear single quantum coherence (HSQC) spectroscopy, where "heteronuclear" refers to nuclei other than ^{1}H. As these nuclei must be visible on an NMR spectrum, the naturally-abundant ^{12}C and ^{14}N in organic molecules will not be visible. As such, proteins produced for NMR spectroscopy must be isotopically labeled with ^{13}C or ^{15}N to be visible in such spectra, which is why certain proteins were produced in ^{15}N-containing media in section 2.2.1. In theory, the heteronuclear single quantum correlation has one peak for each ^{1}H bound to a heteronucleus (usually ^{13}C or ^{15}N)[154]. This technique uses a signal resolution enhancement method known as INEPT (Insensitive Nuclei Enhanced by Polarization Transfer); this method involves the transfer of nuclear spin polarization from spins with large population differences (such as ^{1}H) to nuclear spins of interest with low population differences (such as the heteronucleus) via indirect dipole-dipole coupling.[155] Indirect dipole–dipole coupling (commonly called J-coupling) is an indirect interaction between two nuclear spins that arises from the interactions between the nuclei and local electrons. J-coupling contains information about bond distance and angles, thus

providing information on the connectivity of molecules.[156] The net effect is the non-selective polarizarion transfer from ^{1}H to another nucleus (such as ^{15}N, in this example) with the appropriate coupling, creating a peak for every ^{1}H-^{15}N bond in a molecule.

As each amino acid in a protein contains a N-H bond within the backbone (with the exception of proline, where the sidechain cyclizes back into the backbone, resulting in no proton on the nitrogen), a ^{1}H-^{15}N HSQC spectrum can offer a "fingerprint" of a protein, with each point in the spectrum correlating to each visible amino acid backbone residue[153]. If the backbone residues are in a consistent environment (as they would be in a folded protein, as proteins adopt one consistent lowest energy state when folded properly, with the exceptions of intrinsically disordered[157] or metamorphic proteins[158]) there should be a clear peak for every residue; in cases of non-homogeneously-folded samples, unfolded peaks tend to arise in the center of the spectrum, resulting from the overall averaging of the unfolded peaks being found in different, non-consistent positions. Side chains such as tryptophan, asparagine, and glutamine, which contain nitrogen in their sidechains, also give rise to additional signals, though these signals are usually found in specific parts of the spectrum[159] (see Appendix A, Figure A-5). Analysis of the ^{15}N-HSQC allows researchers to evaluate whether the expected number of peaks is present and thus to identify possible problems due to multiple conformations or sample heterogeneity. However, it is not possible to assign peaks to specific atoms from the heteronuclear single quantum correlation alone; more complex experiments must be used to know which backbone peak correlates to which amino acid. However, as the unbound spectra for both vCCI[136] and vMIP-II[160] have already been previously assigned, these previous "fingerprints" are used to assign the unbound peaks in this work.

All NMR samples were run in 20 mM sodium phosphate buffer, 100 mM NaCl with 10% D_2O (Added to allow the sample to "lock" onto the deuterium frequency to compensate for natural drift and inhomogeneity of the magnetic field during signal acquisition), 5 µM 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) (the reference compound; as the chemical shift of nuclei changes in relation to the strength of the magnetic field, a reference compound must be used to keep spectra consistent, setting the position of the reference compound at "0" ppm), and 0.02% NaN_3 (a preservative that will show no NMR signal during HSQC), with a final pH of 7.0. ^{15}N-labeled, lyophilized vMIP-II was first resuspended into 5 mM NaOP buffer, pH 2.8 in order to dissolve the protein (as chemokines tend to aggregate when introduced suddenly to neutral pH), and then 10 µL was added to 340 µL NMR buffer (either alone or containing 150 µM ^{14}N-vCCI) for final vMIP-II concentration of 50-60 µM. ^{14}N-vMIP-II was also resuspended into 5 mM NaOP, pH 2.8, and 10 uL was added to 340 µL NMR buffer containing ^{15}N-labeled vCCI, for a final concentration of 150 µM VMIP-II in the sample. vCCI was exchanged into NMR buffer as explained in the above section 2.2.1, with final concentrations for ^{15}N samples being 50-60 µM. This resulted in two unbound, ^{15}N-labeled samples, one each for vCCI and vMIP-II (Figures 2.3A,C) and two samples where the proteins were bound to each other (Figures 2.3B,D), though their partner in these samples were ^{14}N-labeled, and thus not detectable on the HSQC, meaning that the spectrum should correlate only to the residues in the ^{15}N-labeled proteins in the samples.

All HSQC NMR data were acquired on a four-channel 600 MHz Bruker Avance III spectrometer equipped with a GRASP II gradient accessory and a TCI cryoprobe with an

actively shielded Z-gradient coil. Spectra with ^{15}N-labeled vMIP-II were measured at 25 °C; spectra with ^{15}N-labeled vCCI were measured at 37°C (as larger proteins tend to tumble more slowly, thus an increase in temperature helps sharpen signal[161]). The chemical shift was referenced relative to internal DSS[162]. The data were processed using NmrPipe[163] and analyzed using PIPP[164]. For HSQC spectra, sweep width was 8474.576(^{1}H) and 1766.784 Hz (^{15}N), with 1280 points in ^{1}H and 128* (256 total) points in ^{15}N.

Once the spectra were obtained (see Figure 2.3), the bound versus unbound spectra were compared to one another via changes in chemical shift perturbation. Chemical shift is the resonance frequency of a particular nucleus; whenever a change to the chemical environment that a particular nucleus is located in changes, such as interaction with another molecule, there is change in the chemical shift, as well[161]. Thus, tracking the changes between an unbound and bound form of a protein can suggest residues that may be involved in the binding of a ligand, either through directly being involved in the binding, or through propagating small changes to larger changes in other parts of the protein, known as an allosteric effect[165]. However, it is important to note that because ^{15}N-HSQC only shows the backbone N-H interactions and a very limited number of N-H containing side chains, this technique may miss certain shifts if they are only caused within the side-chain without shifting the backbone amide.

Chemical shift perturbations can give certain information about the binding of a protein depending on the rate of exchange between the binding protein and its ligand. The utilization of chemical shift perturbations is more generally useful in cases of fast exchange (weaker binding to the ligand), as the gradual addition of ligand allows one to trace the shifting peaks from the unbound position to the bound position. In cases of slow exchange (tight ligand binding), the peak splits between bound and unbound states, allowing for no simple way to trace between an unbound peak and the bound peak[166] (see Appendix A, Figure A-6.) In the case of vCCI and vMIP-II, the binding appears to be quite tight, as the chemical shift perturbations result in a splitting of peaks as more ligand as added, rather than a gradual drift towards the fully-bound position, making it difficult to accurately assign a bound peak to an unbound original peak with perfect confidence. In such situations, where re-assignment of the entire bound spectrum is not possible for the time being, a “minimal chemical shift procedure” may be used, in which assigned peaks in the free protein are linked to the closest peaks (in chemical shift) in the complex[166,167]. This associates each peak with the minimal shift change needed to produce the bound spectrum and may therefore miss some very large chemical shift changes but should not introduce new errors.

As such, the weighted average chemical shift change of the ^{1}H and ^{15}N resonances for each residue in vCCI and vMIP-II upon binding was calculated using equation 2.4[168]:

$$\Delta\delta_{obs} = \sqrt{\frac{\Delta\delta_H{}^2 + \left(\frac{\Delta\delta_N}{5}\right)^2}{2}} \qquad \text{Equation 2.4}$$

where $\Delta\delta_H$ and $\Delta\delta_N$ are the chemical shift changes of the ^1H and ^{15}N dimensions, respectively (with the N chemical shift being divided by a factor of five, as the nitrogen dimension tends to be roughly five times that of the proton dimension). Here, the $\Delta\delta_{obs}$ is the difference between bound and free form of ^{15}N-labeled complexes. As mentioned, due to lack of ^{13}C labeling, bound peak identifications were estimates; to be conservative the nearest residue without a clear origin was assumed to belong to the residues in question, leading to likely underestimates in chemical shift changes in Figure 5A, C.

2.2.3 Isothermal Titration Calorimetry (ITC)

Isothermal Titration Calorimetry (ITC) is a technique that can be used to determine thermodynamic properties of molecules in a solution upon interaction between two or more ligands[169]. This technique can directly give the values for enthalpy (ΔH), the association constant (K_a) and the stoichiometric constant (n) by converting the integrated heat of injection data into a Wiseman plot[170] (explained further below). These variables can then be used to find other thermodynamic variables through the relationship:

$$\Delta G = -RTlnK_a = \Delta H - T\Delta S \qquad \text{Equation 2.5}$$

where R is the gas constant, T is the temperature, ΔG is Gibbs Free Energy, and ΔS is the entropy[171].

The ITC is composed of two identical cells and equipped with sensitive heat-detecting circuits that can detect very small temperature changes between a sample cell (which contains the macromolecules whose thermodynamic properties are being explored) and a reference cell (filled with buffer or water to act as a control for temperature changes.) As ligand is titrated into the sample cell (causing an exothermic or endothermic reaction, and thus, heat release or absorption to or from the aqueous environment), the equipment uses power to compensate and maintain constant temperature with the reference cell.[169] The observations of these power changes are then plotted against time, resulting in a spike of heat for every injection (See Appendix A, Figure A-7A). The spikes are then integrated with respect to molar ratio, giving total heat exchanged per injection (See Appendix A, Figure A-7B). In this graph (known as a Wiseman Plot), the height of the curve gives ΔH, the inflection point gives n, and the slope of the inflected area of the curve gives K_a[170], allowing for the utilization of equation 2.5 above to calculate other thermodynamic variables.

To begin these experiments, a baseline of “buffer into buffer” was taken in order to ascertain whether there was any thermal component purely due to buffer mixing, so that it could be subtracted from final experimental data. A Nano ITC Low Volume Isothermal Titration Calorimeter (TA Instruments) reference cell was flushed with degassed water three times and 300 µL degassed water was finally added as a reference. Flushing aids in removal of bubbles before final solution is added[172]. Usually, when a solution is heated, gas bubbles may form as the solubility of dissolved gases is decreased with the increasing temperature. If gas bubble formation occurs in the ITC cells during the run, the resulting data will be rather noisy due to the abrupt changes that result from the liquid displacement effects of the bubbles[173]. The sample cell was flushed with degassed buffer (20 mM NaOP, 100 mM NaCl, pH 7.0) three times and 300 µL of degassed 20 mM NaOP 100 mM NaCl, pH 7.0 was loaded into the sample cell (the minimal volume for the nano ITC instrument). Slightly over 50 µL (to allow for a slight

excess in the syringe) of the same degassed buffer (20 mM NaOP, 100 mM NaCl, pH 7.0) was loaded into the syringe to allow for 20 separate 2.5 µL injections at 300 second intervals, with a 350 rpm stirring speed (appropriate for gold cells). Stirring too slowly may result in broad titration peaks that require longer intervals between injections. Reactions that have higher enthalpies (as was expected with vCCI and vMIP-II) can be run at faster stirring speeds, which then allow shorter injection intervals[172]. The interval time must be long enough to allow for the heat change to return back to equilibrium after an injection (which in this case, 300 seconds was more than long enough, but this time can be adjusted for shorter experiments or in binding events that result in wider heat spikes). Once the machine was set up, it was given 1 hour to equilibrate to assure a stable reading, and the experiment was run with the aforementioned conditions (identical to the conditions to be used in the true experiment later).

Solutions containing the proteins vCCI and vMIP-II were then degassed, as the proteins are known to be stable under various conditions; less stable proteins should forego the degassing, should there be structural damage caused to the proteins during the process, and instead be added as powder directly to degassed solution if necessary (though this may cause a possible pH change that should be monitored for upon protein addition[172]). Once samples were prepared, the sample cell was flushed with degassed buffer (20 mM NaOP, 100 mM NaCl, pH 7.0) three times and 300 µL of 10 µM vCCI in 20 mM NaOP, 100 mM NaCl, pH 7.0 was added to the sample cell. Slightly over 50 µL (again, to allow for slight excess) of 100 µM vMIP-II in vCCI in 20 mM NaOP, 100 mM NaCl, pH 7.0 was brought up into the syringe and loaded onto a machine. Ligand concentration should be kept roughly 10-fold higher than macromolecule concentration[174], but lower concentrations of vCCI made discerning heat changes difficult. Stirring was set to 350 rpm and the machine was allowed to equilibrate for an hour to assure a stable thermal reading. Once stable, 20 separate 2.5 µL injections at 300 second intervals were performed and the data was collected (see Appendix A, Figure A-7A). The experiment was repeated three separate times.

The buffer-into-buffer blank was subtracted from the data, and peak-by-peak manual integration was performed using NanoAnalyze software (TA Instruments). The data was fit to the one-site independent binding site model[174], which was provided by the software. (See Appendix A, Figure A-7B). The K_d was found to be below detectable limits (10-10 nM) (Table 2.1, below), confirming a very tight binding between vCCI and vMIP-II. Typically, when trying to find optimal conditions for an ITC experiment, the equation:

$$c = K_a[M]n \qquad \text{Equation 2.6}$$

(where the experimental binding isotherm can be characterized by the unitless value c, which is the product of the association constant, K_a, the concentration of macromolecule [M] (vCCI in this case), and the stoichiometry of the reaction, n) can be used to determine if the concentration for the macromolecule is correct[170]. A value between 1 and 1000 is recommended[174], meaning that assuming an n of 1, with K_a of over 1^{10}, the concentration of vCCI would have to be at most 0.1 µM for an accurate result. However, when the concentration of vCCI was brought down to 1 µM, the experimental data was too close to noise to easily discern peaks. This suggested that a slightly different

experiment would be needed to determine an accurate K_d for the binding of vCCI and vMIP-II. Although it would have been possible to quantify on the ITC by utilizing a competition experiment[175] (such as incubating vCCI with a different chemokine of known, but higher K_d and then seeing how much vMIP-II is necessary to outcompete the weaker-binding chemokine off of vCCI), because our laboratory has already quantified the dissociation constants of several chemokine variants complexed vCCI utilizing fluorescent anisotropy, this technique was used again to maintain consistency with previously-published results[41] (see section 2.2.4).

Table 2.1: Results of isothermal titration calorimetry, titrating vMIP-II into a solution containing vCCI

ΔH (kJ/mol)		ΔS (J/mol)		ΔG (kJ/mol)		n		K_d (M)	
Avg[1]	SD	Avg[1]	SD	Avg[1]	SD[2]	Avg[1]	SD	Avg[1,3]	SD[2]
-51.83	0.85	17.61	2.87	-57.08	NM	0.759	0.067	≤1.00E-10	NM

[1]The experiment was done in triplicate.

[2]NM: not meaningful. The values were the same, thus there was no apparent standard deviation.

[3]1.00E-10 is the low limit of K_d detection for the ITC.

2.2.4 Fluorescence Anisotropy

Fluorescence anisotropy experiments were very kindly carried out by Nai-Wei Kuo with data processed by Cindy DePuer. Fluorescence anisotropy is the phenomenon where the light emitted by a fluorophore (a molecule absorbs a photon and gets excited to a higher energy state, and then releases the energy in the form of heat and another photon at a different frequency) has unequal intensities along different axes of polarization[176]. Fluorescence anisotropy can be used to measure the binding constants and kinetics of reactions that cause a change in the rotational time of the molecules (due to their change in total size once bound together). If the fluorophore is bound to a small molecule, the rate at which the small molecule tumbles can decrease significantly when it is bound tightly to a large protein, resulting in a change in measurable anisotropy. The degree of binding is calculated by using the difference in anisotropy of the partially bound, free, and fully bound (large excess of binding protein) states measured by titrating the two binding partners together. The larger (bound) product has a lower rotational diffusion coefficient and a higher fluorescence anisotropy, so as the fluorescently-labeled ligand binds, there should be an increase in fluorescence anisotropy.[177]

Once the binding constant of the fluorescent ligand (in this case, eotaxin-K63C, which is well-characterized in previous studies[41]; the mutated cysteine residue is capable of binding to the fluorophore fluorescein-5-maleimide) to the binding partner (vCCI) is known, a competition assay using unlabeled ligand can be used. Fixed concentrations of the tagged molecule bound to its binding partner have the untagged ligand (vMIP-II) titrated in at specific concentrations to measure the change in fluorescent and apparent binding affinity of the fluorescently-labeled ligand, which can be used to extrapolate the binding constant of the untagged ligand.[178]

The experiments were done in three independent experiments as described in [41], at 25 °C and pH of 7.0 utilizing a PC1 spectrofluorimeter and VINCI software (ISS, Champaign, IL), with an excitation wavelength of 497 nm and emission wavelength of 524 nm. The obtained data were then fit to a system of mass conservation equations as well as the following equation 2.7:

$$\theta = \frac{[L]_{free} \times K_a}{1 + [L]_{free} \times K_a} \qquad \text{Equation 2.7}$$

where θ is the fraction of bound eotaxin-K63C, $[L]_{free}$ is the concentration of unbound vCCI, and K_a is the association constant for the complex.

For the competitive binding experiment, a 1:1 ratio of the vCCI·eotaxin-K63C (with eotaxin-K63C labeled with fluorescein-5-maleimide) complex was prepared at a concentration of 8 nM. 500 µl of this complex was then mixed with varying amounts of unlabeled vMIP-II and incubated 30 minutes at 25 °C. Anisotropy measurements were taken and the values were normalized so that 1 represents the 100% bound state. The resulting data were fitted using Scientist software (Micromath, Salt Lake City, UT) to a system of equations described previously[179,180] (Figure 2.4).

2.2.5 Molecular Dynamics

All Molecular Dynamics simulations were very kindly performed by Mike Colvin. Simulations were performed in 11.5^3 nm^3 cubes and contained about 48000 water molecules, using a cubic cell so that atoms leaving one edge reappear on the opposite side of the box. All-atom molecular dynamics was performed on three vCCI-MIP complexes. All three complex structures were based on the NMR structure of the VCCI:MIP-1β complex (PDB code 2ffk), which has three mutations from the wild type MIP-1β sequence, (positive residues in ^{45}KRSK^{48} were replaced by Ala to allow for solubility, resulting in the mutant MIP-1β-^{45}AASA^{48}). The original wild type structure was recreated by in silico editing of the 2ffk experimental structure. The vCCI:vMIP-II starting structure was created from the vCCI:MIP-1β structure by computationally superimposing the experimental vMIP structure (PDB code 1vmp) on the MIP-1β chain to minimize the average difference between the corresponding Cα backbone atoms (Figure 2.2). The net charge (-11 in the vCCI:vMIP-II complex, -26 in the vCCI:MIP-1β mutant complex, -23 in the vCCI:MIP-1β wild type complex) on the complexes was neutralized by adding Na^+ ions and additional Na^+/Cl^- pairs (~60) were added to yield an ion concentration of approximately 70 millimolar. After short equilibration runs, a full 1 microsecond of MD simulation was run using Gromacs 5.0.7[181–183] using the NPT ensemble, the Verlet cutoff scheme and a 2 fs timestep. All bonds to hydrogen were constrained to their equilibrium length using the LINCS algorithm[184]. Temperature was maintained at 300K using the Bussi et al. thermostat[185] and pressure at 1 bar using the Parrinello-Rahman barostat[186]. The simulations were performed using the AMBER99SB-ILDN force field for the protein[187] and the TIP3P water model[188].

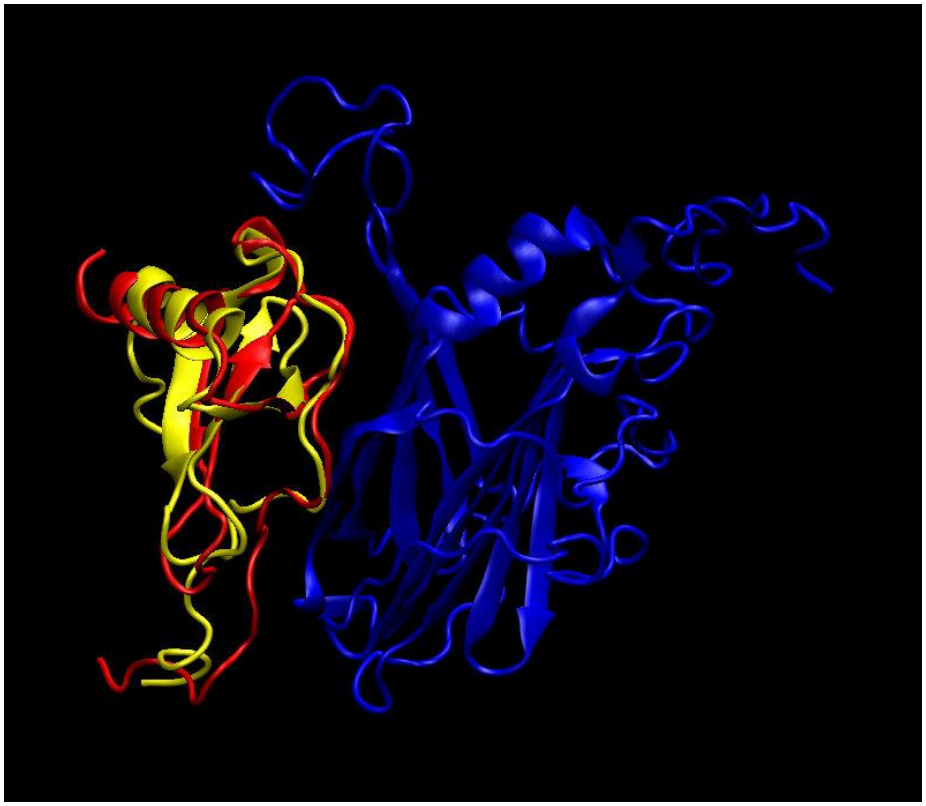

Figure 2.2: vCCI:vMIP-II starting point for MD simulations. The vCCI:vMIP-II trajectory was based on the vCCI:MIP-1β-^{45}AASA^{48} structure (PDB 2ffk; ^{46}) with vCCI in blue, but the MIP-1β (red) was replaced with the experimental structure of vMIP-II (PDB 1vmp) (yellow), superimposed to minimize the difference in positions of the backbone atoms.

2.2.6 Figure Preparation

All structure figures were prepared by using UCSF Chimera[20], with help from Laura Showalter.

2.3 Results

2.3.1 Folded vCCI can be produced from E. coli.

Despite the interest in vCCI, due to its chemokine-binding affinity and use as a possible biotherapeutic, it has been relatively time consuming to produce *in vitro*, as bacterial expression results had failed to produce properly-folded protein. High expression of proteins from this family has been described from yeast[41,46,123], baculovirus[48,49], and an Fc-linked vCCI was produced from 293T mammalian cells[47]. Each expression system is useful, but the lack of an *E. coli* expression protocol has limited study of the protein and, in particular, limited the ability of investigators to easily make a wide range of mutants. Now, we have developed a technique (see Appendix A, Figure A-2) to express and purify vCCI from *E. coli* that results in successfully refolded protein comparable to previous results (Figure 2.1). Briefly, the cells are disrupted in 6 M guanidine hydrochloride at pH 8.0 under reducing conditions, and purified by his-tag

affinity chromatography, using a nickel chelating column. After further reducing agent is added, the solution is slowly added to 20x volume of a cold low-salt refold buffer containing L-Arginine, sucrose and glutathione, and incubated for 1 day. The solution is then dialyzed at pH 7.4 followed by addition of a protease to allow cleavage of the fusion tag. Final purification is carried out on an anion-exchange column (see Materials and Methods section 2.2.1 for details).

The final product of refolding and purification results in a ^{15}N-HSQC spectrum of vCCI that is essentially identical to that produced from *P. pastoris* yeast expression (Figure 2.3A and Figure 2.1). Further, this protein forms a complex upon addition of chemokine analog vMIP-II (Figure 2.3B and D), showing its functionality.

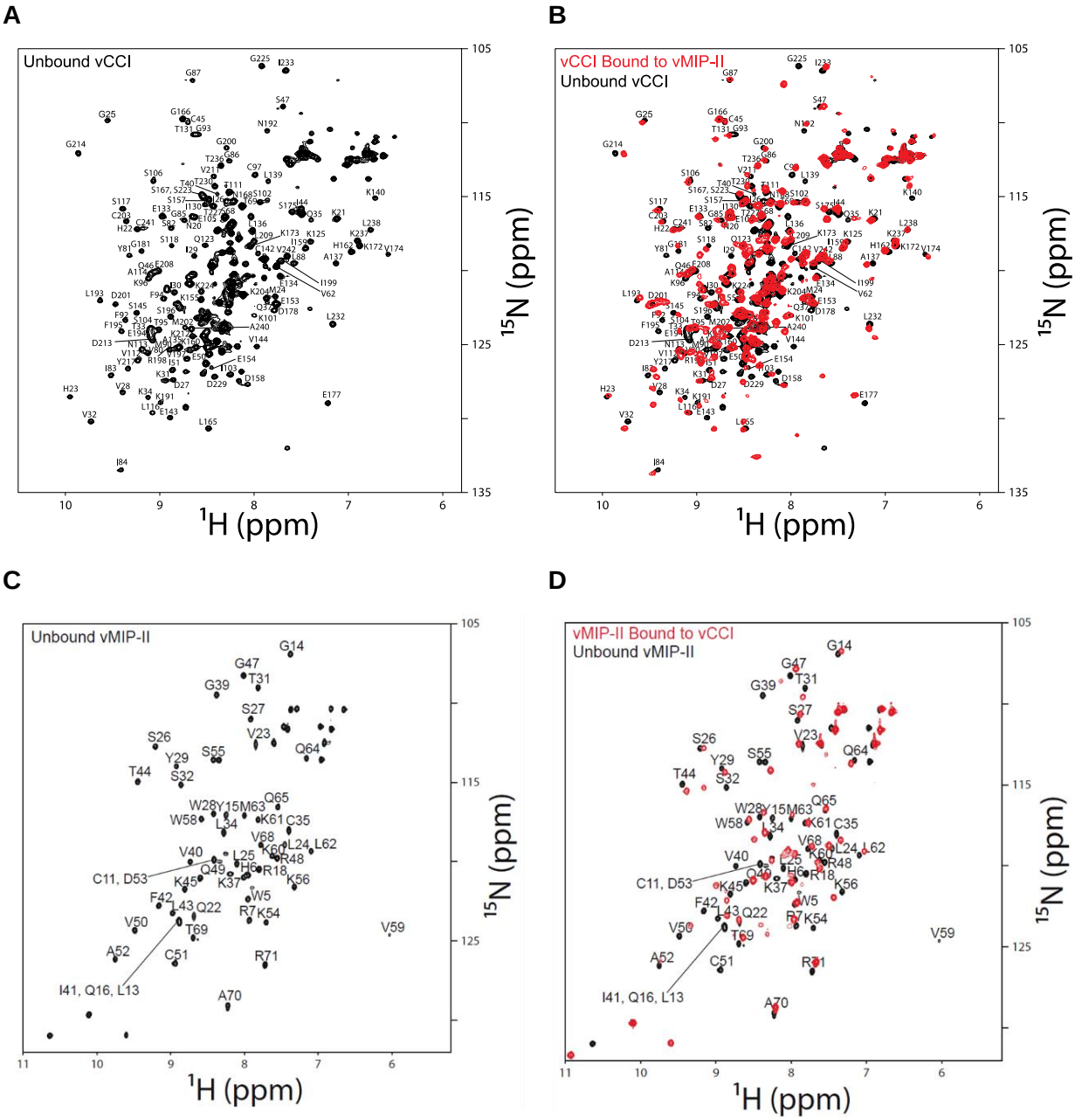

Figure 2.3: Bound versus unbound vCCI and vMIP-II Spectra (**A**) ^{1}H-^{15}N HSQC spectrum of unbound ^{15}N-labeled vCCI in 100 mM NaCl, 20 mM NaOP (sodium phosphate) pH 7.0, measured at 37°C. (**B**) Overlay of the ^{1}H-^{15}N HSQC spectra of free ^{15}N-labeled vCCI (black) and ^{15}N vCCI:^{14}N vMIP-II (red) with a ratio of 1:3, measured under the same conditions as in (A). The concentration of vCCI was 50-60 µM. (**C**) ^{1}H-^{15}N HSQC spectrum of unbound ^{15}N-labeled vMIp-II in 100 mM NaCl, 20 mM NaOP (sodium phosphate) pH 7.0, measured at 25°C. (**D**) Overlay of the ^{1}H-^{15}N HSQC spectra of free ^{15}N-labeled vMIP-II (black) and ^{15}N vMIP-II:^{14}N vCCI (red) with a ratio of 1:3, measured under the same conditions as in (C). The concentration of vMIP-II was 30-40 µM.

Likewise, vMIP-II was refolded and produced (Figure 2.3C) and also showed a tight complex when bound to vCCI (Figure 2.3D), as demonstrated by slow exchange during chemical shift perturbation (see section 2.2.2), further providing evidence that this *E. coli* produced vCCI has proper functionality and that both these proteins bind to one another as expected.

2.3.2 vCCI:vMIP-II produce a high affinity complex

vCCI has a remarkable ability to bind almost all chemokines from the CC subfamily, and a qualitative measure of its binding with herpes-virus encoded vMIP-II has been reported[40]. To more quantitatively investigate the affinity between vCCI and vMIP-II, we used Isothermal Titration Calorimetry, in which vMIP-II was titrated into a solution of vCCI. This technique can provide several thermodynamic parameters and often a dissociation constant. However, analysis of the titration data indicated that the K_d of the complex was below $1x10^{-10}$ M (Table 2.1; see section 2.2.3). This is the lowest detectable limit of the instrument, and although a competitive binding technique could have been used with ITC to find a more accurate K_d[175], an alternate method was used to obtain a more accurate binding constant, as the technique had been used previously for other chemokines[41] and would allow for a direct comparison between the previous and current results.

The alternative method of obtaining affinity involves a competition technique in which vCCI is bound to a fluorescently labeled chemokine (eotaxin-1/CCL11), and the competitor (vMIP-II in this case) is titrated into the solution, with the resulting change in fluorescence anisotropy providing the dissociation constant for the interaction[41]. This showed a K_d of 0.06 nM ± 0.006 nM for the vCCI:vMIP-II interaction (Figure 2.4), which is among the tightest measured vCCI:chemokine interactions[41].

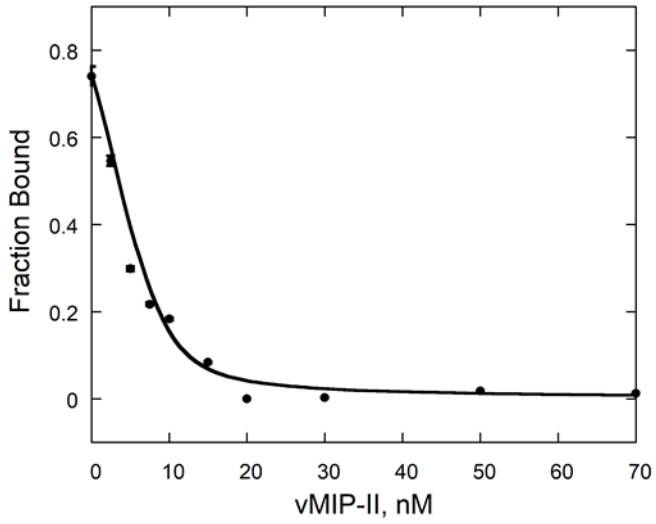

Figure 2.4: Competition fluorescence anisotropy of the vCCI:vMIP-II interaction. vMIP-II was added to a complex of vCCI pre-incubated with fluorescently labeled eotaxin-1 (CCL11) as seen in [41]. Error bars are shown, but are sometimes within the size of the data point. Experiments were kindly performed by Nai-Wei Kuo with data interpretation kindly performed by Cindy DePuer.

2.3.3 Changes in chemical shift suggest vCCI:vMIP-II interaction is similar to other vCCI:chemokine complexes

To determine the amino acids that are likely involved in the vCCI:vMIP-II complex, a comparison of chemical shift changes between the free and bound forms of both vCCI and vMIP-II was carried out (Figures 2.3A and B). In the case of vCCI, ^{15}N-HSQC spectra of the free and bound forms using ^{15}N labeled vCCI (with non-isotopically labeled vMIP-II for the bound) were measured, and compared to determine the level of peak movement upon complex formation. (Assignments of the unbound vCCI were obtained from [136] and BMRB databank 6809.) Chemical shift changes in the ^{15}N vMIP-II HSQC spectrum upon binding non-isotopically labeled vCCI were also determined. (Figures 2.3 C and D). (Assignments of unbound vMIP-II were obtained from [160]). In cases where unambiguous assignment of the peak in the bound spectrum was not possible, conservative assessments of peak movements were made, indicating that actual peak movement could be greater than shown. Figure 2.5 shows the residue-by-residue chemical shift change estimates. See Table 2.2 for definitions of chemical shift perturbation categories for this figure and section 2.2.2 for more information on chemical shift perturbations.

As shown in Figure 2.5A, the areas of greatest chemical shift change for vCCI upon binding to vMIP-II are located in the region of amino acids in the 80's, 140's, and 190's, with changes also observed in the 30's, 170's-180's and 220's. These areas are shown on the structure of vCCI (shown without vMIP-II) in Figure 2.5B, and indicate a binding surface similar to those shown previously with native chemokines[41,46,48,136], comprised of negative charges (D141, E143, and likely the negatively charged loop in the 52-77 region that is not clearly assignable, likely due to flexibility) as well as interaction in the early 80's region. Figure 2.5C indicates that vMIP-II interacts with vCCI using residues from its N-loop region (residues 12-19, see Figure 1.3 for general chemokine structure), and with the second beta strand in the 30's region, as well as with residues in the early 50's. Figure 2.5D shows vMIP-II with presumed interacting regions highlighted.

It is important to note that because the chemical shift perturbations were assigned utilizing a "minimal chemical shift procedure"[166,167], it is very likely that many larger chemical shifts cannot be visualized on the data, as only the closest residues were used for the estimates. As a result, the average chemical shift is relatively low compared to previous assignments[41,189] and may be disproportionately representing the importance of allosteric effects propagating through the beta sheets as a result of the binding. Also, it is important to note that ^{15}N-HSQC chemical shifts rely on the backbone N-H group of a protein; there may still be effects on the functional groups containing no nitrogen that do not result in a direct shift in the backbone itself, and thus, cannot be visualized using this technique.

So, due to the limitations of this technique without re-assigning all the backbone residues in the bound state of the proteins, a molecular dynamics simulation was done in an attempt to elucidate more information about the bound structure of the vMIP-II:vCCI complex.

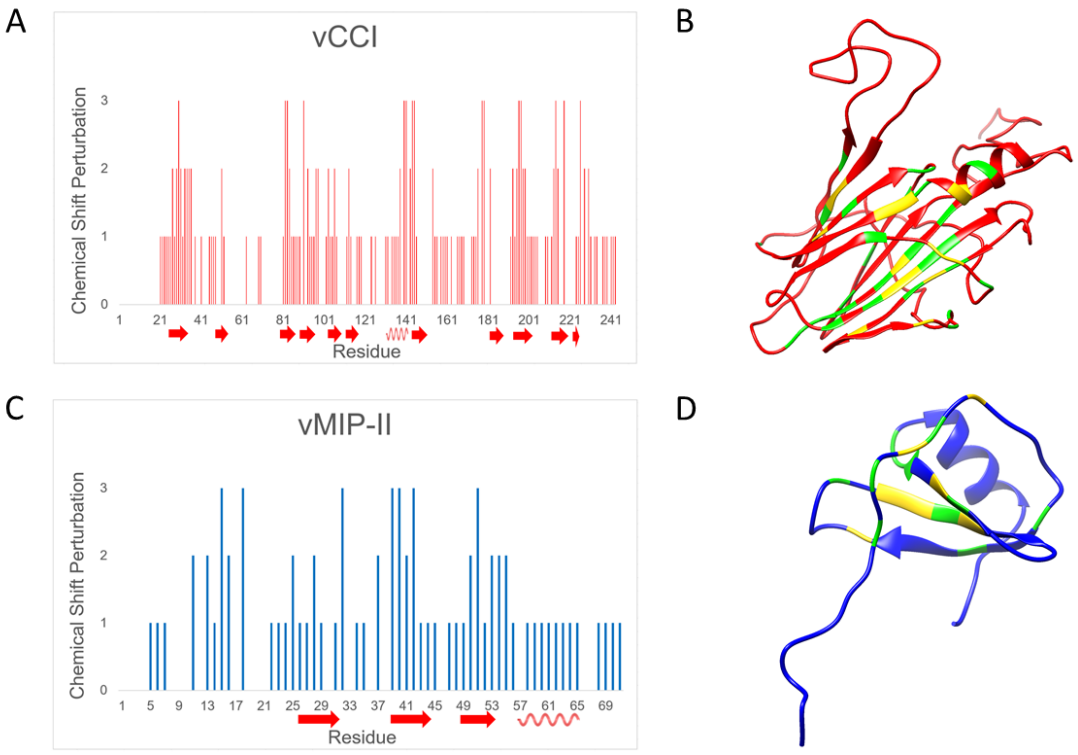

Figure 2.5: Changes in chemical shift upon complex formation. (**A**) Changes in chemical shift in vCCI upon binding to vMIP-II. See Table 2.2 (below) for definitions of "0", "1", "2", "3". (**B**) Structure showing changes in vCCI (Protein Data Bank code 2ffk) upon binding vMIP-II. Green indicates greater than average, up to 1 standard deviation away from the average; yellow indicates over 1 standard deviation away from the average. (**C**) Changes in chemical shift in vMIP-II upon binding to vCCI. See Table 1 for definitions of "0","1","2","3". (**D**) Structure showing those changes in vCCI upon binding vMIP-II (Protein Data Bank code 1vmp). Green indicates greater than average, up to 1 standard deviation away from the average; yellow indicates over 1 standard deviation away from the average. All structure figures were prepared by using UCSF Chimera[20].

Table 2.2: Definitions of the chemical shift perturbation categories in Figure 2.5.

Chemical Shift Perturbation	Definition	vCCI Chemical Shift	vMIP-II Chemical Shift
0	No confirmable change	No peaks visible	No peaks visible
1	Less than or equal to average	$\Delta\delta_{obs} \leq 0.045$	$\Delta\delta_{obs} \leq 0.100$
2	Greater than average, up to one standard deviation above average	$0.045 < \Delta\delta_{obs} \leq 0.086$	$0.100 < \Delta\delta_{obs} \leq 0.178$
3	Greater than one standard deviation above average	$\Delta\delta_{obs} \geq 0.086$	$\Delta\delta_{obs} \geq 0.178$

2.3.4 Molecular Dynamics simulations on vCCI:vMIP-II

To further clarify the structural components and the likely interacting surfaces of the complex and to gain insight into the extraordinarily tight binding between vCCI and vMIP-II, molecular dynamics (MD) simulations were generously carried out by Michael Colvin to create 1 μsec trajectories of both the vCCI:MIP-1β (simulated wild type) complex and the vCCI:vMIP-II complex. Both trajectories are based on the reported vCCI:MIP-1β-^{45}AASA^{48} structure (PDB code 2FFK), but for MIP-1β the residues were changed to reflect the wild type chemokine as opposed to the triple mutant used in that structure determination[46]. The vCCI:vMIP-II trajectory was created from the same complex structure, but by computationally superimposing the vMIP-II structure onto the MIP-1β-^{45}AASA^{48} chain to minimize the average difference between the corresponding Cα-backbone atoms ([46]; Figure 2.6A-B and Figure 2.2; see section 2.2.5). (A third simulation was also included that used the 2FFK structure directly, keeping MIP-1β as a triple mutant (K45A/R46A/K48A) instead of changing it to wild type. However, this third simulation is not emphasized in this work; see the Materials and Methods section 2.2.5 for more details on the simulations.)

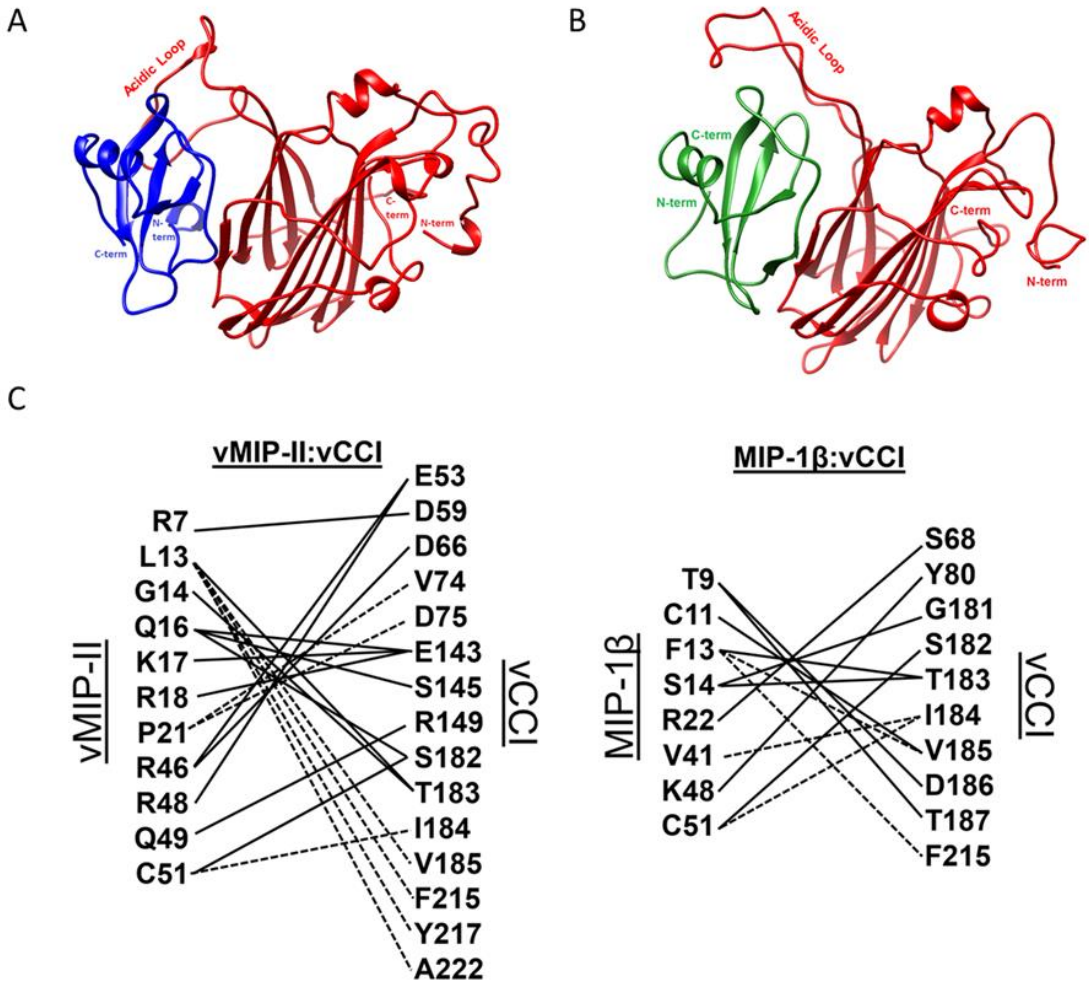

Figure 2.6: Complexes after 1 µsec molecular dynamics simulation of vCCI:chemokine. For all figures, vCCI is in red ribbon and the bound chemokine is either blue ribbon (vMIP-II) or green ribbon (MIP-1β). (**A**) Structure of vCCI:vMIP-II after 1 µsec trajectory. (**B**) Structure of vCCI:MIP-1β in complex after 1 µsec trajectory. (**C**) Interactions between residues of vCCI and vMIP-II, as well as vCCI and MIP-1β. H-bonds (solid lines) are shown if they appear in at least 50% of the last 300 ns of the molecular dynamics simulation. Dashed lines indicate non-H-bond interactions between residues, obtained from PDBePISA. These are defined as residues whose access to solvent is occluded upon complex formation at least 50% of the time, and that are within 2.8 Å of the partner residue on the other protein in at least 50% of the structures sampled every 20 ns for the final 500 ns of the trajectory.

Analysis of the secondary structure of the complexes during the trajectory shows that all of the α-helices and β-sheets are preserved thoughout the 1 µsecond runs for all three complexes (Figure 2.7), suggesting that there are no large-scale shifts that would result in changes in overall secondary structure, which corresponds with the generally-small chemical shift changes seen in Table 2.2.

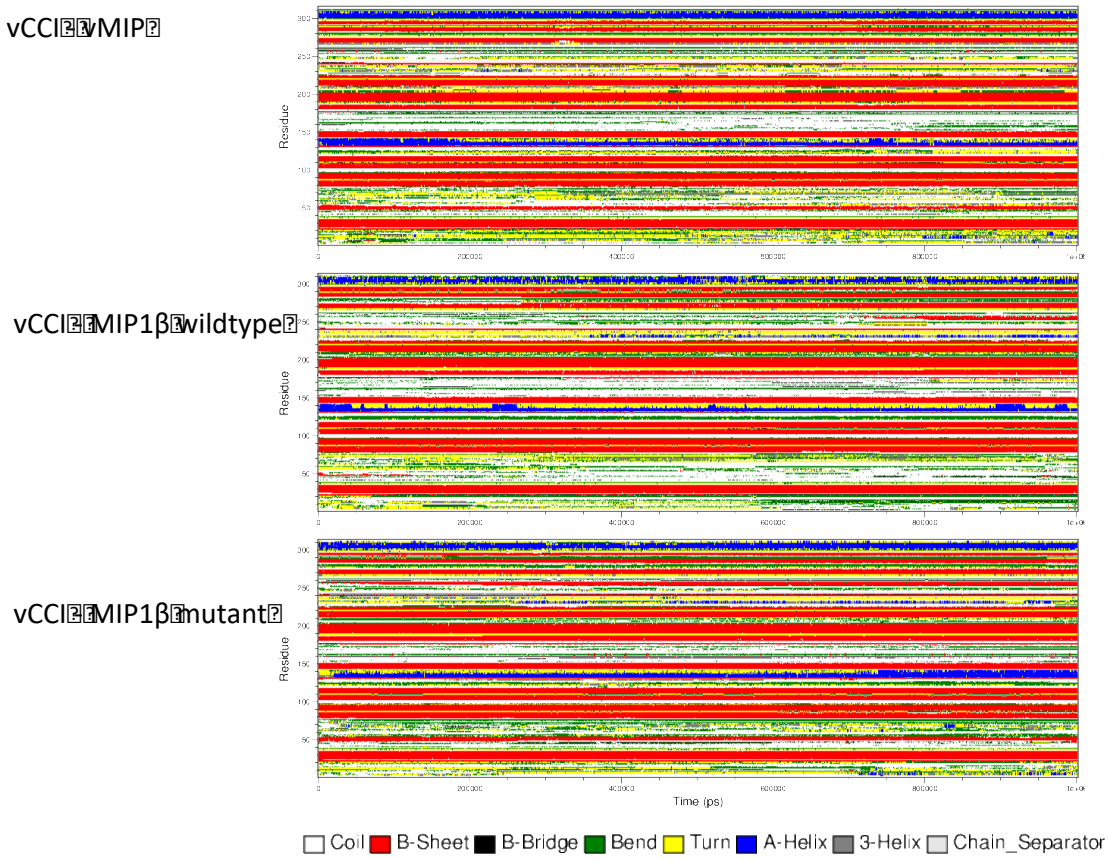

Figure 2.7: Secondary structure of the complexes vCCI:vMIP-II, vCCI:MIP-1β, and vCCI:MIP-1β-K45A/R46A/L48A throughout the MD trajectory[190]. The y-axis shows the residue number in the complex, where residue 1 through 242 is vCCI and 243-313 (vMIP-II) or 243-311 (MIP-1β or MIP-1β-^{45}AASA^{48}, the mutant) is the chemokine, while the colors represent the different secondary structures as shown in the legend. Changes in color over time indicate a change in secondary structure at that point of the trajectory.

The Root Mean Squared Deviation (RMSD) in the position of the backbone Cα atoms for the entire complex is less than 1 nm over the entire run, and the RMSD for the individual vCCI and MIP-1β chains is less than 0.8 nm, indicating no gross protein disordering over the simulation, despite the highly flexible regions in both the vCCI and the chemokines. The residue-level fluctuations are the root-mean square fluctuations (RMSF) of each residue around the average protein structure for a trajectory. The RMSF values plotted vs sequence location and calculated for the final 750 ns of the trajectories are shown in Figure 2.8. These show significant fluctuations of the vCCI N-terminal residues in all three complexes as well as in the vCCI loop at residues 52-77. As described below, the loop acts as an "arm" that folds down on the bound chemokine. The MIP-1β structures show significant fluctuations at both the N- and C-termini, while vMIP-II shows fewer fluctuations at its termini, which will be discussed further in section 2.4.

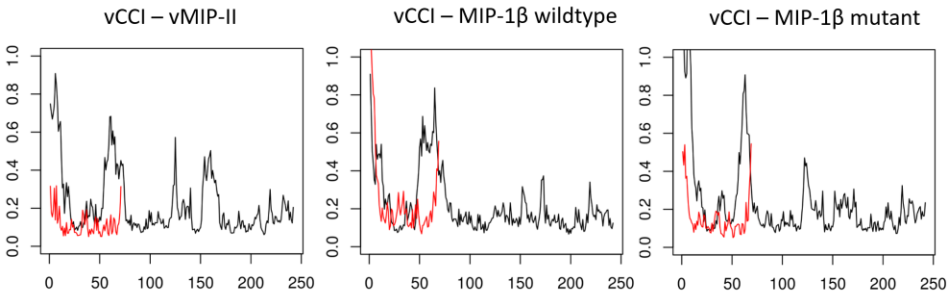

Figure 2.8: Root mean square fluctuation values for vCCI:Chemokine Variant Complexes. RMSF (in units of nm) plotted vs sequence location and calculated for the final 750 ns of the trajectories. The red line represents the chemokine variants (69-71 residues in length), while the black line represents vCCI (241 residues in length). Higher values show regions of increased protein flexibility, including both backbone and side-chain motion.

Figure 2.9 plots the total number of inter-strand (between vCCI and the chemokine variants) hydrogen bonds over the 1 microsecond simulation for each of the chemokine variants. Of the three complexes, the vCCI:vMIP-II complex has significantly more interstrand hydrogen bonds than either of the MIP-1β complexes, as seen by the higher blue peaks in the figure; the wild type MIP-1β has more hydrogen bonds than the MIP-1β-^{45}AASA^{48} mutant, as seen with the green (wild type) compared to black (MIP-1β-^{45}AASA^{48}). Figure 2.6C shows some of the specific the H-bonds formed in the vCCI complex with both vMIP-II and MIP-1β for the at least 50% of the final 300 ns of the trajectories, which are most likely to represent a more "settled" state for the complexes compared to the beginning of the trajectories.

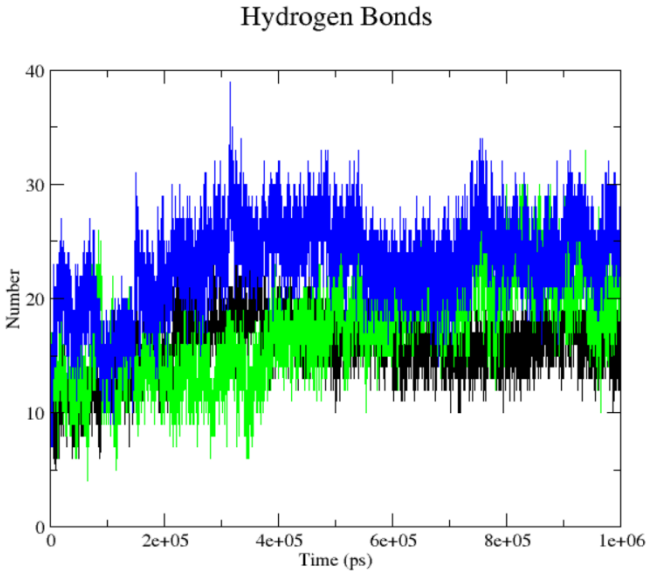

Figure 2.9: Interstrand H-Bonds between vCCI and chemokine variants throughout MD trajectory. Number of interstrand (vCCI:chemokine variant) H-bonds observed throughout the MD trajectory for vCCI:vMIP-II (blue); vCCI:MIP-1β (green) and MIP-1β-^{45}AASA^{48} (black).

We analyzed the effect of individual residues in vCCI and the chemokine variants on solvent exposure using the server-based program, PDBePISA[191–193]. The program computes the solvent accessible surface area for both the complex and the computationally-separated fragments and then reports the solvent-exposed surface area of each residue in the separated proteins as well as the amount of area that is buried when the complex is formed. This analysis was performed every 10 nanoseconds for the final 500 nanoseconds of the simulation for a total of 51 structures analyzed per complex. The three complexes show different total amounts of buried surface area, as shown in Table 2.3, showing that the vCCI:vMIP-II complex is most heavily occluded from water compared to the other two complexes.

Table 2.3: Buried surface area between vCCI and chemokine variants. Buried surface area averaged over 51 structures analyzed during the final 500 nsec of simulation.

Complex	vCCI buried surface area (Å^2)	Chemokine buried surface area (Å^2)
vCCI:vMIP-II	1473	1528
vCCI:MIP-1β wild type	1355	1392
vCCI:MIP-1β K45A/R46A/K48A (variant used in 2FFK structure determination)	1020	1060

Additionally, specific residue-residue contacts vary between the three different complexes; Figure 2.10 shows interactions of residues that are occluded during the simulation upon complex formation at least 50% of the time, and that are within 2.8 Å of the partner residue in at least 50% of the structures sampled every 20 ns for the final 500 ns of the trajectory. The final 500 ns are chosen, since it most likely represents a more “settled” structure with vCCI and the chemokine variants compared to the beginning of the trajectory. While there is considerable overlap between the residues used between the two complexes, several differences in interactions are seen between the complexes, which we then focused on more specifically.

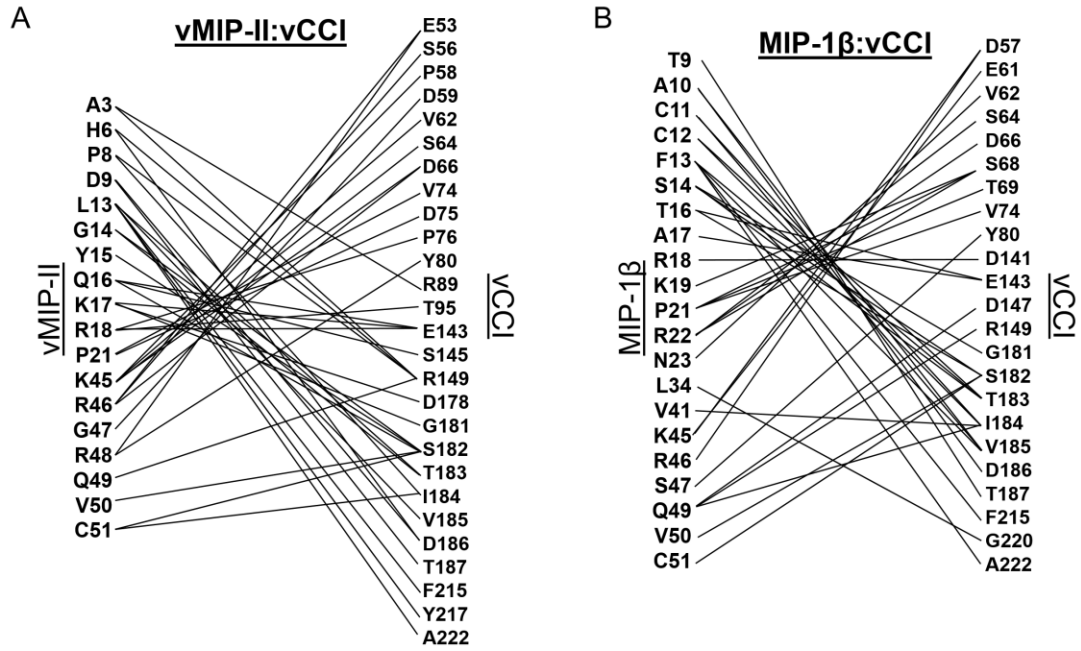

Figure 2.10: Occlusion maps for vCCI with vMIP-II or MIP-1β. (**A**) Interaction maps between residues of vCCI and vMIP-II and (**B**) interaction maps between residues of vCCI and MIP-1β. Interactions are shown for residues that are occluded during the simulation upon complex formation for at least 50% of the final 500 ns, and that are within 2.8 Å of the partner residue in at least 50% of the structures sampled every 20ns for the final 500 ns of the trajectory.

A comparison of the dynamics trajectories of vCCI binding to vMIP-II and to MIP-1β shows some striking differences, in particular revealing several possible interactions that may account for the approximately 10-fold tighter binding for vMIP-II to vCCI than to MIP-1β (whose K_d has been found to be 1.2 ± 0.17 nM[41] using the same Fluorescent Anisotropy technique as seen used for Figure 2.4 with the vCCI:vMIP-II complex). First, at the end of the simulation, the total buried surface area for vMIP-II in complex is 1528 $Å^2$, while the buried surface area for MIP-1β in complex is 1392 $Å^2$ (Table 2.3). Second, as shown in Figure 2.9 and Figure 2.6C, during the time course of the trajectory, vCCI shows an overall larger number of H-bonds with vMIP-II than with MIP-1β. Third, the flexible, negatively charged loop in the 52-77 region of vCCI (between beta strands 2 and 3) makes more contact (including a larger number of H-bonds over the course of the trajectory) with vMIP-II than with MIP-1β (Figure 2.6C and Figure 2.10). And finally, during the 1 μsec trajectory, vMIP-II shows overall more interactions with vCCI than does MIP-1β, in particular at the N-terminus of the chemokine where a large portion of that region of vMIP-II lays across the vCCI binding face, while the MIP-1β N-terminus does not (Figure 2.10).

The vCCI:vMIP-II trajectory shows several individual interactions that illuminate aspects of their binding and complementary interactions, including significant contact throughout the trajectory between residues E143 on vCCI and residue R18 in vMIP-II (Figure 2.11A) and the interaction between the negatively charged loop between strands β2 and β3 in vCCI with K45/R46 in vMIP-II (Figure 2.11C). A similar trajectory is seen for vCCI:MIP-1β, with MIP-1β residue R18 showing interactions with E143 as well as D141 of vCCI (Figure 2.11B). However, vMIP-II's R18 residue also shows interaction

with vCCI residue D75 through almost half the time steps in the trajectory, while no such interaction is observed with MIP-1β, illuminating a difference between these interactions.

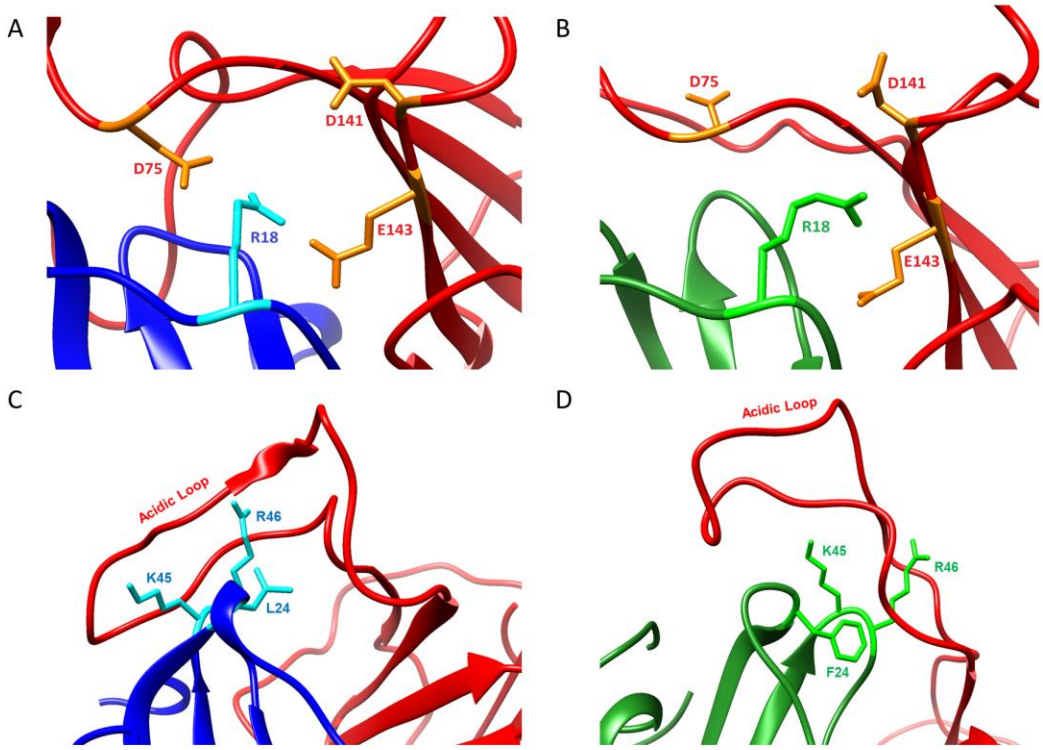

Figure 2.11: Interactions between vCCI and vMIP-II or MIP-1β. vCCI is in red ribbon and the bound chemokine is either blue ribbon (vMIP-II) or green ribbon (MIP-1β). (**A**) A close-up of the interaction between R18 of vMIP-II (cyan) and D75, D141, E143 of vCCI (orange). (**B**) A close-up of the interaction between R18 of MIP-1β (bright green) and D75, D141, E143 of vCCI (orange). (**C**) A close-up of the interaction between the β2 and β3 loop of vCCI (red, maskred 'Acidic Loop') and K45 and R46 of vMIP-II (cyan). (**D**) A close-up of the interaction between the β2 and β3 loop of vCCI (red, marked 'Acidic Loop') and K45 and R46 of MIP-1β (bright green).

The trajectories also show both vMIP-II and MIP-1β have their 24/45/46-position residues clustered together and interacting with the vCCI loop between β2 and β3, but the interaction is much more extensive in the vCCI:vMIP-II complex (Figure 2.11C,D, Figure 2.6C, Figure 2.10, Table 2.3). The N-termini of each of the chemokines also behave very differently in the trajectories. vMIP-II shows considerably more interaction with vCCI throughout its N-terminus for much of its trajectory, while the MIP-1β trajectory shows an N-terminus that does not appear to interact consistently with vCCI, with main contacts to the binding partner not starting until residue 8 (Figures 2.6C and 2.10). In total, the simulation results point to possible reasons why vCCI shows different binding constants to various CC chemokine partners, which will be discussed further in the following section.

2.4 Discussion

The ability to modulate the immune system, and in particular, to reduce the inflammatory response, has great potential in health and medicine. Protein therapeutics have been approved for this purpose[194,195], and investigation has continued into other potential sources of anti-inflammatory proteins. Both herpesviruses and poxviruses have evolved to produce proteins that subvert the mammalian chemokine system, and these include both chemokine binding proteins as well as chemokine homologs[118,196,197]. This chapter investigates the unusually high affinity interaction between the vCCI (also known as 35K) CC chemokine binding protein from rabbitpox, and vMIP-II, a CC chemokine analog from herpesvirus HHV-8, using a combination of biophysical and molecular dynamics techniques, in an attempt to elucidate more about their structures.

vCCI was successfully produced and purified from *E. coli*. This fairly efficient procedure will greatly expand the range of experiments that can be carried out with vCCI, from quickly making large quantities of the protein (and any desired variants) for X-ray crystallography, to inexpensive isotopic labeling that can lead to a variety of NMR experiments, including a full structure determination. Isothermal titration calorimetry indicated a high affinity for the vCCI:vMIP-II complex (Table 2.1), and this was confirmed by fluorescence anisotropy (Figure 2.4), which revealed a K_d of 0.06 nM ± 0.006 nM that is roughly 10-100 times lower than the K_d observed for vCCI with other chemokines using the same method[41]. Other groups have investigated the binding constant of vCCI with various chemokines using other methods, including early qualitative work that suggests that vCCI bounds vMIP-II more tightly than it does most other chemokines[40]. Others have used a scintillation proximity assay[197] and surface plasmon resonance[44,48,49] to show that vCCI binds a variety of chemokines at levels ranging from sub-nanomolar to 20 nM.

Analysis of chemical shift perturbation analysis by NMR indicates that vCCI interacts with vMIP-II using residues similar to those that have been shown to be important in binding by vCCI to other chemokines, MIP-1β/CCL4 and eotaxin-1/CCL11[41,46], such as acidic residues in the 141/143 area (Figure 2.3). Similarly, chemical shift perturbation of vMIP-II upon binding vCCI shows chemical shift changes in generally known areas, including the region near R18 as well as the area near the so-called N-loop of the chemokine, where hydrophobic L13 is located (Figure 2.5C, D). While NMR chemical shift perturbation is a powerful tool, there are two main drawbacks. First, while a perturbation suggests a locus for protein-protein interaction, and one can infer regions of interaction between proteins, it does not confirm a pairwise interaction with the binding partner. Second, 2D ^{15}N HSQC spectra can be ambiguous in terms of assigning peaks upon movement. To resolve ambiguity would require ^{13}C-labeling of the protein(s) and a series of 3D NMR experiments[41]. Therefore, we chose to pursue molecular dynamics simulations which provide a high resolution “movie” (within certain approximations) of the structure and motions of all atoms in the protein and surrounding solvent, and can delineate specific interactions and provide insight into differences in affinity.

Atomistic classical molecular dynamics (MD) is a well-established tool for studying protein structure and dynamics[198]. In typical protein MD, the motions of all atoms are simulated using empirical force fields that approximate the forces due to bonded and non-bonded interactions. The resulting output is a high-resolution series of atomic motions that can be analyzed to characterize the structure and dynamics of the

protein and infer the energy causes of the observed behavior. With modern computers and MD software, it is feasible to routinely run simulations of moderately large proteins (including a solvation shell of water and ions) for microsecond timescales, with the largest published MD simulations reaching millisecond times[199]. The accuracy is limited by the approximate nature of the force field and the limitation that bonds are not broken or formed during the simulations (including protonation and deprotonation of acid and base sites), but MD has been shown able to accurately predict protein properties, such as the folded conformation of small proteins[200].

In the investigation of the vCCI:vMIP-II complex, an MD trajectory up to 1 microsecond was run, providing great insight into likely interactions that were not observable and/or confirmable by our NMR experiments to this point. In general, the hypothesis that these two viral proteins may be a near-ideal binding pair is supported by the MD trajectories, which show that that vCCI:vMIP-II complex has a larger buried surface area (including the vMIP-II's N-terminus) and a greater number of H-bonds throughout the trajectory, including more interactions between the chemokine and the negatively charged flexible loop of vCCI than the vCCI:MIP-1β complex. The MD simulations also provide context for specific regions of interaction that may be useful in general for a vCCI:chemokine complex. For example, D141 and E143 in vCCI were observed to contact R18 in the vCCI:MIP-1β structure[46] and this R18 was found to be critical for vCCI binding in other chemokines[41,44,46,49]. However, mutational studies on vCCI/35K showed E143 to be more important than D141[47]. The MD trajectory provides an explanation, showing significant, continuous interaction between E143 (of vCCI) and R18 (of vMIP-II), while almost no close interaction across the trajectory is observed with D141 (Figure 2.11A). In the vCCI:MIP-1β trajectory, significant H-bonding interaction (although below the 50% threshold for Figure 2.6C) is observed between R18 of the chemokine and both E143 and D141 of vCCI, although the interaction with E143 predominates (Figure 2.11B).

The MD trajectory also provides a possible explanation for other unexplained mutational results. In the original structure of the complex between vCCI and MIP-1β /CCL4, it was observed that both Y80 and R89 in vCCI appeared close in space to the 48th position of MIP-1β[46]. In many chemokines, this position contains a large, basic residue that could be expected to both sterically and electrostatically clash with those groups. It had been noted that mutation of this residue to the smaller Ala48 increased affinity for a similar chemokine MCP-1/CCL2[44,49]. Indeed, MIP-1β was mutated from a Lys to an Ala at that position in the structure, and that mutation was attributed to tighter binding to vCCI. In an attempt to design a vCCI that was better able to bind chemokines, White et al. mutated each Y80 and R89 to Ala in Vaccinia vCCI, hypothesizing that a smaller, uncharged residue in these positions would better interact with the large basic residue of a chemokine[47]. Interestingly, while the vCCI R89A mutation did lead to a better chemokine binding ability, Y80A completely abolished the activity of the protein. The MD trajectory of Y80 in the vCCI:MIP-1β complex shows the tyrosine side chain of vCCI consistently forming an H-bond with the backbone nitrogen of Lys48 of MIP-1β (Figure 2.6C and Figure 2.12B). This Y80 H-bond was not consistently observed in the vCCI:vMIP-II trajectory, although the trajectory shows consistent contact between these residues (Figure 2.10). In either case, the Y80 in this crowded area of the protein shows little motion and appears to be holding open the negatively charged loop in vCCI. The Y80 residue in vCCI has also been mutated by Arnold et al who replaced tyrosine with

arginine; this mutation did also lead to loss of chemokine binding ability, although in this case the cause is likely placing a basic Arg on vCCI near the Arg48 of a chemokine[48].

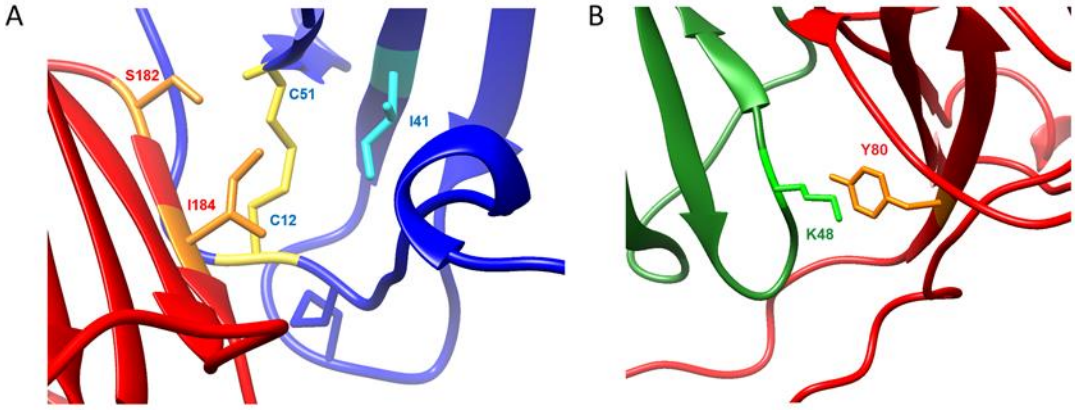

Figure 2.12: Interactions illuminated by molecular dynamics simulation of the vCCI:MIP-1β and vCCI:vMIP-II complex. vCCI is in red ribbon, bound vMIP-II is in blue ribbon, and bound MIP-1β is in green ribbon. (A) Interaction between I184 of vCCI (orange) and I41 of vMIP-II (cyan), as well as S182 of vCCI (orange) with the backbone of C51 of vMIP-II (yellow). (B) Interaction between Y80 of vCCI (orange) and the backbone of K48 of MIP-1β (green).

The MD trajectory has helped reveal several interesting interactions that NMR alone had difficulty explaining. Arnold et al. observed that the mutation of residues S182 and I184 (using the present vCCI numbering) resulted in substantial loss of binding activity, especially for I184 [48]. However, previous chemical shift assignments for vCCI with other chemokines[41,46] as well as the current work with vMIP-II are unable to quantify the shifts to these residues in vCCI. The MD trajectory, however, reveals an interaction between VCCI I184 and vMIP-II Ile41 and Cys51, seemingly to help anchor vCCI to the chemokine throughout the trajectory (Figure 2.12A); this interaction is also seen in the vCCI:MIP-1β trajectory, explaining large chemical shift changes in chemokines in the region 39-42. vCCI Ser182, meanwhile, appears to have formed a hydrogen bond to the backbone N-H of C51 in both vMIP-II and MIP-1β. This may explain the large chemical shift changes to that region of the spectra. This was also observed in the chemical shift changes of eotaxin in the vCCI:eotaxin complex[41].

Overall, a combination of biochemical, biophysical and computational experiments have been used to provide a comprehensive explanation of the basis for the high affinity interaction between vCCI/35K and vMIP-II. These proteins each exemplify a highly evolved mechanism to mimic and/or subvert the CC chemokine branch of the mammalian immune system, and an understanding of their interaction will be useful in both their development as possible therapeutics and in general protein design for immunomodulation. Further studies must be done in order to elucidate the importance of the residues focused on in this chapter, but the vCCI purification technique as well as the MD trajectories offer solid footing for beginning such experiments.

Chapter 3

The effect of N-terminal cyclization on the function of the HIV entry inhibitor 5P12-RANTES

This work was published as part of Nguyen et. al (2017)[201] with slight modifications. Funding was provided by NIH R01AI112011, as well as funding provided by UC Merced SNS Summer fellowships. Thank you to Mike Jian and Meghan Schill for their assistance with DNA and protein purification work.

3.1 Introduction

HIV, the virus that causes AIDS, infects about 2 million people each year, mostly throughout the developing world[202]. While effective treatments are available, particularly in economically advanced countries, there is no cure, and prevention remains a critical issue. A major strategy to prevent HIV infection is the development of microbicides, substances that could be used topically to prevent the sexual spread of HIV. Ideally, a microbicide would be active and remain functional in both the vaginal compartment (pH 4.0-4.5)[203] and the rectal compartment (pH 7)[204] to be capable of being helpful during two of the most common pathways for viral entry.[205]

One of the most promising proteins under consideration for use as a microbicide is 5P12-RANTES. This protein was developed using random mutagenesis at the N-terminus of the chemokine RANTES (also called CCL5) [63], itself a weak HIV inhibitor (Figure 1.6). 5P12-RANTES binds tightly to the chemokine receptor CCR5, which is the major co-receptor used by HIV to gain entry to the cell during initial infection[206,207]. 5P12-RANTES has been shown to be stable at biological temperatures and in the presence of relevant bodily fluids[208], and it was fully protective as a topical microbicide when tested in macaques[76]. More recently, it has been reported as being under clinical development[209].

The N-terminal region of chemokines are known to be critical for function, with slight alterations or truncations known to cause significant differences in their interaction with their respective receptors[83,85,88,210]. For example, the simple addition of an N-terminal methionine caused RANTES/CCL5 to act as a CCR5 antagonist rather than an agonist[85], and truncation of the N-terminal residues of chemokines likewise led to proteins with the ability to bind but not activate CCR5[83,211]. In developing chemokines as HIV inhibitors, chemically modifying the N-terminus of RANTES was found to be particularly effective. For example, one potent analog was AOP-RANTES, in which an organic extension replaced the first residue of RANTES[212]. This modified protein was not only a potent HIV inhibitor, but it also prevented chemotaxis of human monocytes; this is a valuable property because it is important to avoid bringing immune cells to a site

where they could potentially be infected with HIV[213]. Researchers then used this design as a scaffold and tested various synthetic modifications to the N-terminus, resulting in the discovery of the even more potent PSC-RANTES; both AOP-RANTES and PSC-RANTES induce internalization of the CCR5 receptor[87,214]. However, neither of these variants could be produced recombinantly; to circumvent this issue, Hartley et al. developed a random mutagenesis/phage display technique with changes to the N-terminal residues of RANTES that led first to the production of P2-RANTES[88] and then to the discovery of 5P12-RANTES[63]. The mutations that produced 5P12-RANTES were focused on just the first ten amino acids in RANTES (Figure 1.6) and showed over two orders of magnitude variation in inhibitory potency[63].

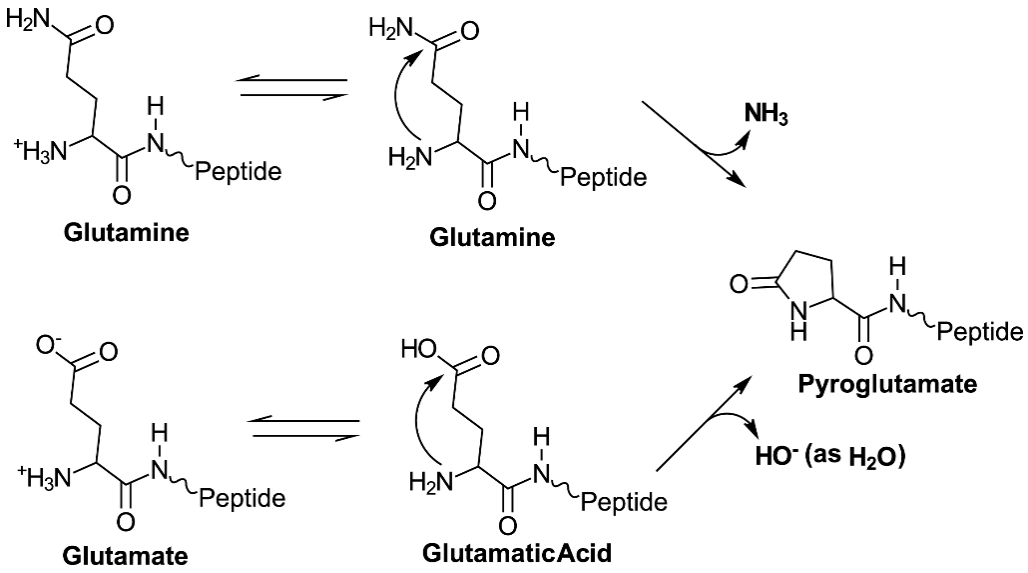

Figure 3.1: **Cyclization reactions of N-terminal glutamine and glutamate residues in a polypeptide chain.** Conditions such as pH are important factors in the rate of cyclization; low pH leads to a higher proportion of a good leaving group, while high pH leads to better nucleophilicity in the attacking amino group.[215,216]

Interestingly, 5P12-RANTES as well as the other resultant inhibitors carried forward from the study, such as 5P14-RANTES and 6P4-RANTES, have a glutamine as the first amino acid (termed Q0 because this is an additional amino acid added to the wild type RANTES sequence, giving 69 total residues rather than the original 68). N-terminal glutamines are able to chemically cyclize [217–219](Figure 3.1), leading to possible heterogeneity in the protein product with two forms (cyclized and uncyclized) that are significantly different in charge and other chemical properties. This has also been a topic of significance in the development of antibodies as therapeutics, where N-terminal Gln and Glu are both common[215,217,220,221]. The increasing use of antibodies as therapeutics has led to several reports about the importance of cyclization of both Gln and Glu at the N-terminus of these and other clinically-relevant proteins[218,220–222]

While it has been recognized that 5P12-RANTES is likely to undergo N-terminal cyclization, the functional impact of this modification at such a critical region of the protein, as well as the details of the kinetics of cyclization, have not yet been reported[63,209,223]. Such information is critical both from the standpoint of understanding

the mechanism of HIV inhibition and from the standpoint of identifying all forms of a molecule that may enter the clinic.

This chapter reports a study on the N-terminal cyclization of 5P12-RANTES, including the effect of cyclization on CCR5 receptor function and HIV inhibition in two representative viral strains, as well as the rate of cyclization under various conditions. We also show the effect of mutating the N-terminal "Q0" to other amino acids, such as the residue glutamate, which cyclizes to form an identical molecule to "wild type" cyclized 5P12-RANTES (Figure 3.1).

3.2 Materials and Methods

3.2.1 Protein Purification

Please see sections 2.2.1 and 5.2.1 for a more detailed explanation on the purification of proteins from *E. coli*. For an overview of the purification of 5P12-RANTES variants, see Appendix B, Figure B-1.

In brief, the following method was used (with the exception of wild type RANTES/CCL5, which required some differences, which are noted): Plasmids were transformed into *Escherichia coli* BL21(DE3) (Novagen) competent cells and expressed in minimal media with $^{15}NH_4Cl$ as the sole nitrogen source. Protein production was induced with the addition of isopropyl β-d-1-thiogalactopyranoside (IPTG) and incubated with shaking at 22°C for 20 h (wild type RANTES/CCL5 was induced with IPTG followed by shaking at 37°C for 6 h). The cell pellet was resuspended in 6 M guanidine hydrochloride, 200 mM NaCl, 10 mM benzamidine, and 50 mM Tris (pH 8.0) and was then disrupted by French press and then centrifuged at 27000 × *g* for 1 h. The soluble portion was loaded onto a nickel chelating column (Qiagen) equilibrated with the resuspension buffer. Proteins were eluted from the column using a pH gradient with 6 M guanidine hydrochloride, 200 mM NaCl, and 60 mM NaOAc (pH 4) followed by addition of 10 mM Beta-mercaptoethanol with stirring for one hour at room temperature. The proteins were then refolded by dropwise addition to 10x volume of refolding buffer (550 mM L-Arginine Hydrochloride, 200 mM NaCl, 1 mM EDTA, 1 mM reduced glutathione (GSH), 0.1 mM oxidized glutathione (GSSG), 50 mM Tris, pH 8), and then allowed to stir overnight at 4°C. The solution was dialyzed twice into 4 liters 200 mM NaCl, 2 mM $CaCl_2$, 20 m Tris, pH 7.4 buffer at 4°C (into 4L of 200 mM NaCl, 20 mM Tris pH 8 for wild type RANTES).

To cleave fusion tags from the purified protein, the samples were incubated for 24 hours with 650 nM of the protease enterokinase (5P12-RANTES-Q0E required 3 days at a 2 μM; 100 nM ULP-1/SUMO protease was used to cleave wild type RANTES/CCL5; protease purification described below.) The protein solution was then centrifuged to remove precipitated material and added onto a second nickel chelating column (Qiagen), with unbound eluent (containing the cleaved 5P12-RANTES or its variants) being collected as cleaved thioredoxin tag with the 6xHis tag remained bound (see Appendix B, Figure B-1). (This second nickel column is not necessary for wild type RANTES/CCL5 purification, as it elutes from a C_4 column at a different percentage of AcN on a reverse phase column from its fusion tag, unlike the other variants). The samples were then dialyzed into 80 mM NaCl and 20 mM Tris, pH8 and then purified on a C_4 reversed-phase chromatography column (Vydac, Hesperia, CA), using an acetonitrile gradient. Wild type RANTES/CCL5 was also purified with the C_4 column, allowing separation from the SUMO tag. Overall, samples were near neutral pH for

about a day during proteolytic cleavage (3 days for 5P12-Q0E due to slow cleavage caused by the negative charge after the enterokinase cleavage site) and a day for dialysis to prepare conditions for the C4 column, during which time N-terminal cyclization was possible (see section 3.2.4 and Figures 3.3 and 3.4A for NMR visualization of cyclized versus uncyclized forms of the protein.) However, dialysis was carried out at 4°C to minimize this reaction, and as noted in Section 3.3.1, little cyclization was observed after this point. The fractions were analyzed on an SDS-PAGE gel to confirm purity and then lyophilized in a Labconco freeze-dry system.

N-terminally cyclized proteins used for functional assays (below) were prepared as follows: purified and lyophilized 5P12-RANTES was incubated for 36 hours at 50°C in 20 mM sodium phosphate buffer pH 2.8 to ensure 100% cyclization (Figure 3.4). 5P12-RANTES-Q0E was incubated for 50 days at 50°C in 20 mM sodium phosphate buffer pH 2.8 to fully cyclize (Figure 3.2C). Both samples were tested by NMR to confirm structural integrity and full cyclization.

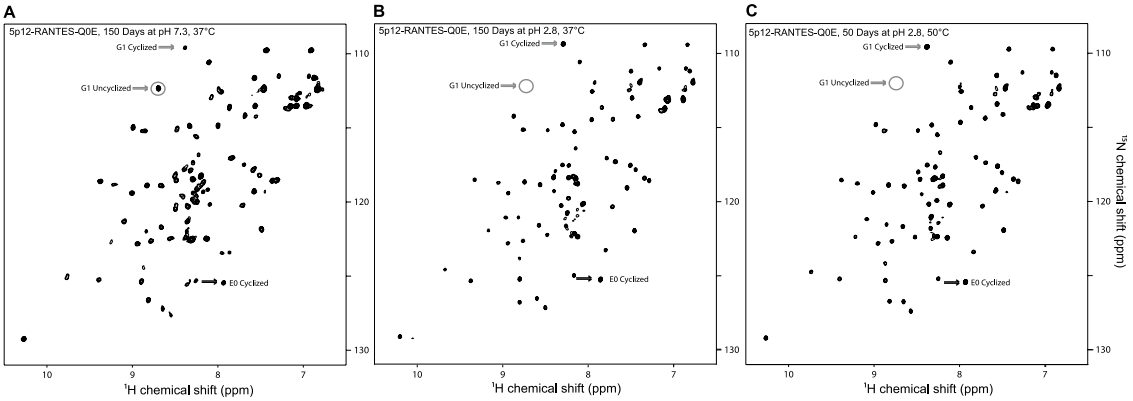

Figure 3.2: **HSQC spectra of ^{15}N-labeled Q0E, after varying incubation conditions**. All spectra were measured at pH 2.8. A) 150 days incubation at pH7.3, 37°C; NMR measured at 25°C B) 150 days incubation at pH 2.8, 37°C; NMR measured at 37°C. C) 50 day incubation at 50°C; NMR measured at 25°C. NMR was carried out in 20 mM sodium phosphate at pH 2.8. Cyclization results in a shift of the G1 residue (grey arrows), as well as an appearance of the N-terminal pyroglutamate residue (lower black arrow), E0. Figure B shows slight deviation in peaks due to temperature at which NMR was measured.

The proteases used in these purifications were produced and purified in our laboratory as briefly described: ULP1 or enterokinase protease were proteins were expressed in LB medium using a pET-28a vector and the cells were collected and French pressed. The ULP1 protease from the supernatant was purified using a nickel chelating column[144]. The enterokinase was found in the inclusion body and resuspended in 6M guanidinium chloride buffer before being purified using a nickel chelating column. Enterokinase was then dialyzed in buffer to allow for refolding and tested for activity through self-cleavage of the fusion tag (manuscript in preparation).

3.2.2 Obtaining rates of 5P12-RANTES and 5P12-RANTES-Q0E Cyclization

See section 3.2.4 for a description of how to quantify cyclization and Figures 3.3 and 3.4A for NMR visualization of cyclized versus uncyclized forms of the protein. See section 2.2.2 for a more detailed explanation about NMR and HSQC techniques.

Samples incubated at pH 2.8 were prepared by adding ^{15}N-labeled lyophilized protein (5P12-RANTES or 5P12-RANTES-Q0E) to 20mM sodium phosphate buffer with 10% D_2O and 5 µM DSS, resulting in a final pH=2.8. Samples were placed into Shigemi NMR tubes (Shigemi, Inc., Allison Park, PA) and incubated at 37°C. NMR spectra were periodically measured at 37°C for these samples (see below), with protein still showing folding even after 150 days of incubation (Figure 3.2B).

Samples at pH=7.3 (the same pH as the pseudoviral assay medium; see section 3.2.7) were prepared by adding ^{15}N-labeled lyophilized protein (5P12-RANTES or 5P12-RANTES-Q0E) to 5 mM Sodium Phosphate buffer with 0.02% sodium azide with a final pH of pH 7.3. Proteins were incubated at 37°C, and samples were removed at selected time points, flash-frozen using liquid nitrogen, and lyophilized. Samples were then dissolved in 20 mM Phosphate buffer, 5% D_2O, and 5 µM DSS with a final pH=2.8, and then measured by NMR at 25°C (see section 3.2.4). Proteins still showed folding, even after 150 days of incubation (Figure 3.2A). All measurements for the pH of protein samples were taken with a micro electrode (Thermo Scientific/Orion) to assure correct pH before running the samples and that no change in pH resulted from the addition of protein to the solution.

3.2.3 NMR Spectroscopy

For a more detailed explanation on NMR spectroscopy and HSQC spectra, see section 2.2.2.

^{15}N-labeled lyophilized proteins were added into 20 mM Sodium Phosphate buffer with 10% D_2O and 5 µM 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS), with a final pH=2.8 for all samples. Concentration of protein ranged from 30 uM to 80 uM. Sample preparation for cyclization time points are further described above in section 3.2.2. All NMR data were acquired on a four-channel 600 MHz Bruker Avance III spectrometer equipped with a GRASP II gradient accessory and a TCI cryoprobe with an actively shielded Z-gradient coil. Spectra were measured at 25 °C or 37 °C. The chemical shift was referenced relative to internal DSS[162]. The data were processed using NmrPipe[163] and analyzed using PIPP [164]. For HSQC spectra, sweep width = 8474.576(^{1}H) and 1766.784 Hz (^{15}N), with 1280 points in ^{1}H and 128* (256 total) points in ^{15}N.

3.2.4 Quantifying N-terminal Cyclization

N-terminal cyclization of 5P12-RANTES or 5P12-RANTES-Q0E results in the appearance of an N-terminal lactam peak “Gln0” (see Figure 3.3 for uncyclized 5P12-RANTES, Figure 3.4A for Cyclized 5P12-RANTES for comparison), as well as a shift in the Gly1 backbone amide peak.[209,224] The appearance of two Gly1 peaks results from there being two populations in the sample, cyclized and uncyclized 5P12-RANTES variant. Peak height comparison of the Gly1 amide when the N-terminus of the protein is cyclized versus uncyclized was used to estimate the percent of the protein cyclized in the sample as follows: NmrPipe was used to obtain peak heights for the Gly1 amide, and then the percent cyclization was obtained by dividing the height of the Gly1 peak when the N-term was cyclized by the sum of the heights of both cyclized and uncyclized Gly1 peaks in a spectrum (to find percent of Gly1 in the cyclized position out of the total Gly1 in solution). Similar values were obtained when using the peak height of the cyclized Q0 peak as it appeared compared to when fully cyclized to measure the cyclized amount, though utilizing a changing value that can be observed at any point in the reaction (Gly1

cyclized versus uncyclized) seemed more likely to be consistent compared to comparing separate NMR files to a final point (Gln0), as the uncyclized N terminus of proteins tend to not be visible on NMR spectra due to exchange with the solution[159]. Using peak volumes rather than height did not appreciably change the results, though both can give relative population differences.

3.2.5 Cell Lines and Viruses

HeLa-TZM-bl cells were obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, National Institutes of Health; the HeLa-TZM-bl cell line (a cell line stably expressing human receptors CD4, CXCR4, and CCR5) was from Dr. John C. Kappes, Dr. Xiaoyun Wu, and Tranzyme Inc[225–229]. 293FT cells were originally obtained from Invitrogen; CHO-K1 cells stably expressing CCR5, $G_{\alpha 16}$, and apo-aequorin were a kind gift from Dr. Marc Parmentier from the Institute of Interdisciplinary Research of the Free University of Brussels (ULB) Medical School, Brussels, Belgium.[230] HeLa-P5L Cells (a cell line stably expressing human receptors CD4 and CCR5) was a kind gift from Dr. M. Alizon (Cochin Institute, Paris, France).

Viral plasmids used to create pseudovirus, including HIV-1 PVO clone 4 (SVPB11), ZM53M.PB12(SVPC11), and pSG3$^{\Delta env}$, were obtained from the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, National Institutes of Health; PVO,[227] clone 4 (SVPB11) was from Dr. David Montefiori and Dr. Feng Gao[231]; ZM53M.PB12, SVPC11 was from Drs. C.A. Derdeyn and E. Hunter[232]; plasmid pSG3$^{\Delta env}$ was obtained from Drs. John C. Kappes and Xiaoyun Wu[232–234].

Utilization of these viral plasmids and cell lines are described in the following sections.

3.2.6 Single Round Pseudovirus Production

293FT cells were co-transfected with the pSG3$^{\Delta env}$ plasmid (containing a defective *vpu* gene and truncated, nonfunctional envelope protein gene (*env*), thus only capable of forming naked (non-enveloped) virus) [232], and an envelope gene plasmid (such as PVO clone 4 or ZM53M.PB12 mentioned in section 3.2.5; this gene determines which strain of the virus will be expressed on the surface of the single-round virion, as this genetic information is missing from pSG3$^{\Delta env}$) using the XTreme Gene Transfection Reagent (Roche/Sigma-Aldrich). 48 hours post-transfection, the supernatant containing single-round virus (virions incapable of infection due to the nonfunctional *env* gene received from pSG3$^{\Delta env}$ plasmid, resulting in naked virions with no envelope proteins capable of ushering new infection) [232–234] was collected, centrifuged, and filtered with a 0.45-µm syringe filter. The viral stocks were then stored at −80 °C for use in pseudoviral assays, as described in section 3.2.7.

3.2.7 Single-Round Pseudoviral Assay

HeLa-TZM-bl cells were grown in culture media (10% FBS, penicillin/ streptomycin, 25 mM HEPES in DMEM). These cells stably express CD4 (the HIV receptor), CCR5 (the HIV co-receptor for initial infection), and CXCR4 (the HIV co-receptor for late-stage infection) and contain genes for luciferase and β-galactosidase under the control of an HIV Long Terminal Repeat (LTR) promotor.[227] This means that these cells, when infected by virus (or single-round pseudovirus) will begin the transcription of these proteins once the promotor interacts with the HIV viral protein Tat,

which is packed in HIV virions with the viral RNA (see section 5.2.3 for more information).

10^4 cells per well were seeded into a 96-well plate and allowed to incubate at 37°C overnight to allow cells to adhere to the wells. The number of cells added is meant to result in 90-100% confluency after five day of incubation (the total length of the assay), so that before assay is complete, cells do not overcrowd and begin apoptosis. Media was then aspirated and replaced with 50µL fresh media and allowed to incubate for 2 hours as inhibitor dilutions were prepared. For preparation of inhibitor, a deep-well dish was made with varying concentrations of inhibitor diluted into PBS and mixed thoroughly to create a decreasing gradient of inhibitor concentration in the wells. 20µL of inhibitor from these dilutions (or 20 µL PBS for positive and negative controls) were added to the TZM cell-containing wells, and the plate was rotated for 1 min to mix, before incubating for 30 min at 37°C. This incubation imitates use of these inhibitors as a topic microbicide, as they would be applied before situations that may result in challenge by HIV. 30µL of single-round virus (described above in section 3.2.6) in media was then added (with virus-free media added to the negative control) for a final volume of 100µL/well and the plate was rotated for 1 min to mix. This "layering" method differs from other similar pseudoviral assays, where the inhibitor is mixed in-well. In our hands, mixing in-well leads to lower IC_{50} (apparent better inhibition), but current experiments use the "layering" addition to more closely mimic conditions under which a topical microbicide may be used.

After 20 hours, media was aspirated and replaced with 100µL/well of fresh media. After 48 more hours (giving cells enough time to incubate and produce the β-galactosidase enzyme in cells that were infected by the pseudovirus), the media was aspirated and the cells were lysed using 30µL/well of 0.5% NP-40 in PBS and allowed to incubate at room temperature for 25 min to ensure total lysis. Substrate was added to the lysed cells (30µL/well; 8mM chlorophenol red-β-d-galactopyranoside (CPRG), which when cleaved by β-galactosidase turns from yellow to red, 10 mM β-mercaptoethanol, 20 mM KCl in PBS) and monitored for color change from yellow to red (See Appendix B, Figure B-2). The absorbance signals at wavelengths 570 and 630 nm were measured, and the 570:630 ratio for each well was calculated. The 570 wavelength quantifies the difference between the yellow (uncleaved CPRG in cells that were not infected by the virus) and red (cleaved CPRG in cells that expressed β-galactosidase, and thus, were infected by pseudovirus) of the substrate in the wells, while the 630 reading gives a background reading correlating to the number of cells in a particular well (as more cells would produce more enzyme, which the 570:630 ratio thus corrects for).

All inhibitors were tested in triplicate, repeated three times. 570:630 ratios from positive control and inhibitor-containing wells had negative control values subtracted from them, and then inhibitor-containing values were divided by the positive control value to give percent infection. Data were plotted using Microsoft Excel, and the IC_{50} value was determined using a linear equation fitted between the two data points surrounding 50% inhibition found in the previous step and then using this equasion to solve for 50% inhibition.

3.2.8 CCR5 Functional Assays in CHO-K1 cells

The functional response of CCR5-expressing cells to RANTES and its variants was analyzed by measuring the luminescence of aequorin in response to release of Ca^{2+}

as described previously[235]. CHO-K1 cells, which stably express CCR5 (the RANTES receptor and HIV-1 co-receptor), $G_{\alpha 16}$ (a g-protein that couples with CCR5 and will allow for downstream signaling upon activation of CCR5, resulting in calcium release; see section 1.1), and apo-aequorin (a protein that, upon binding the cofactor coelenterazine H, is capable of luminescing upon calcium binding[230],) were cultured in Ham's F12 medium (Corning, Cellgro) supplemented with 10% FBS (Life Technologies), 100 units/mL penicillin, 100 µg/mL streptomycin, and 400 µg/mL of the selective agent G418 (Life technologies). The cells were then removed from plates by incubating at 37°C with versene (Gibco), and then adding 3x volume Ca^{2+}- and Mg^{2+}-free DMEM (GE Healthcare) to gently resuspend. As the receptors on the surface of the cell (CCR5) are vital for the assay, using a protease to remove the cells, such as trypsin, would have rendered the cells unusable for the assay, thus requiring the more gentle cell-removing solution, Versene, which removed cells by binding to calcium ions in the solution, breaking cadherins which hold the cells together and causing them to separate from one another[236].

The cells were gently pelleted for 4 min at 600 × *g* and then resuspended at a density of 5×10^6 cells/mL using Ham's F12 Medium supplemented with 10% FBS, penicillin/streptomycin, 25 mM HEPES. Cells were then incubated in the dark at room temperature for 1 h with 5 µM coelenterazine H (Promega) in order to form the non-apo, fully-functional form of aequorin. Cells were then diluted five-fold using Ham's F-12 medium, and 50 µL (50,000 cells) were added per well in a 96-well plate. 50 µL of each inhibitor diluted in Ham's F-12 medium were added to each well at varying concentrations. Luminescence was measured after 30 seconds using an Orion II microplate luminometer (Berthold Techniques, Germany), repeated in triplicate, and data were plotted in Figure 3.8.

3.2.9 CCR5 Activity in HeLa-P5L

In order to confirm the results seen in the CHO-K1 functional assay, a similar assay to those previously used with 5P12-RANTES was utilized. The functional response of CCR5-expressing cells HeLa-P5L to RANTES and its variants was analyzed by measuring the fluorescence using Fluo-4 dye (Thermo Scientific) according to the manufacturer's suggested protocol, to be able to directly compare to previous receptor-activation results[63] performed by others. Briefly, HeLa-P5L cells were cultured in RPMI-1640 (Invitrogen) supplemented with 10% FBS, and 100 units of penicillin and 0.1 mg/ml streptomycin. The expression of CCR5 was selectively expressed by adding zeocin (Invitrogen) at 0.5 mg/ml. Cells were then removed from plates by incubating at 37°C with versene solution (Gibco) to maintain cell surface receptors and then gently resuspended with culture media. The cells were gently pelleted for 4 min at 600 × g and then resuspended at a density of 1×10^6 cells/mL and 100uL was plated onto 96 well plates with black polystyrene wells and a flat, micro-clear bottom (Greiner CELLSTAR). After a 15 hour incubation at 37°C, wells were washed three times with sterile Hank's buffered salt solution (HBSS, 137 mM NaCl, 5.4 mM KCl, 0.25 mM Na_2HPO_4, 0.44 mM KH_2PO_4, 1.3 mM $CaCl_2$, 1.0 mM $MgSO_4$, 1.0 mM $MgCl_2$, 10 mM glucose, 10 mM HEPES, with pH adjusted to 7.4 using 1N NaOH). 50 µL of 2 µM Fluo-4-AM Ester (Invitrogen) in HBSS was then added to the wells. The plate was incubated in the dark at room temperature for 30 min, then washed three times with HBSS supplemented with 2 mM Probenecid (Tocris), then left to incubate with 80 µL/well of HBSS with 2 mM

Probenecid for 30 min at 37°C. Fluorescence was measured after addition of 20 µL of chemokine variants in HBSS for a final chemokine concentration of 300 nM/well. Fluorescence reading was done on a Cytation 5 Cell Imaging Multi-Mode Reader (Biotek) with an absorption/emission at 494/516 nm at 35 seconds, with wild type RANTES/CCL5 as a positive control and HBSS containing no inhibitor as a negative control.

3.3 Results

3.3.1 5P12-RANTES N-terminal cyclization

^{15}N labeled 5P12-RANTES was produced from *E. coli* with an N-terminal fusion partner to disallow N-terminal cyclization of Q0 for most of the purification process. Cleavage of the fusion tag by enterokinase was carried out at pH 7.4 at 4°C followed by reversed-phase chromatography and lyophilization of the pure protein so that it could be stored as dry powder to inhibit cyclization. The purified 5P12-RANTES was solubilized at pH 2.8 for NMR spectroscopy, where it was observed that less than 5% of the protein had undergone N-terminal cyclization during the purification process (Figure 3.3).

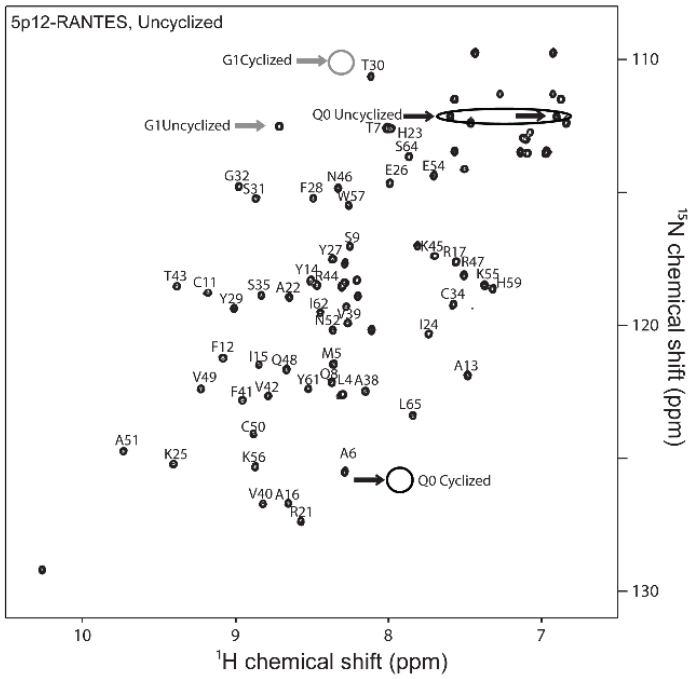

Figure 3.3: Uncyclized 5P12-RANTES. HSQC NMR spectrum of ^{15}N-labeled 5P12-RANTES directly after dissolution in pH 2.8 20 mM sodium phosphate buffer at 25°C. Little or no cyclization is observed at this time. Cyclization of Q0 results in a shift of the G1 peak (labeled, grey arrows; cyclized position circled) which can be used to quantify the amount of cyclized 5P12-RANTES in solution. Cyclization also results in loss of Q0 side chain amide peaks (black arrows, circled) and appearance of cyclized Q0 lactam peak (black arrow, circle). Chemical shift assignments from Wiktor et. al, 2013[224]; no assignments are shown for region near E66, where these authors used a different 5P12-RANTES variant. Also not shown are side chain assignments for W57 and Asn/Gln (except for the relevant N-terminal side chain amide). Percent cyclization was determined by peak height at lower contour level than shown.

The cyclization rate of the N-terminal Gln of 5P12-RANTES at 37°C was monitored by NMR at pH 7.3 and at pH 2.8 as shown in Figure 3.4B. A clear indication of cyclization at the N-terminal "Q0" position is the peak position of the backbone amide of glycine 1. This peak shows a clear shift from 8.7 ppm (^{1}H) and 112.5 ppm (^{15}N) to 8.4 ppm (^{1}H) and 109.5 ppm (^{15}N) as its neighboring side chain cyclizes[209] (Figures 3.3, 3.4A). Concomitantly, the lactam peak of the cyclized pyroglutamate (derived from Q0) is observed to grow in at 7.9 ppm (^{1}H) and 125.5 ppm (^{15}N) upon cyclization. At pH 7.3 these peaks are not discernable in an HSQC spectrum, likely due to faster amide exchange. Therefore NMR analysis for the pH 7.3 incubation was carried out at pH 2.8.

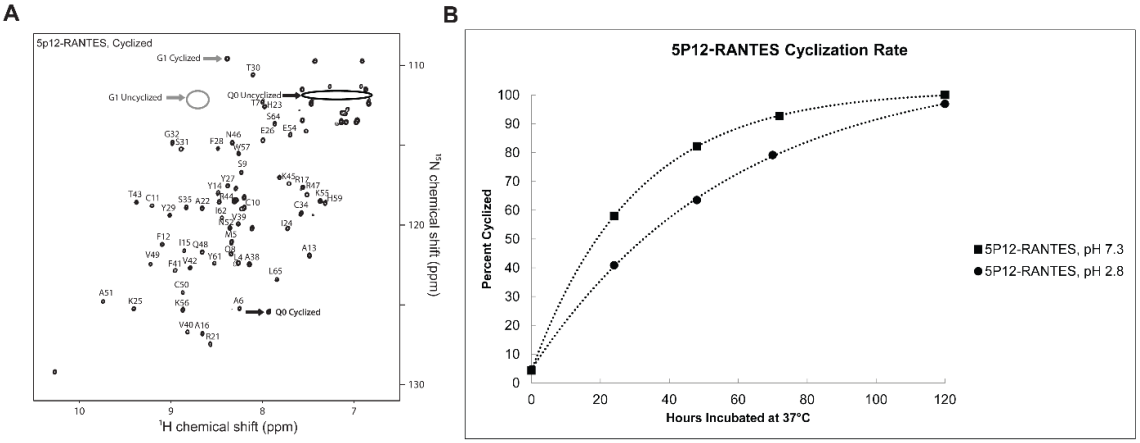

Figure 3.4: 5P12-RANTES cyclization. (**A**) HSQC spectrum of cyclized ^{15}N-labeled 5P12-RANTES after being incubated at 37°C for 5 days at pH 2.8. NMR was performed in 20 mM sodium phosphate at pH 2.8, 25°C. Cyclization results in a shift of the G1 residue (grey arrows; G1 resonances denoted by gray arrows and circles), as well as an appearance of the N-terminal pyroglutamate residue (black arrow; Q0 resonances denoted with black circles and arrows) as well as loss of Q0 amide side chain peaks (black arrow, circled). Assignments are not shown for certain areas as described in Figure 3.3. (**B**) Cyclization over time of 5P12-RANTES at pH=7.3 and pH=2.8, incubated at 37°C. Amount of cyclization was determined by obtaining peak heights of the amide of G1 when the N-terminus of the protein (Gln0) was cyclized and uncyclized using NMRPipe, and dividing the cyclized peak height by the total of all G1 (cyclized and uncyclized) peak heights.

In order to monitor cyclization at pH 7.3, lyophilized powder of 5P12-RANTES was dissolved at pH 7.3, incubated at 37°C, and aliquots were removed at various time intervals. These aliquots were lyophilized to halt further cyclization. Then immediately prior to measurement by NMR, each dry aliquot was dissolved in pH=2.8 buffer. In this way, incubation occurred at neutral pH, but spectroscopy occurred at low pH where key peaks were visible in the spectrum, chemokine solubility is optimal, and on a time scale that would not allow significant further cyclization.

As shown in Figure 3.4B, Q0 cyclization of 5P12-RANTES at pH 7.3 (37 °C) exhibits a half-life of about 20 hours. When cyclization is measured with constant incubation at pH 2.8 (37°C), the half-life is about 33 h, with cyclization essentially complete at 120 hours.

3.3.2 5P12-Q0E also cyclizes to produce mature 5P12-RANTES

To further investigate the properties of the N-terminus of 5P12-RANTES, we produced 5P12-RANTES-Q0E, in which the Gln is replaced by the amino acid Glutamate, which is also capable of cyclizing when placed at the amino terminus of a protein (Figure 3.1). While the cyclization is expected to be slower due to the poor leaving group on the Glu side chain, the product of cyclization should be identical to cyclized Q0. This variant allows further study of cyclization and also provides an alternate route to the presumed mature product i.e. fully cyclized 5P12-RANTES containing a pyroglutamate at the N-terminus.

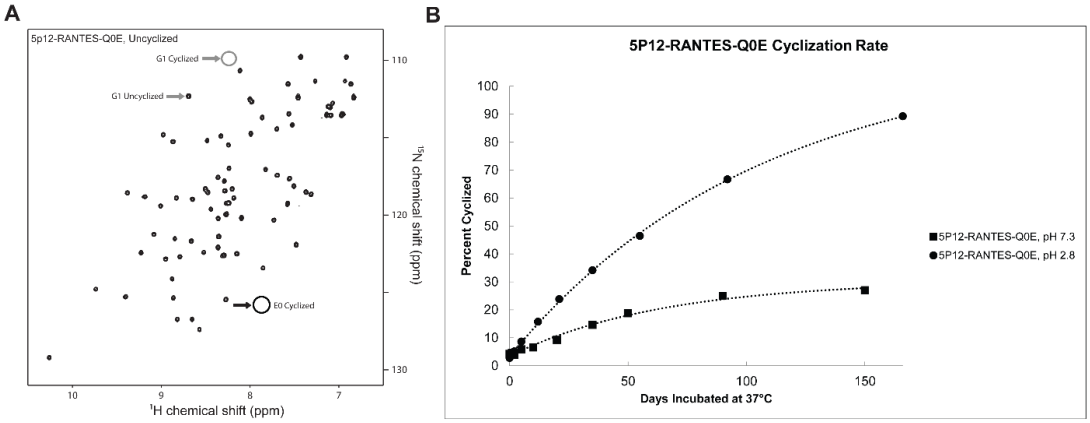

Figure 3.5: 5P12-RANTES-Q0E. (A) HSQC spectrum of ^{15}N-labeled 5P12-RANTES-Q0E. The spectrum is identical to that of 5P12-RANTES except for a slight G1 shift (grey arrows) and a loss of two side chain amide peaks corresponding to the Gln-0 NH_2 group and no cyclized peak (black arrow, circled, lower right). Cyclization results in a shift of the G1 residue (labeled) which can be used to quantify the amount of cyclized 5P12-RANTES-Q0E in solution. Spectrum was measured in 20 mM sodium phosphate, pH 2.8, 25°C. Percent cyclization was measured by peak height at lower contour levels than shown. (B) Cyclization over time of 5P12-RANTES-Q0E in pH=7.3 and pH=2.8, incubated at 37°C. Spectrum at the 150 day time point shows folded protein, with some degradation (Figure 3.2A, B.)

Isotopically labeled ^{15}N 5P12-RANTES-Q0E was produced with an N-terminal fusion protein to disallow cyclization until cleavage near the end of purification, in a similar manner as with 5P12-RANTES. Initial solubilization of purified, lyophilized protein followed by NMR revealed that less than 4% of this protein was cyclized during purification (Figure 3.5A). Incubation of ^{15}N 5P12-RANTES-Q0E was carried out at pH 7.3 and pH 2.8 and monitored by NMR as described above. As shown in Figure 4B, the reaction to produce N-terminal pyroglutamate takes many days, with a half-life of greater than 150 days at pH 7.3 and a half-life of about 60 days at pH 2.8. This faster cyclization at low pH is expected due to improvement of the leaving group upon protonation. When cyclization is complete, the spectrum is identical to cyclized “wild type” 5P12-RANTES as expected (Figure 3.6)

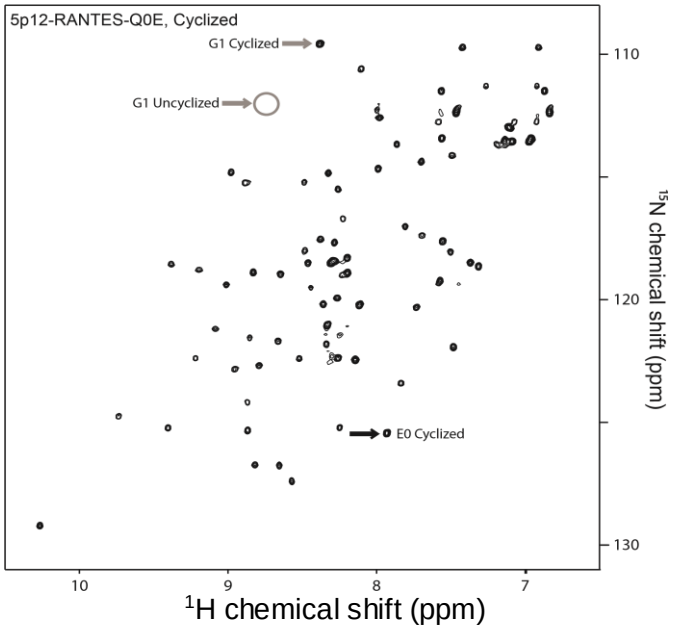

Figure 3.6: **Cyclized 5P12-RANTES-Q0E.** HSQC spectrum of cyclized ^{15}N-labeled 5P12-RANTES-Q0E after being incubated at 50°C for 50 Days at pH 2.8; used for functional assays. NMR was performed in 20 mM sodium phosphate at pH 2.8, 25°C. N-terminal cyclization of E0 results in a shift of the G1 residue (grey arrows), as well as an appearance of the N-terminal pyroglutamate residue (lower black arrow), E0. This is in the same position as the peak corresponding to cyclized Q0 in 5P12-RANTES (Figure. 3.4A).

3.3.3 High potency HIV inhibition does not require N-terminal cyclization or N-terminal glutamine

To determine the relative importance of N-terminal cyclization on the anti-HIV potency of 5P12-RANTES, single round viral assays were carried out using two representative strains of HIV. Since these assays necessarily use mammalian cells, conditions are constrained to physiological pH (pH 7.3). In these assays, cells are aliquoted into 96 well plates and allowed to divide overnight. Various dilutions of inhibitor (freshly made from lyophilized powder and minimally cyclized) were incubated with the cells for 30 minutes. Virus was then added and the cells incubated overnight. Medium containing virus and inhibitor were removed the following day as medium was changed and cells were allowed to grow an additional two days with fresh medium. Infection was monitored with a standard β-galactosidase readout as described in Section 3.2.7. The total time the inhibitor is in solution and able to cyclize is 20 hours at pH 7.3 in this assay.

As shown in Table 3.1, “immature” 5P12-RANTES that was essentially uncyclized at the start of the assay gave an IC_{50} of 1.51 ± 0.10 nM for strain PVO.4 and 1.61 ± 0.12 nM for strain ZM53. Given the length of the assay, it is estimated that roughly 50% of the immature 5P12-RANTES did cyclize during the course of the assay. However, virus entry has been demonstrated to take place over the course of minutes[237], so a significant amount of virus-cell interaction could be expected before the inhibitor was significantly cyclized. “Matured” 5P12-RANTES (that was incubated for 36 hours,

pH 2.8 at 50°C to allow full cyclization prior to use in the HIV assay) showed essentially equal potency, with an IC_{50} of 1.27 ± 0.10 nM for PVO.4 and 1.46 ± 0.15 nM for ZM53.

Table 3.1: **IC_{50} Values of 5P12 Variants.** IC_{50} values of 5P12-RANTES variants (in nM) against a Clade B HIV viral strain (PVO.4) or Clade C viral strain (ZM53) in single-round pseudoviral infection assays. Assays were performed in triplicate and repeated three times.

Table 3.1. IC_{50} Values of 5P12 Variants		
5P12 Variant	*PVO.4 (nM)*	*ZM53 (nM)*
5P12-RANTES Uncyclized	1.51 ± 0.10	1.61 ± 0.12
5P12-RANTES Cyclized	1.27 ± 0.10	1.46 ± 0.15
5P12-RANTES-Q0E Uncyclized	6.30 ± 0.14	4.04 ± 0.29
5P12-RANTES-Q0E Cyclized	1.09 ± 0.06	1.23 ± 0.09
5P12-RANTES-Q0N	2.20 ± 0.33	1.31 ± 0.19
5P12-RANTES-Q0I	1.21 ± 0.24	1.43 ± 0.11
5P12-RANTES-Q0F	1.26 ± 0.19	1.84 ± 0.12
5P12-RANTES-Q0L	1.31 ± 0.12	1.45 ± 0.15

The mutant 5P12-Q0E showed somewhat poorer inhibition in its immature, uncyclized form, with an IC_{50} of 6.30 ± 0.14 nM for PVO.4. This variant has a negatively charged glutamate at position 0, a part of the protein that is presumed to interact with the receptor CCR5 in or near the hydrophobic membrane of the cell[238]. Given the half-life of cyclization described above, it is expected that during the assay, under 5% of this protein would be cyclized. When the matured, cyclized 5P12-Q0E (prepared as described in Section 3.2.1) was used in these inhibition assays, its potency was essentially identical to wild type, cyclized 5P12-RANTES (Table 3.1), as expected for these two now-identical proteins.

As a control for uncyclized Q0 (“wild type”) 5P12-RANTES, we produced and purified 5P12-Q0N, in which the Gln is replaced by Asn. This amino acid is identical to Gln but has a shorter side chain so cannot similarly cyclize. As expected, the NMR spectrum of ^{15}N 5P12-Q0N shows no indication of a cyclized N-terminus (Figure 3.7A). In HIV inhibition assays, this protein shows high potency, with an IC_{50} of 2.20 ± 0.33 nM against PVO.4 and 1.31 ± 0.19 against ZM53, essentially the same as for the wild type (Q0) protein, indicating that the presence of a cyclized pyroglutamate is not necessary for the anti-HIV activity of 5P12-RANTES.

Since both uncyclized and cyclized 5P12 variants were shown to be potent inhibitors of HIV, we made other mutations at the Q0 position of the protein to determine the effect of changes at this position on HIV inhibitory ability. For these experiments the Gln0 was changed to Phe, Ile, and Leu. Each of these variants was expressed and

shown to be folded by NMR (Figures 3.7B, C, D); none of these side chains are expected to undergo significant chemical transformation upon incubation. As shown in Table 3.1, each variant was highly potent and very similar in activity to 5P12-RANTES, with IC_{50} values ranging from 1.21 ± 0.24 nM to 1.31 ± 0.13 nM (PVO.4) and 1.43 ± 0.11 nM to 1.84 ± 0.12 nM (ZM53).

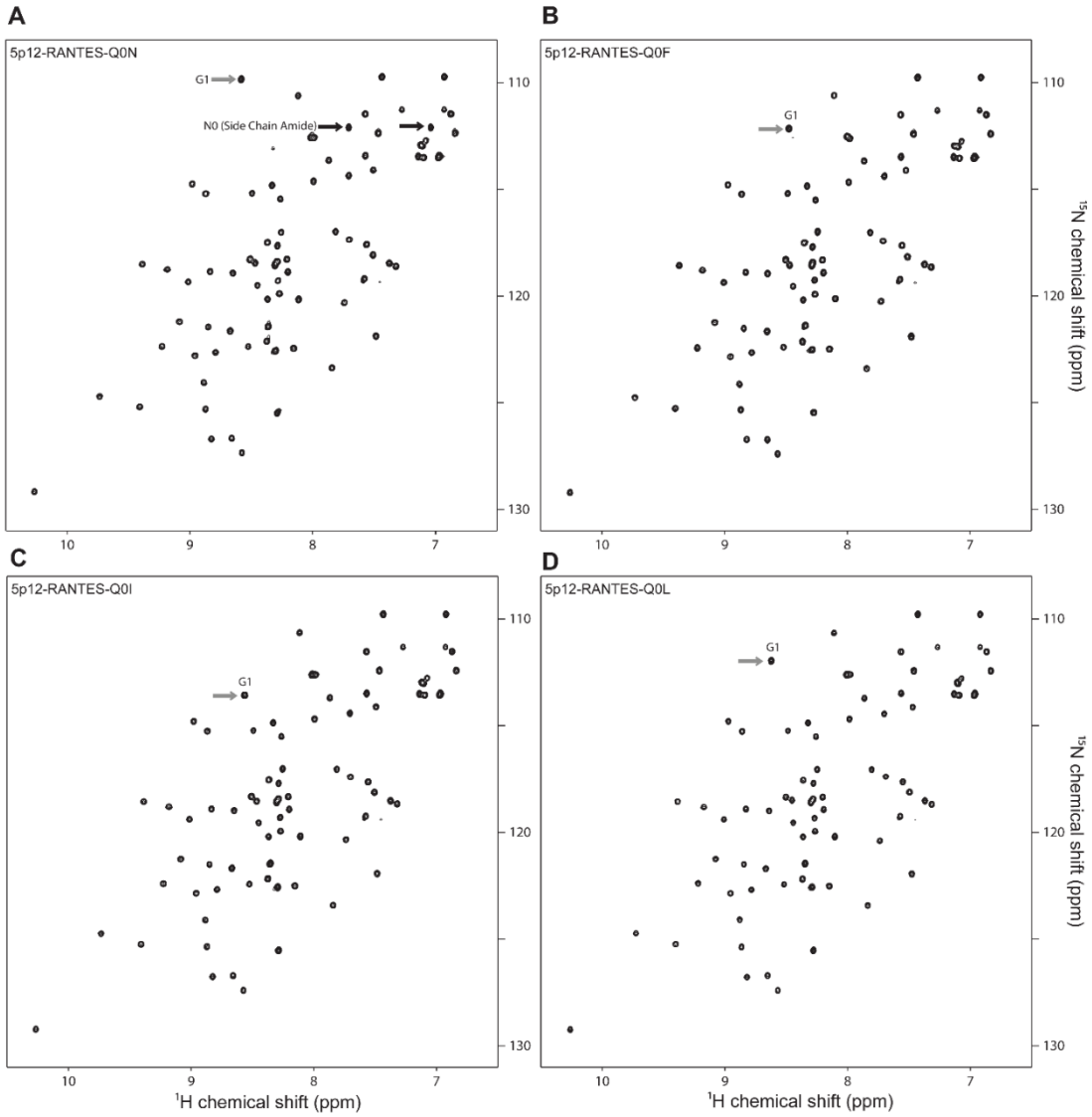

Figure 3.7: Other 5P12-RANTES N-terminal Mutant Controls. HSQC spectra of ^{15}N-labeled **A**) 5P12-RANTES-Q0N, **B**) 5P12-RANTES-Q0F, **C**) 5P12-RANTES-Q0I, and **D**) 5P12-RANTES-Q0L, (pH 2.8, 20 mM sodium phosphate buffer, 25°C). The spectra are essentially identical to the 5P12-RANTES spectrum (Figure 2) except for a slight G1 shift (grey arrows) and a difference in the amide side chain peaks. B, C, and D do not have that amide side chain. A shows a slight shift in the amide side chain peaks that correspond to the Asn-0 NH2 group (black arrows). While N-terminal Glu and Gln can cyclize, Asn does not favorably cyclize due to the shorter side chain.

3.3.4 The effect of the N-terminal amino acid and its state of cyclization on activation of CCR5

5P12-RANTES has been reported to be an antagonist of CCR5, not inducing calcium mobilization; lack of calcium mobilization is a property that is favorable (and likely necessary) in the context of therapeutic HIV inhibition. To determine whether the cyclization of Q0 or the placement of other amino acids at the N-terminus affect activation of CCR5, calcium flux assays were carried out on CHO cells expressing CCR5, $G_{\alpha 16}$, and apo-aequorin, a calcium-responsive protein. As shown in Figure 3.8, neither N-terminally cyclized nor N-terminally uncyclized 5P12-RANTES (i.e. having Q0) caused a calcium flux until concentrations reached 500 nM, while wild type RANTES/CCL5 showed calcium release at low nM concentrations as expected. Further, the N-terminal variants of 5P12-RANTES (-Q0E cyclized and uncyclized, -Q0N, -Q0I, -Q0F, -Q0L) also exhibited no ability to cause the release of calcium until reaching 500 nM concentration (Figure 3.8). These results were further confirmed with a fluorescence assay using human HeLa-P5L cells, which showed similar results (Table 2.2).

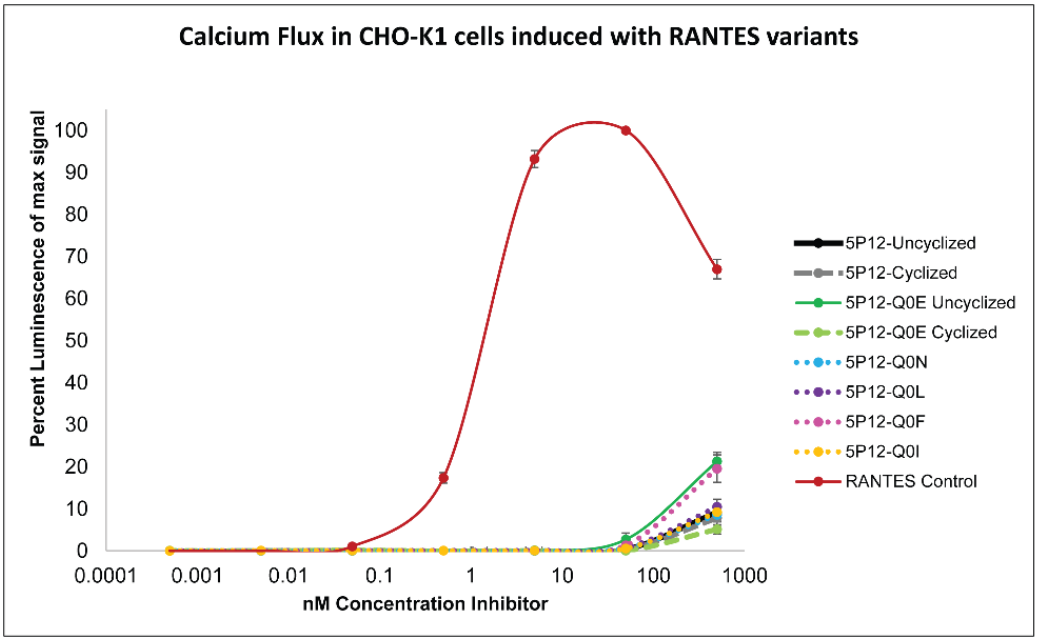

Figure 3.8: Calcium Flux Assay in CHO-K1 cells induced with RANTED variants. CHO-K1 cells expressing CCR5 on their surface were incubated with various concentration of chemokine (either wild type RANTES/CCL5 or a 5P12-RANTES variant) and monitored for luminescence of aequorin upon calcium release. At very high "supraoptimal" concentrations, RANTES/CCL5 exhibits aggregation with alternate effects on receptor activation.[239]

Table 3.2: CCR5 signaling activity of 5P12-RANTES variants based on fluorescence assay using Fluo-4 AM and HeLa-P5L cells. Percent activity was determined by comparing to the maximal signal of positive control, wild type RANTES. Assays were performed in triplicate and repeated three times, at 300 nM concentration.

Table 3.2. CCR5 Signaling Activity (Fluorescence)	
5P12 Variant	*CCR5 signaling activity (% Max)*
RANTES, Wild Type	100%
5P12-RANTES Uncyclized	-0.4 ± 0.3 %
5P12-RANTES Cyclized	0.3 ± 0.2%
5P12-RANTES-Q0E Uncyclized	6.5 ± 0.6%
5P12-RANTES-Q0E Cyclized	-0.8 ± 0.2%
5P12-RANTES-Q0N	2.3 ± 0.5%
5P12-RANTES-Q0I	4.1±0.8%
5P12-RANTES-Q0F	5.0 ± 0.1%
5P12-RANTES-Q0L	-0.4 ± 1.5%

3.4 Discussion

There are at least four highly potent classes of protein HIV entry inhibitors, each with great potential for clinical use in preventing viral infection. The first of these is broadly neutralizing antibodies[240], which bind gp120 or gp41 and show great promise, particularly when used in combination[241], but are generally produced in eukaryotic cells.[242,243] A second class of entry inhibitors includes lectins, which bind to glycosylation sites on HIV env and include the highly studied Griffithsin and Cyanovirin-N, both of which are being evaluated for potential clinical trials[244,245]. A third group are variations of peptides that are derived from HIV gp41, called fusion inhibitors[246], that are particularly effective in combination with other inhibitors[93,247].

Finally, in the fourth group, certain chemokine variants have been shown to inhibit entry of HIV due to their ability to bind the chemokine receptors that act as co-receptors for HIV. The most effective chemokine variants have been derived from the chemokine RANTES, originally by using synthetic modification at the N-terminus to produce PSC-RANTES, for example, which was shown to be effective in protecting macaques from infection[76,87]. Structural studies have shown that the N-terminus of the chemokine is likely to interact with the receptor at or near the cell membrane[248], and work on synthetic RANTES variants supported these results, showing that hydrophobic groups at the N-terminus were more effective at interacting with the CCR5 receptor than similar variants with hydrophilic modifications[87].

More recently, random mutagenesis/phage display was employed to select for RANTES N-terminal variants with enhanced ability to inhibit HIV. This work first led to P2-RANTES[88], which was used as a starting point for later selection of a series of even

more highly potent, fully-recombinant RANTES variants[63]. This latter set of inhibitors included 5P12-RANTES, 5P14-RANTES, and 6P4-RANTES, each of which has slight differences in its 10 N-terminal amino acids, leading to differing ability to activate and/or internalize the CCR5 receptor, and again, providing evidence that the N-terminal region of the chemokine is critical in interacting with the receptor and affecting its conformational changes in the membrane. The most clinically promising of these inhibitors is generally considered to be 5P12-RANTES, due to its combination of high potency and inability to induce downstream signaling or internalization of the CCR5 receptor. Therefore, this inhibitor can inhibit HIV without activating the immune system or mediating an influx of immune cells (and by virtue of being a CCR5 antagonist, may actually reduce immune activity).

5P12-RANTES has been shown to have many properties necessary for use as a topical microbicide, including stability at elevated temperatures, a range of pH conditions, and within environments containing bodily fluids[208,249]. This inhibitor demonstrated the ability to prevent infection in macaques[76] and has shown effectiveness against many strains of HIV with little or no indication that HIV is able to mutate to lose sensitivity to it[89]. Therefore, this protein is a top candidate for a clinical HIV microbicide.

A critical issue for any protein moving to the clinic is a clear understanding of its function, as well as characterization of any variation in its chemical or structural composition. For instance, antibodies often have Glu or Gln as N-terminal residues, and the growing importance of antibodies as therapeutics has led to the study of these residues' propensity to cyclize [215,217,220,221]. 5P12-RANTES has an N-terminal Gln (referred to as Q0), an amino acid that is known to cyclize when at this position in a protein [217,250], leading to an extra cyclization step during clinical-grade production[223]. However, a detailed study of the rate and functional effect of this cyclization in 5P12-RANTES has not been reported.

Wiktor et al have provided NMR chemical shift assignments and the structure of a non-aggregating variant of 5P12-RANTES [209,224]. This group estimated a half-life for Q0 cyclization as approximately two days at pH 3.8, although the sample was refrigerated during part of this time, since the primary goal of the work was not to study N-terminal cyclization but rather to investigate the overall structure of the protein and its interactions with detergents.

We show here that the N-terminus of 5P12-RANTES cyclizes to form pyroglutamate with a half-life of roughly 20 hours at pH 7.3 and 33 hours at pH 2.8 (37 °C). This indicates that the protein, if purified without regard to cyclization, would still cyclize on the time scale of a typical HIV assay (20 hours at pH 7.3), leaving inconclusive whether high potency is reserved for only the cyclized form. However, when Q0 is replaced with N, such cyclization is not observed, and this variant (5P12-Q0N) shows essentially identical HIV-inhibitory ability as 5P12-RANTES, indicating that a cyclized N-terminus is not important for the function of this inhibitor. Other hydrophobic N-terminal variants, Q0I, Q0F, Q0L, also show very high potency, nearly equivalent to 5P12-RANTES (Table 3.1).

To further investigate the importance of the cyclized form of 5P12-RANTES, we made 5P12-Q0E, in which glutamate replaces glutamine at the N-terminus of the protein. Glutamate can also be expected to cyclize and form the same pyroglutamate as is formed by glutamine (Figure 3.1), although a poorer leaving group is expected to

cause this reaction to be slower. Upon cyclization, this variant will yield a product that is identical to cyclized 5P12-RANTES (Figures 3.3, 3.6).

It was observed that at 37 °C, 5P12-Q0E cyclization occurs with a half-life of well over 150 days at pH 7.3 (with still only 25% cyclization observed at 150 days) and roughly 60 days at pH 2.8. The fully cyclized version of the variant (prepared by prolonged incubation at higher temperatures) was tested against two strains of HIV and shown to be essentially identical in function to cyclized 5P12-RANTES containing the original Q0. Interestingly, the uncyclized 5P12-Q0E also showed good inhibitory properties despite the timescale of the assay not allowing significant cyclization and therefore leaving a negatively charged Glu at the terminus of the protein throughout the assay (Table 3.1).

These results clarify the conditions necessary to cyclize 5P12-RANTES if it were to be used clinically in a homogeneous form. They further show that not only is cyclization not necessary for anti-HIV activity, but also that a variety of amino acids at the N-terminus would be expected to provide protection against HIV infection. This allows a greater range of amino acid sequences as potential inhibitors, which could lead to flexibility in determining which sequence most easily results in production of the large amounts of protein required for clinical use.

In conclusion, 5P12-RANTES is a highly potent HIV inhibitor that could be used as a microbicide to prevent HIV infection. However, chemical cyclization of its N-terminal glutamine causes this protein to exist as a heterogeneous mixture when expressed recombinantly. We show that the protein is still a highly effective inhibitor of HIV in both its cyclized and uncyclized forms and have determined the rate of cyclization under various conditions, with relevance both for the manufacture and clinical use of this protein. We also show that several amino acids are suitable replacements for the N-terminal glutamine if changes in the sequence are necessary or desirable.

Chapter 4

Preliminary Results for Next Generation Chimeric 5P12-Linker-C Peptide HIV Entry Inhibitors

Funding for this work was provided by NIH R01AI112011, as well as funding provided by UC Merced SNS summer fellowships. Thank you to Kimberly Nguyen and Ricardo Ramos for their assistance with production of pseudoviral env DNA.

4.1 Introduction

Human Immunodeficiency Virus (HIV) is the virus that causes Acquired Immune Deficiency Syndrome (AIDS), a condition that causes an individual's immune system to fail, allowing for opportunistic infections to prove fatal[53]. Though considerable progress has been made in the field, especially for those currently living with the virus, there are still just under 2 million infections every year[56]. As there has still been no true success with a vaccine[251], one of the leading strategies for hopes of combating this virus around the world is through preventing new infections entirely. Current preventative therapies mainly focus on inhibitors that take effect only after the therapeutic is able to enter the cell[58], including the inhibitor used in the newly-popular HIV pre-exposure prophylaxis (PrEP)[59]. However, there is a different class of HIV inhibitors known as entry or fusion inhibitors that can inhibit the virus at an even earlier step of initial infection, viral entry into the human cell. By blocking entry into the cell, these inhibitors prevent the viral content from ever having the opportunity to even interact with the inner machinery of the human cells[60].

As a majority of the countries with the highest rate of new HIV infection are found in the developing world[56], for the inhibitors to be easily available to the population, they must be safe, stable, and cost-effective.[61] Many HIV entry inhibitors are based on antibodies or small molecules, which tend to be expensive to synthesize (antibodies) or can be rendered impotent by vital mutation (small molecules and certain antibodies), and thus, would be difficult to make readily available.[62] So instead, some groups have focused on peptide or small protein-based HIV entry inhibitors, which could be synthesized in a more cost-effective manner compared to a protein as complex as an antibody[63].

HIV entry occurs in several steps (Figure 1.5), allowing for several points at which the process can be impeded. An HIV virion has a viral spike made up of two heavily glycosylated proteins, gp120 and gp41.[66] First, the virus uses gp120 to bind to its main receptor, the human protein CD4 (Cluster of Differentiation 4)[67], a receptor usually found on the surface of cells such as certain types of T-cells, macrophages, and dendritic cells (cells that are typically involved in immune response and accumulate at the site of infections)[68]. The gp120:CD4 interaction induces a slight conformational

change that allows for gp120 to then bind to its co-receptor, the human chemokine receptor CCR5[69] or CXCR4[70]. Once both the receptor and a co-receptor are bound, gp120 goes through a larger conformational change, revealing gp41, which previously was buried and protected inside the trimer core[71]. For the final steps, gp41 goes through a conformational change and extends the fusion peptide found at its N-terminus towards the host cell to pierce the human cell membrane. The C-terminal helices and the N-terminal helices of gp41 then come together into a six helix bundle that pulls the two membranes together, allowing the virus to physically enter the cell[72].

Because the co-receptors for HIV are chemokine receptors, much research has been done on the effects of chemokines on HIV entry. There have been no reported cases of new infections through virus subtypes that use the CXCR4 receptor[70], which is connected to later-stage HIV infection (progression into AIDS), so those focusing on preventing initial HIV infection generally have focused on the CCR5 co-receptor. Soon after the discovery of these co-receptors, scientists noticed that the native chemokines for CCR5 (MIP-1β/CCL4, MIP-1α/CCL3, and RANTES/CCL5) could slightly inhibit HIV infection,[81] and it was soon discovered that the initial ten or so residues on these chemokines are important to their function. Removal of these N-terminal residues from RANTES[82] or MIP-1β[83] resulted in a molecule that could bind to CCR5 and block HIV infection (weakly), but did not activate the signal transduction pathway that led to inflammatory response. This was intriguing, since unnecessary inflammatory response can bring $CD4^+$ cells to the site of HIV infection, the cell types susceptible to HIV infection[18]. Because of this ability to alter signaling and binding properties, eyes turned to these chemokines, especially RANTES, as a likely scaffold for an HIV microbicide that could block HIV from CCR5 interaction. Although it may seem worrisome to block an chemokine receptor, there are healthy individuals who are born with a mutation that prevents them from expressing functional CCR5, a mutation known as CCR5Δ32.[84] This suggests that there should be no negative effects to blocking this receptor, and clinical trials of small-molecule CCR5-blockers have shown this to be true[84].

One of the leading HIV entry inhibitors is an N-terminal variant of RANTES known as 5P12-RANTES; this chemokine analog not only shows sub-nanomolar HIV inhibition with no signs CCR5 activation[63], it has also been shown that HIV has difficulty evolving around this inhibitor[89] while also displaying full inhibition in vaginally-challenged macaque models[76]. But due to HIV's propensity to mutate and lose sensitivity to drugs[90], it is difficult to bring such an inhibitor through clinical trials due to fear of the virus mutating into highly-resistant strains of HIV. Because 5P12-RANTES only binds to CCR5, viruses that use CXCR4 as its co-receptor will not be blocked by this inhibitor.[89] So, there is a worry that inhibitors that only block CCR5 may exert an evolutionary pressure to push the virus towards using CXCR4 instead, which has been seen in some individuals using a small molecule CCR5 inhibitor, maraviroc[91]. HIV usage of CXCR4 has been linked to many debilitating consequences of late-stage AIDS[92], and thus, a push towards mutations that would preferentially utilize this receptor is something that HIV inhibitor candidates would need to attempt to avoid.

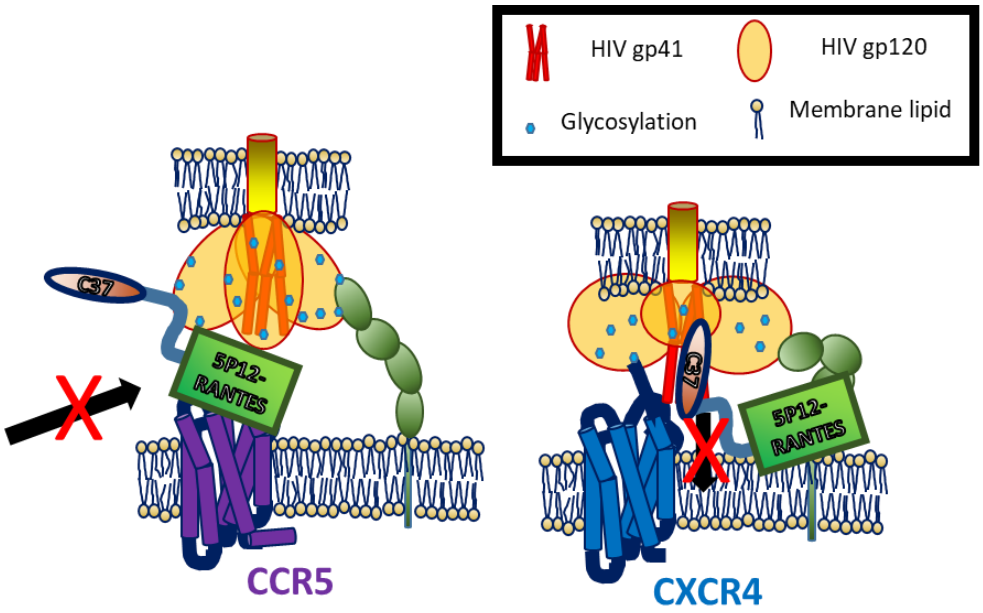

Figure 4.1: 5P12-Linker-C37 inhibiting CCR5- or CXCR4-utilizing strains of HIV. With HIV strains using CCR5 for entry (Purple 7TM receptor on left), 5P12-RANTES is able to block co-receptor interaction with gp120, blocking HIV entry. However, in strains utilizing CXCR4 (blue 7TM receptor on right) 5P12-RANTES cannot bind to the receptor. However, C37 is capable of binding to the N-terminal half of the fusion peptide gp41, blocking membrane fusion, instead.

In an attempt to get around this disadvantage, the P. LiWang group created a chimeric version of 5P12-RANTES by utilizing a flexible amino acid linker to covalently join it to the HIV C-peptide analogue C37, creating a protein called 5P12-Linker-C37 (Figure 4.1).[93] C-peptide analogs received their name because they mimic the sequence of the C-terminal alpha helix found in HIV gp41.[94] By mimicking this region of gp41 (The C-Heptad Repeat region, CHR, Figure 4.2), these C-peptide analogs are able to form a coiled coil with the N-terminal alpha helices of gp41 once they are exposed by gp120, preventing them from forming the six-helix bundle necessary for membrane fusion[95] (Figure 1.5B). Unlike gp120, gp41 is less prone to mutations[252], allowing C-peptides to affect a broader range of viral strains, despite their window of possible activity being limited to the HIV entry step right before six-helix bundle formation. As a result, one of these C-peptide analogs, T20 (also known as enfuvirtide, Figure 4.2), has become an FDA approved drug[96] to help aid those who have developed resistance to other HIV inhibitors.[78] Because of this, the chimeric protein 5P12-Linker-C37 is able to inhibit HIV at two points of the entry process – both the co-receptor binding to gp120 and fusion membrane fusion (Figure 4.1). Because these C-peptides interact with gp41, a highly-conserved part of the HIV virus, they are very broadly-acting with both CCR5 and CXCR4-utilizing clades of virus and can compensate for some of the disadvantages seen with 5P12-RANTES alone, with 5P12-Linker-C37 not only showing more potent inhibition than unlinked 5P12-RANTES and C37, but also showing the ability to inhibit HIV infection even against HIV strains that do not use CCR5 as a co-receptor[93].

```
T2635   ---TTWEAWDRAIAEYAARIEALIRAAQEQQEKNEAALREL---------------
T1144   ---TTWEAWDRAIAEYAARIEALLRALQEQQEKNEAALREL---------------
C37     --HTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQEL------------------
T20     ---------------YTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWF-----
CHR     WNHTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFNITNW
```

Figure 4.2: Sequence comparison of C-peptide analogs with C Heptad Repeat of HIV. Next generation C-peptide analogs are in red and orange; as shown, unlike C37 (Green) or T20 (blue), the next-generation inhibitors contain several mutations to allow for better helical structure and difficulty for gp41 to mutate around them, unlike C37 and T20, which match the natural HIV C-heptad repeat (CHR) seen in black.

However, it has been shown that there are several mutations that gp41 can evolve that the C-peptide C37 (the C-peptide used in 5P12-Linker-C37) are weak against[97], including several mutations that render T20 entirely ineffective, as well. C37 has also been shown to have poor pharmacological properties, making it less-than-ideal for use in patients,[98] which is one of the reasons why it has not been FDA approved, unlike T20. However, a new generation of C-peptides have been created and optimized to be more resistant to the mutations that T20 is weak against (several of which C37 has also shown weakness to)[99]. Two next-generation C-peptides, T-1144 and T-2635 (Figure 4.2), both show good pharmacokinetic properties as well as an increased level of difficulty for HIV to escape their inhibition through mutation.[100,101] These traits suggest that they are viable candidates for creating a next-generation set of chimeric inhibitors that will be more broadly-acting than the previous C37 variants, while still maintaining the same low-cost benefits that could allow them to be available to countries in most-need of such inhibitors.

4.2 Methods and Results

4.2.1 Expression and Purification of 5P12-Linker-T1144 and 5P12-Linker-T2635

Please see sections 2.2.1 and 5.2.1 for a more detailed explanation on the design and purification of proteins from *E. coli*.

The first step of the process was expressing 5P12-Linker-T1144 and 5P12-Linker-T2635 and purifying the protein, to see if they result in folded protein, as well as creating 5P12-Linker-C37 as a control. Please see section 5.2.1 for a more detailed explanation on the purification of proteins from *E. coli*. The genes for 5P12-Linker-T1144 and 5P12-Linker-T2635 (ordered from GenScript) were placed into a pET-28a vector with a 6x-His tag and SUMO fusion tag, then transformed into *Escherichia coli* BL21(DE3) (Novagen) competent cells. 5P12-Linker-C37 was in an identical construct and purified the same as 5P12-Linker-T1144 and 5P12-Linker-T2635. The cells were then grown in minimal media with $^{15}NH_4Cl$ as the sole nitrogen source until OD_{600} reading reached 0.7-0.75. Protein production was induced by adding IPTG to a 1 mM concentration and incubated with shaking at 22°C for 20 h.

The cell pellets were resuspended in 6 M guanidine hydrochloride, 200 mM NaCl, 50 mM Tris (pH 8.0) and were lysed by French press, then centrifuged at 27000 × *g* for 1 h. The soluble portion was then loaded onto a nickel chelating column

(Qiagen) equilibrated with the resuspension buffer. The column was then washed with 50 mL resuspension buffer, and then washed with 50 mL 6 M guanidine hydrochloride, 200 mM NaCl, 80 mM NaOP (pH 7.2). Proteins were eluted from the column using a pH gradient with 6 M guanidine hydrochloride, 200 mM NaCl, 60 mM NaOAc (pH 4), with the pH 4 fractions pooled together. 10 mM βME was then added and the solution was allowed stir for 2 hours at room temperature to allow for disulfide reduction of the 5P12-RANTES portion of the chimeric proteins. The proteins were then refolded by dropwise addition into 10x volume of refolding buffer (550 mM L-Arginine Hydrochloride, 200 mM NaCl, 1 mM EDTA, 1 mM reduced glutathione (GSH), 0.1 mM oxidized glutathione (GSSG), 50 mM Tris, pH 8), and then allowed to stir overnight at 4°C to continue refolding. The solution was dialyzed three times into 4L of 200 mM NaCl, 20 mM Tris pH 8 buffer at 4°C, ideal conditions for Ulp-1 cleavage of the SUMO tag.

To cleave the SUMO fusion tag from the purified protein, the samples were incubated for 12 hours with 100 nM of the protease Ulp-1 (The protease used in these purifications was produced and purified in our laboratory as briefly described: ULP1 was expressed in LB medium using a pET-28a vector. The cells were then collected and French pressed, and ULP1 protease from the supernatant was purified using a nickel chelating column[144]). An SDS-PAGE gel was run to confirm protein cleavage. The samples were then dialyzed into 80 mM NaCl, 20 mM Tris pH8 and then purified on a C_4 reversed-phase chromatography column (Vydac, Hesperia, CA), using an acetonitrile/water gradient with 0.1% trifluoroacetic acid (TFA) in both buffers. The fractions were analyzed on an SDS-PAGE gel to confirm correct protein size and purity, and then were lyophilized on a Labconco freeze-dry system.

5P12-RANTES was produced as described in Section 3.2.1, while N-Acetylated and C- Amidated version of T-1144 and T-2635 peptides were purchased from GenScript.

4.2.2 NMR Spectroscopy of 5P12-Linker-T1144 and 5P12-Linker-T2635

For a more detailed explanation on NMR spectroscopy and HSQC spectra, see section 2.2.2.

NMR samples of 52 µM 5P12-RANTES, 68 µM 5P12-Linker-C37, 57 µM 5P12-Linker-T1144 and 58 µM 5P12-Linker-T2635 were run at 25°C in 20 mM sodium phosphate buffer with 10% D_2O and 5 µM 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) with a final pH of 2.8 to check proteins for proper folding. All HSQC NMR data were acquired on a four-channel 600 MHz Bruker Avance III spectrometer equipped with a GRASP II gradient accessory and a TCI cryoprobe with an actively shielded Z-gradient coil. The chemical shift was referenced relative to internal DSS[162]. The data were processed using NmrPipe[163] and analyzed using PIPP[164]. For HSQC spectra, sweep width was 8474.576(^{1}H) and 1766.784 Hz (^{15}N), with 1280 points in ^{1}H and 128* (256 total) points in ^{15}N.

After several refolding attempts, 5P12-Linker-T1144 and 5P12-Linker-T2635 showed folded spectra (Figure 4.3) that overlap almost perfectly with 5P12-RANTES spectra (76/80 peaks overlapping in 5P12-Linker-T1144 and 77/80 peaks with 5P12-Linker-T2635, including tryptophan and amide side chains, with the largest shift seen with L65 (compare to Figure 3.3), as is expected with a residue near the C-terminus, as

it is closest to the linker region), suggesting that 5P12- RANTES is fully folded within the spectrum. The spectra also show at least 42 more backbone peaks not corresponding to 5P12-RANTES, as well as some slightly unfolded regions in the center, which corresponds with a flexible 10-amino acid linker and a 39-residue peptide known for having some flexibility[100,101], especially with the proper size for the full-length protein being seen on an SDS-PAGE gel.

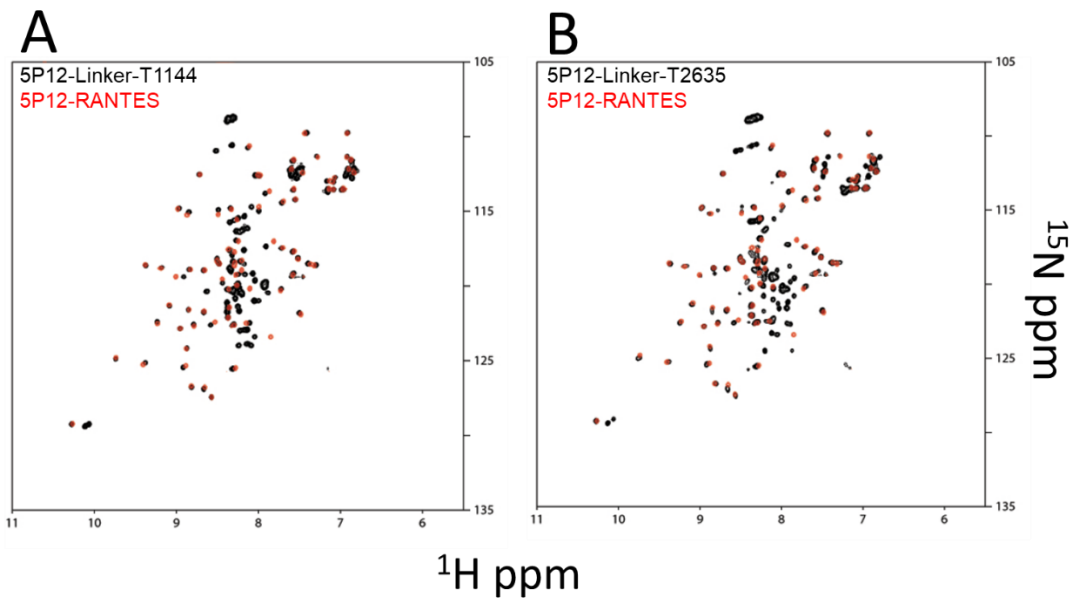

Figure 4.3: HSQC NMR spectra of 5P12-Linker-T1144 and 5P12-Linker-T2635. (**A**) ^{15}N-Labeled HSQC spectrum of 57 µM 5P12-Linker-T1144 in black with 52 µM 5P12-RANTES spectrum overlaid in red. (**B**) ^{15}N-Labeled HSQC spectrum of 58 µM 5P12-Linker-T2635 in black with 52 µM 5P12-RANTES spectrum overlaid in red. All spectra were run at 25°C in 20 mM NaOP Buffer, pH2.8.

As the proteins appear to be folded, testing their ability to inhibit HIV compared to 5P12-Linker-C37's results was the next step, the preliminary results of which will be reported in the following two sections

4.2.3 Preliminary Pseudoviral Assays comparing 5P12-Linker-T1144 and 5P12-Linker-T2635 with 5P12-Linker-C37

See sections 3.2.5-3.2.7 for more information on pseudoviral assays.

In order to produce HIV pseudovirus (which is incapable of replication after one round of infection), the DNA backbone pSG3$^{\Delta env}$ must be transfected with an envelope gene (*env*, which contains gp41 and gp120) inside of a eukaryotic cell line. Maintaining the same backbone DNA while altering the *env* gene allows for different strains of pseudovirus to be produced. 293FT cells were co-transfected with the pSG3$^{\Delta env}$ plasmid and an envelope plasmid using the XTreme Gene Transfection Reagent (Roche/Sigma-Aldrich). 48 hours post-transfection, the supernatant was collected, centrifuged, and filtered with a 0.45-µm syringe filter. 293FT cells were originally obtained from Invitrogen. The viral stocks were then stored at −80 °C. Viral plasmids used to create pseudovirus, including HIV-1 PVO clone 4 (SVPB11), ADA, Ba-L.1, CAP210, HxB2, and pSG3$^{\Delta env}$,

were obtained from the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, National Institutes of Health; PVO, clone 4 (SVPB11) was from Dr. David Montefiori and Dr. Feng Gao[231]; ADA (R5) was received from Howard Gendelman[253–256]; the plasmid containing full length *env* and *rev* genes of CAP210.2.00.E8, SVPC17 from Drs. L. Morris, K. Mlisana, and David Montefiori[257]; HIV-1 HxB2 Env Expression Vector (pHxB2-env) from Dr. Kathleen Page and Dr. Dan Littman[258]; pSG3$^{\Delta env}$ DNA containing a defective *vpu* gene and truncated, nonfunctional *env* was obtained from Drs. John C Kappes and Xiaoyun Wu.[232–234]

Once the virus was obtained, it was possible to begin pseudoviral assays. HeLa-TZM-bl cells were obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, National Institutes of Health; the HeLa-TZM-bl cell line (a cell line stably expressing human receptors CD4, CXCR4, and CCR5) was from Dr. John C. Kappes, Dr. Xiaoyun Wu, and Tranzyme Inc. HeLa-TZM-bl contain a reporter gene (β-galactosidase) under control of an HIV Long Terminal Repeat (LTR), which begins transcription when it interacts with the HIV transactivator of transcription protein (Tat). As the pseudoviruses contains Tat, infection by these pseudoviruses will result in production of the β-galactosidase enzyme in TZM-bl cells.

HeLa-TZM-bl cells were grown in culture media (10% FBS, penicillin/streptomycin, 25 mM HEPES in DMEM). 10^4 cells per well were seeded into a 96-well plate and allowed to incubate at 37°C overnight. Media was then aspirated and replaced with 50µL fresh media and allowed to incubate for 2 hours. For preparation of inhibitor, a deep-well dish was made with varying concentrations of inhibitor diluted into PBS and mixed thoroughly. 20µL of inhibitor from these dilutions (or 20 µL PBS for positive and negative controls) were added to the TZM cell-containing wells, and the plate was rotated for 1 min to mix, before incubating for 30 min at 37°C. 30µL of single-round virus (described above) in media was then added (with non-viral media to the negative control) for a final volume of 100µL/well and the plate was rotated for 1 min to mix. This “layering” method differs from other similar pseudoviral assays, where the inhibitor is mixed in-well. In our hands, mixing in-well leads to lower IC_{50} (apparent better inhibition), but the “layering” addition leads to more consistency while more closely mimicking conditions under which a topical microbicide may be used. After 20 h, media was aspirated and replaced with 100µL/well of fresh media. After 48 more hours, the media was aspirated and the cells were lysed using 30µL/well of 0.5% NP-40 in PBS. Cells were incubated at room temperature for 25 min to allow for full lysing. Substrate was added then added, 30µL/well (8mM chlorophenol red-β-d-galactopyranoside, 10 mM beta mercaptoethanol, 20 mM KCl in PBS) to be cleaved by β-galactosidase if enzyme was produced. The absorbance signals at wavelengths 570 and 630 nm were measured, and the 570:630 ratio for each well was calculated. All inhibitor combinations were tested in triplicate, repeated two times. The data were plotted using Microsoft Excel, and the IC_{50} value was determined using a linear equation fitted between two data points surrounding 50% inhibition (Table 4.1)

Table 4.1: Preliminary IC^{50} Values of 5P12-Linker Series (in nM) Using Pseudoviral Assays on HeLa-TZM-bl cells. Table showing preliminary IC_{50} values of various viruses inhibited by chimeric proteins, 5P12-RANTES, C-peptide analogs, or a combination of unlinked 5P12-RANTES and the C-peptide analogs variants. Ba-L.1 is a laboratory-created R5, B-clade strain. ADA and PVO.4 are patient-derived R5, B-Clade strains. CAP210 is a patient-derived R5, C-clade strain. HxB2 is a patient derived X4, B-clade strain. HeLa-TZM-bl cells contain both CXCR5 and CCR5 on their surface, which are known to dimerize; in this cell line, it has been seen previously[93] that this dimerization aids 5P12-linker proteins in inhibiting X4 strain viruses, as the 5P12-RANTES portion binds to the dimerized receptor. All experiments were performed in triplicate, two times.

Preliminary IC_{50} Values of 5P12-Linker Series (in nM) Using Pseudoviral Assays on HeLa-TZM bl cells					
	ADA Virus	**Ba.L.1 Virus**	**PVO.4 Virus**	**CAP210 Virus**	**HxB2 Virus**
5P12-L-T-1144	0.420 ± 0.035	0.100 ± 0.006	0.485 ± 0.065	0.107 ± 0.008	0.0210 ± 0.0019
T-1144 Peptide:	41.3 ± 6.0	35.1 ± 3.1	139 ± 5	24.7 ± 2.4	5.55 ± 0.44
5P12 and T-1144, Unlinked	1.77 ± 0.48	1.02 ± 0.06	2.44 ± 0.14	2.48 ± 0.63	5.63 ± 0.41
5P12-L-T-2635	0.467 ± 0.020	0.115 ± 0.007	0.394 ± 0.091	0.180 ± 0.050	0.0194 ± 0.0033
T-2635 Peptide	42.1 ± 2.7	38.1 ± 2.88	142 ± 11	27.3 ± 3.6	5.88 ± 0.28
5P12 and T-2635, Unlinked	1.77 ± 0.48	0.923 ± 0.076	2.59 ± 0.12	3.53 ± 0.91	5.51 ± 0.70
5P12-RANTES	3.24 ± 42	1.48 ± 0.07	2.98 ± 0.48	4.32 ± 0.24	>1000
5P12-L-C37	0.201 ± 0.048	0.0854 ± 0.0038	0.367 ± 0.091	0.119 ± 0.012	0.0114 ± 0.0041

Much like 5P12-Linker-C37[93], the next-generation proteins 5P12-Linker-T2635 and 5P12-Linker-T1144 show consistent improvement over the individual components alone (5P12-RANTES and T-1144 or T-2635) or even in combination (5P12-RANTES in solution with an equal amount of T-1144 or T-2635), showing that linking the inhibitors together results in a synergistic effect. However, despite being T-1144 and T-2635 being newer inhibitors, 5P12-Linker-C37 consistently showed a slightly higher IC_{50} value compared to the next-generation chimeric inhibitors, except for a slight difference with the C-clade virus CAP210 (Figure 4.4). However, the difference between the 5P12-Linker variants is very small, meaning that these new variants are still very potent inhibitors and very viable for clinical use. There is a need, however, to see if there are any advantages over these new chimeric inhibitors versus the older 5P12-Linker-C37.

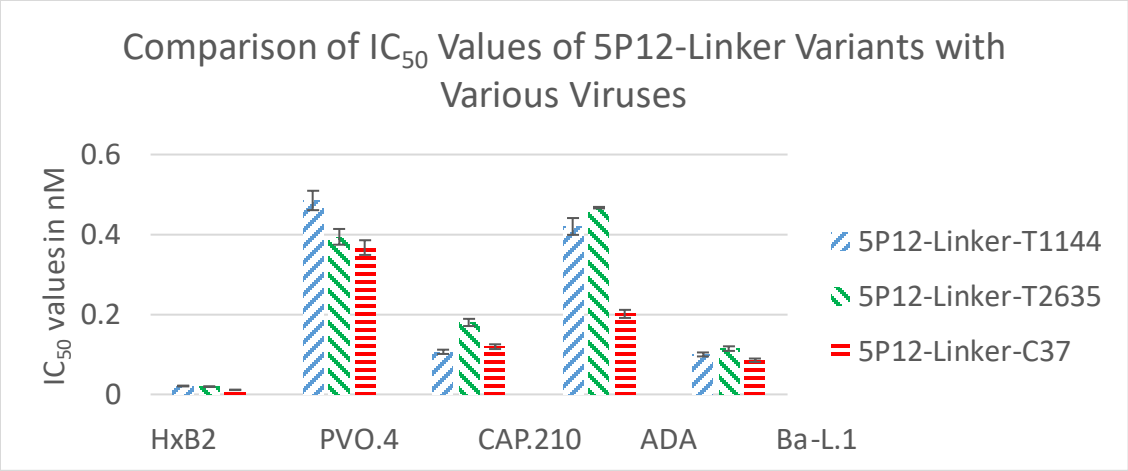

Figure 4.4: Comparison of IC_{50} values of 5P12-Linker Variants with various viruses. 5P12-Linker-T1144 in blue stripes, 5P12-Linker-T2635 in green stripes, 5P12-Linker-C37 in red stripes. Error bars represent standard deviation from two triplicate experiments.

4.2.4 Preliminary Pseudoviral Assays comparing 5P12-Linker-T1144 and 5P12-Linker-T2635 with 5P12-Linker-C37 with C37-Resistant gp41 Mutants

Seeing as how there currently seems to be little advantage of these next-generation chimeric inhibitors over the previous generation in concern with usual viral strains (though more testing is necessary to confirm these results), there is a need to test against C37-resistant viruses to determine whether T-1144 or T-2635 can compensate for the disadvantages of the previous-generation inhibitor. There are several mutations that take place in gp41 that can cause resistance to both T20 (which T-1144 and T-2635 were specifically produced to compensate for[100,101]) and C37; several of such mutations are compiled together in Table 4.2.

Table 4.2: List of gp41 mutations that result in T20/C34 Resistance. A list of gp41 mutations on the virus strain NL4-3 that would result in Slight Sensitivity (0-10 fold different in EC_{50}), Sensitivity (10-100 fold difference in EC_{50}), High Sensitivity (100-200 fold difference in EC_{50},) or Resistance (over 200 fold difference in EC_{50}) to either C34 (a version of C37 missing its three final amino acids often used in studied) or T20. Mutations come from references [259,260].

Mutation	T20 Sensitivity	C34 Sensitivity
D36G/V38A	Resistant	Slightly Sensitive
D36G/V38A/N42D	Resistant	Sensitive
D36G/N42T/N43K	Resistant	Sensitive
D36G/N38E/N42S	Resistant	Sensitive
D36G/V38A/N42T	Resistant	Sensitive
137K	Resistant	Sensitive
I37K/N126K	Sensitive	Sensitive
D36G/I37K/N126K/L204I	Sensitive	Highly Sensitive
D36G/I37K/N126K	Sensitive	Highly Sensitive

Though many HIV *env* genes can be ordered from the AIDS Reagent Program, these specific mutants are either not available or only available in the form of active virus (while our laboratory currently only utilizes single-round virus). So, instead, available HIV *env* DNA was mutated using a two-step PCR method (due to the considerable length of the genome), and the resulting DNA was then utilized to produce pseudovirus for testing using the same method described in section 4.2.3. Although the original mutations in Table 4.2 were done in the viral strain NL4-3[259,260], this *env* gene proved difficult to mutate due to issues regarding non-specific cleavage of the vector. So, instead, the strain PVO.4 was used, as it already contains the D36G mutation necessary for many of these resistance strains, and otherwise contains the same amino acids in the same positions as NL4-3 where the resistance mutations take place. (Figure 4.5)

```
NL4-3       -AVGIGALFLGFLGAAGSTMGAASMTLTVQARQLLSDIVQQQNNLLRAIEAQQHLLQLTV
PVO.4       AVGTLGAMFLGFLGAAGSTMGAASVTLTVQARQLLSGIVQQQNNLLKAIEAQQHMLQLTV
             .  :**:****************:***********.*********:*******:*****

NL4-3       WGIKQLQARILAVERYLKDQQLLGIWGCSGKLICTTAVPWNASWSNKSLEQIWNNMTWME
PVO.4       WGIKQLQARVLAIERYLKDQQLLGIWGCSGKLICTTAVPWNTSWSNKSFNKIWDNMTWME
            *********:**:******************************:******:::**:******

NL4-3       WDREINNYTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFNITNWLWYIKLFIMIVG
PVO.4       WEREIDNYTGLIYNLLEESQNQQEKNEQDLLALDKWESLWNWFSITKWLWYIKIFIMIVG
            *:***:***.**:.*:************:** **** ******.**:******:******

NL4-3       GLVGLRIVFAVLSIVNRVRQGYSPLSFQTHLPIPRGPDRPEGIEEEGGERDRDRSIRLVN
PVO.4       GLIGLRIVFAVLSIVNRVRQGYSPLSFQTHLPTSRGPDRPEGIGGEGGERDRDRSGPLVD
            **:****************************  *********  **********  **:
```

Figure 4.5: NL4-3 and PVO.4 with locations of resistance/sensitivity mutations from highlighted. Mutations mentioned in Table 4.2 highlighted in various colors to show sequence identity of these residues between viral strains PVO.4 and NL4-3. Natural PVO.4 D36G mutation highlighted in blue with red font. Mutation location from reference[260] highlighted in yellow, while mutation locations from reference[259] highlighted in pink.

Of the mutants available from Table 4.2, several of these have produced successful virus, and we have begun preliminary testing with 5P12-Linker-C37, 5P12-Linker-T1144, and 5P12-Linker-T2635 as described in section 2.4.3. Though the current mutants are generally considered mainly T20 resistant, as shown in Table 4.2, C34/C37 has previously shown sensitivity to such mutations with mutated NL4-3. Even though has been previously shown that C34/C37 only has a sensitivity to these mutations, 5P12-Linker-C37 does begin to lose its advantage over the next-generation chimeric inhibitors due to several of these mutations, as shown in Figure 4.6, with the N42T/N43K mutant showing the largest difference between the next-generation inhibitors and the previous-generation inhibitor. Though these experiments must be repeated to confirm the results and continued for other resistant viral strains, it is the beginning of a trend, showing that 5P12-Linker-C37 displays slightly worse inhibition than the next generation inhibitors in cases of certain gp41 mutants.

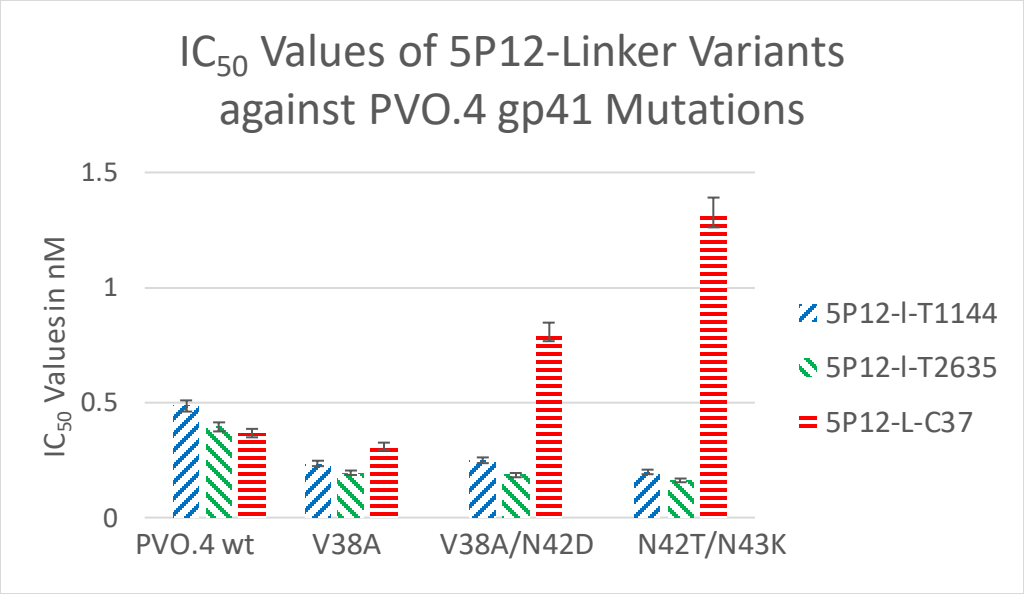

Figure 4.6: IC_{50} of 5P12-Linker variants with PVO.4 Sensitivity Mutants. 5P12-Linker-T1144 in blue stripes, 5P12-Linker-T2635 in green stripes, 5P12-Linker-C37 in red stripes. Error bars represent standard deviation from triplicate experiments done twice.

4.3 Discussion and Future Directions

When designing an HIV inhibitor, it is important to consider those infected but unaware of their infection. It is estimated that one in seven people are unaware of their infection[56], and if these individuals were to take a topical microbicide, it would allow the virus that they are harboring an opportunity to accumulate mutations. As a result, there is a need to make HIV microbicides as broadly-acting as possible.

In an attempt to broaden the inhibitory range 5P12-RANTES, the P. LiWang group linked a broadly-acting C-peptide inhibitor to this protein, creating 5P12-Linker-C37[93]. Although 5P12-Linker-C37 is one of the most potent HIV entry inhibitors described, C37's poor pharmacokinetic properties have prevented C37 from entering clinical trials, and instead, it is usually reserved for use as a research tool[261]. However, the next-generation C-peptide inhibitors T-1144 and T-2635 were designed specifically with superior pharmacokinetic properties and to have an advantage over common T20 resistance mutations[100,101], making these peptides viable candidates for designing next-generation chimeric entry-inhibitors with certain advantages over 5P12-Linker-C37.

According to preliminary tests, despite being second generation chimeric inhibitors, 5P12-Linker-T1144 and 5P12-Linker-T2635 show no appreciable improvement over 5P12-Linker-C37 in pseudoviral assays, currently showing IC_{50} values that are slightly higher than the C37 variant (Figure 4.4). However, these chimeric inhibitors do currently show an improvement over certain viral strains that have lost sensitivity to C37 (Figure 4.6), though more of these strains in Table 4.2 must be tested to see the full effect of these new chimeric proteins over the previous-generation inhibitor.

As there is an advantage of 5P12-Linker-T1144 and 5P12-Linker-T2635 over the previous-generation inhibitor with certain gp41 mutations, there is reason to continue experiments and to test the stability of these next-generation chimeric proteins to see if the pharmacokinetic advantages of the peptides alone translate when linked to 5P12-RANTES. The 5P12-Linker proteins can be tested under various temperature conditions (incubating at high temperatures and various conditions, then monitoring them through NMR techniques and/or functional assays), and their resistance to protein cleavage should also be tested using various protease enzymes (running the results on a protein gel to see what changes to band-sizes can be seen due to the cleavage compared to the unlinked proteins or 5P12-Linker-C37). Comparing the solubility of 5P12-Linker-C37 versus 5P12-Linker-T1144 and 5P12-Linker-T2635 would also be useful, as one of the main concerns with C37 is its poor solubility[246].

If 5P12-Linker-T1144 and 5P12-Linker-T2635 continue to compensate for viral mutations while also showing superior pharmacokinetic properties to 5P12-Linker-C37, these next-generation chimeric inhibitors offer a clear advantage over 5P12-Linker-C37 that could make them viable candidates for a possible theraprutic, and hopefully can be offered as an alternative to individuals who have mutated resistance to T20.

Chapter 5

Design and Purification of a CC Chemokine-Based Activator of HIV Latency

Funding for this work was provided by NIH R01AI112011, as well as funding provided by UC Merced SNS summer fellowships. Thank you to Ricardo Ramos for his assistance in protein preparation, and good luck to Laura Showalter in continuing this project.

5.1 Introduction

Although the life expectancy of those infected with HIV has increased drastically since the discovery of AIDS[56], the treatment for the condition (known as combined antiretroviral therapy, or cART) is both costly and necessary to take throughout an entire lifetime. The treatment also results in certain negative long-term health effects that accumulate over an extended period of time,[102] creating a large demand for a true cure for HIV rather than only focusing on treatment. When cART first started to be used with patients, there was a presumption that over time, patients would eventually be able to deplete the viral stores in their bodies. However, time and time again, it has been shown that if those who take cART (even for years with non-detectable viral load, meaning that the therapy was successfully suppressing the virus) stop their regimen, there is a rapid rebound in viral load.[103] This means that, even after years of treatment, the patients are somehow still harboring cells with viral DNA that is capable of replication, meaning that a cure through cART is not possible. For there to be a true cure for HIV, the reason for this viral rebound must be thoroughly understood.

HIV is a retrovirus, meaning that the viral RNA genome is reverse-transcribed into a DNA template that is then integrated into the host DNA.[104] Once HIV enters the cell by utilizing the human receptors CD4 and CCR5[21] (See section 1.3.1) and the viral contents are released from the membrane, the positive-stranded RNA is reverse-transcribed into DNA (known as the HIV provirus) and then permanently integrated into the human genome of the infected cell.[107] Usually, the provirus is integrated into sites of high transcriptional activity[262], so the DNA quickly begins to produce HIV virions. In a small percentage of infections, however, the integration events result in transcriptional silencing, instead.[108]

These transcriptionally silent cells, also known as “latently infected cells”, have been recognized as the main reason for viral rebound after cART[109]. It has become known that the “latent reservoir”, the cells capable of harboring latent HIV DNA, is more diverse than first expected. $CD4^+$ T-cells were the first cells recognized to be capable of becoming latently infected with HIV, but latent HIV DNA has since also been discovered in dendritic cells, macrophages (including the microglial cells), peripheral blood

monocytes, astrocytes and hematopoietic stem cells.[110] If there was some way to remove the latent reservoir from patient, there would be no more ability to create HIV virions, and thus, the patient would be cured of the condition. As a result, there is great interest in being able to somehow locate and remove or kill the cells that are harboring latent HIV provirus.

One of the leading methods being attempted by scientists is the "Shock and kill" technique[111], where the general idea is to "shock" the latently-infected human genome into expressing the HIV provirus by reversing the transcriptional silencing and then aiding the immune system in removing the cells, now that they can be recognized by the immune system. While there has been some success in cell[112] and animal models[113] many strategies are very broadly acting, affecting all cell types, and many attempts still lead to toxic side effects[114,115]. As a result, there is a need for a more targeted “shock and kill” system, one that could at least narrow down the pool of cells affected by a treatment of transcriptional activation.

A possible strategy would be using a targeted molecule to select only for certain cell types. As many cell types in the immune system are recognized through the types of receptors that are found on their surface[1], targeting certain cell surface receptors seems like a viable strategy for cell specificity. Chemokine receptors can be found on specific immune cell types[116], and though many chemokines bind to more than one receptor, certain chemokines are only specific for one target receptor (see Figure 1.2). The natural diversity and robustness of chemokines and chemokine receptors, as well as the extensive library of modified chemokine ligands with various binding and receptor-activation properties[63,263], make them an ideal system for targeting specific cell types within the immune system.

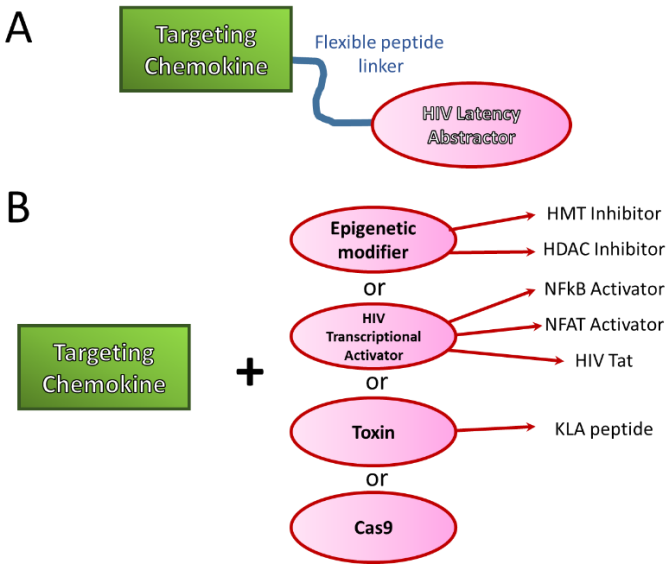

Figure 5.1 Designs for a Chemokine Receptor-Targeting HIV Latency Abstractor. (**A**) A general design for a chemokine receptor-targeting HIV Latency abstractor, where the targeting chemokine is in green and the abstraction agent is in pink, with a blue flexible linker between them. (**B**) Gives specific examples of different ways to remove HIV latency, either by causing transcriptional activation of HIV provirus, cell death, or removing the provirus from the genome.

There is a possibility that a chemokine may be able to be joined to some sort of agent capable of removing the latently infected cell from the reservoir (Figure 5.1A) to result in a chimeric protein capable of both targeting specific cells (through a specific chemokine or chemokine analog) and activate production of HIV virions in the cell (through a transcriptional activator), kill the cell (through a toxin), or remove the proviral HIV DNA (such as by utilizing the CRSPR-Cas9 system) (Figure 5.1B). There are several transcriptional activators that have been shown to be effective in reactivating HIV provirus production in latently-infected HIV cells; Histone Methyltransferase (HMT) inhibitors[264], Histone Deacetylase (HDAC) inhibitors[265], Nuclear Factor Kappa-light-chain-enhancer of activated B cells (NFκB) activators[266], and Nuclear Factor of Activated T-cells (NFAT) activators[267] are some examples. HDAC and HMT can interact with histones, causing them to tighten around DNA and prevent transcription[268]; by inhibiting these proteins, it allows DNA to unwind, allowing transcriptional machinery access to the DNA and allowing production of HIV virions. NFκB is a protein complex that controls transcription of DNA[269] while NFAT constitutes a family of transcription factors[270]; activating either one of these factors up-regulates transcription in a cell, increasing the chance that HIV provirus will be expressed. There is also the possibility to target cells for destruction without activation of the immune system, by using a toxin such as the KLA peptide, which causes cellular apoptosis[271]. Finally, with recent biotechnological developments, there is an increasing chance that soon, the modified bacterial CRSPR-Cas9 system could successfully be used to remove HIV proviral DNA from human cells[272], in which case, using a chemokine to target certain cell types could help efficiency by avoiding uninfected cells.

However, out of the currently-available possibilities, one of the simplest is the HIV trans-activator of transcription protein (Tat), a small HIV protein responsible for naturally activating transcription of HIV proviral DNA[273]. This is one of the few protein-based molecules that has been shown to re-activate HIV transcription in latently infected cells[274], making production of this molecule in a cell possible, unlike many of the small-molecules available from the other listed strategies. Unlike a toxin, Tat would not necessarily kill the cell it enters[275], which is important, as some of the cells that are entered by such a construct may not contain HIV proviral DNA. It also contains a Nuclear Localization Sequence (NLS) region in its protein sequence, allowing for localization into the nucleus once the cell is entered[276], and Tat has the natural ability to escape endosomes through the process of pinocytosis through the protein transduction domain (PTD)[277], which would aid the construct in exiting the endosome produced by a chemokine targeting-protein interacting with its receptor.

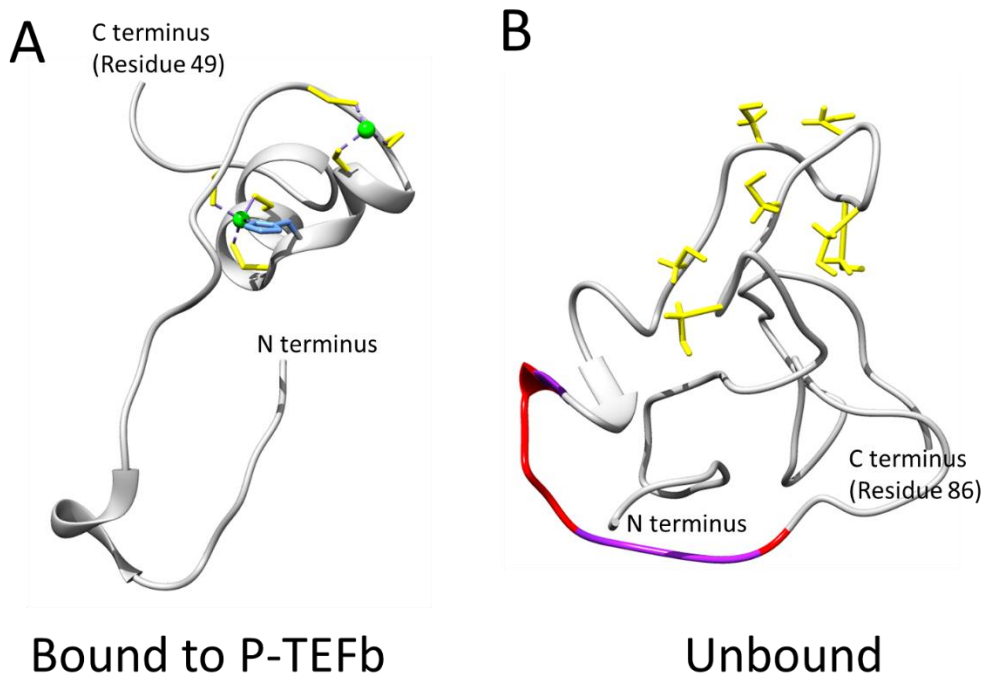

Figure 5.2. Bound and unbound structures of HIV Tat. (**A**) Crystal structure of HIV-Tat when bound to transcription-regulation protein P-TEFb. Cysteines in yellow, histidine interacting with green zinc atom is in blue. Only the first 46 residues formed a shape capable of crystallization when bound to the protein[278]. (PDB code 3MI9). (**B**) Solved NMR structure of unbound HIV Tat protein[279]. (PDB code 1JFW). Cysteine residues are in yellow, with the PTD highlighted in purple, and NLS in red within it.

The Tat protein has been known to be flexible compared to other proteins, making structural studies of it quite difficult. Even when bound to cognate proteins, Tat still resists full crystallization[280,281] (Figure 5.2A), and in solution NMR studies, shows a shape that generally lacks most secondary structure[282,283] (Figure 5.2B). This flexibility, however, appears to result in making purification of the protein from *E. coli* relatively simple[284,285], making it a good candidate for a protein-based cargo for a chimeric chemokine-based HIV latency activator protein.

As for which chemokine to use, many variants with various receptor binding properties have been produced from RANTES[63]. Because CCR5, one of cognate receptors for RANTES, is the main coreceptor for HIV[69], it makes sense for CCR5 to be the target receptor for initial studies of a chemokine-based targeting system. Of the RANTES variants, there is a variant known as 5P14-RANTES, which has been shown to bind CCR5 tightly, but unlike 5P12-RANTES, can also cause internalization of the CCR5 receptor without causing calcium flux from cellular stores[263]. This means that 5P14-RANTES would internalize into the cell, hopefully bringing Tat with it, allowing the PTD region of Tat to aid in escaping the endosome, and then localizing to the nucleus with the NLS.

This chapter details the discovery of a method of purifying the chimeric protein 5P14-Linker-Tat and shows very preliminary results of application of this protein on a Tat-dependent reporter cell strain, HeLa-TZM, usually used for reporting HIV infection[286].

5.2 Methods and Results

5.2.1 Design and Purification of 5P14-Linker-Tat

Many human proteins, especially those that require disulfide bonding, are not capable of proper folding inside of *E. coli*[287]. Disulfide bonds are formed when two cysteine residues come within proximity of one another under oxidizing conditions[137]. In eukaryotic cells, disulfide bonds are formed in the oxidizing environment of endoplasmic reticulum to prepare the proteins for export, organelles which *E. coli* lack, making refolding of proteins containing disulfide bonds inside them difficult[288]. As such, it is traditionally necessary to unfold, reduce, and then refold a protein under the proper redox conditions, if it is to be purified into an active state after expression in *E. coli*.

At times, expressing a recombinant protein inside of *E. coli* can be heavily aided by the use of a fusion tag to the protein of interest. Non-bacterial proteins, especially when overexpressed, tend to be unable to fold properly, which can sometimes lead to internal degradation[289], but expressing these proteins with a fusion tag can allow for part of the conctruct to be folded and avoid degradation, increasing yield of the recombinant protein[290]. Fusion tags are also useful, as they can aid in protein purification, such as by the addition of a 6xHis tag, which would preferentially bind nickel on an affinity chromatography column[291]. There are many fusion tags available, such as the FLAG-tag[292], Thioredoxin[293], Glutathione-S-transferase[294], and Maltose Binding Protein[295], which can help improve expression and solubility of the recombinant protein. One that can be of particular use when attempting to refold a protein for a first time is the SUMO tag. This tag is cut by Ulp-1 protease, which cuts only after the SUMO protein when it shows proper tertiary structure, while many other common cleavage enzymes will often target a specific amino acid sequence, whether it is folded into a proper tertiary shape or not[296]. This not only avoids unspecific cleavage, but by utilizing SUMO as the fusion tag, the protein expression is increased and purification attempts can be monitored for proper SUMO folding before final purification. Being able to judge the likelihood at least part of a protein is folded properly before purification attempts are completely finished can help save considerable time in the search for proper refolding conditions.

Because the 5P14-Linker-Tat protein is a fusion between a protein with two disulfide bonds and another protein with seven free cysteines, utilizing a tag that could suggest proper folding after a cutting reaction while also increasing protein yield was important. As such, the DNA encoding a 6xHis-SUMO-5P14-Linker-Tat construct (GenScript) was placed into the pET-28a expression system in BL21(DE3) *E. coli* cells (see section 2.2.1 for more details about this expression system) to begin attempts at refolding the protein. (Figure 5.3)

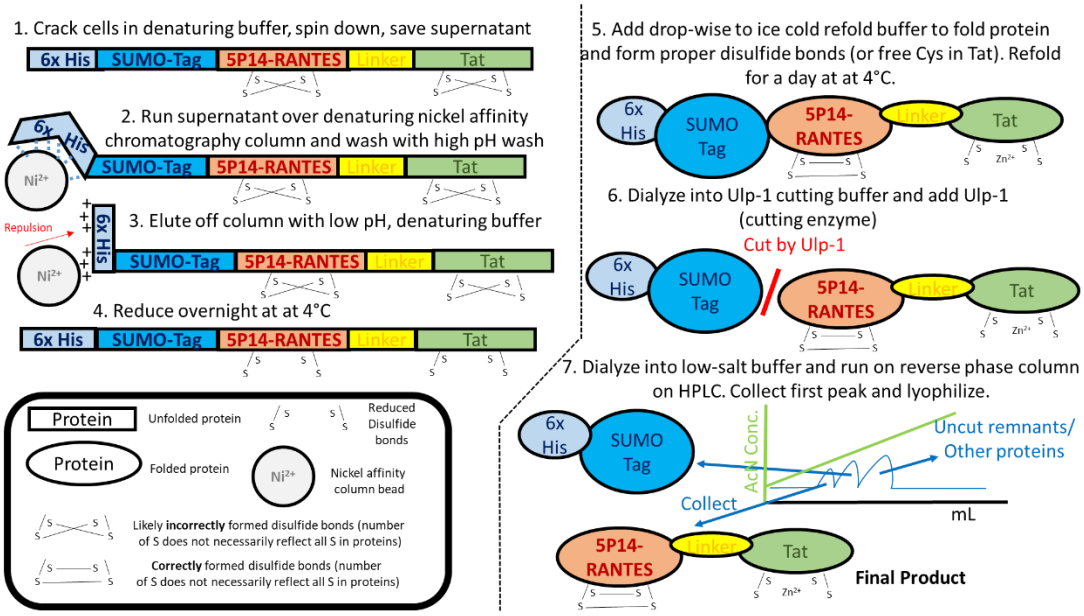

Figure 5.3: General overview of the purification of 5P14-Linker-Tat. A general overview of the purification of 5P14-Linker-Tat. Blue sections represent purification tags, pale orange section represents 5P14-RANTES, yellow section represents the flexible protein linker, and the green section represents HIV-1 Tat from viral strain HxB2. Square sections represent steps where the protein is not yet folded properly, while rounded sections represent properly-folded protein. White boxed section shows legend.

Finding the correct temperature to express a recombinant protein can also be important for final yield[297]. As this is a complex protein, lowering the temperature was expected to yield better results, so 6xHis-SUMO-5P14-Linker-Tat expression at 37°C, 25°C, 22°C, and 18°C degrees were all tested, with 22°C yielding the best results. With lower temperatures, a longer time of incubation is often necessary for optimum protein yield[297], and with 22°C, 20 hours showed best expression of the desired protein.

As human CC chemokines do not naturally fold in *E. coli*, it would be preferred to push the protein into the inclusion body in order to maintain a smaller volume to after lysing the cells. Unfortunately, cell lysing tests in 50 mM, 250 mM, 500 mM, and 1M NaCl with 50 mM Tris pH8 buffers all resulted with a relatively even distribution of the expressed protein between supernatant and inclusion body, so the cells were ultimately resuspended in 6M Guanidine HCl, 200 mM NaCl, and 50 mM Tris pH8 in order to obtain all protein in the lysed supernatant, as well as to unfold all proteases within the cell by utilizing guanidine[298] to protect 5P14-Linker-Tat from protease degradation. The cells were then lysed via three passes on a French press at 16,000 psi until the solution was smooth and then centrifuged at 27000 × *g* for 1 h at 4°C to separate membrane components from the supernatant; cooling the sample before centrifugation aids in precipitation of lipid and removal from the supernatant, which can result in smoother chromatography.

As the SUMO fusion tag was preceded by a 6xHis tag, the lysed supernatant was then passed through a nickel chelating column equilibrated with resuspension buffer. The column was then washed with 50 mL more cracking buffer, and then washed with 50 mL 6M Guanidine, 200 mM NaCl, 80 mM NaOP pH7.2 to wash off nonspecific

binding. As histidine's pK_a is very close to 6, proteins will sometimes come off the column if a pH too close to the pK_a is reached[299], and with 304 amino acids in the final 6xHis-SUMO-5P14-Linker-Tat construct, it was considered too large to risk a lower pH wash, when Nickel columns already showed many non-specific proteins leaving with the flow-through from the original loading. Finally, the protein was eluted from the column using 6M Guanidine HCl, 200 mM NaCl, 60 mM AcO pH4. To pool fractions containing eluted protein together, a nanodrop reading at A_{280} was taken, and fractions showing protein were pooled together for disulfide reduction.

As this protein has many possible disulfide combinations (seven cysteines from Tat and four from 5P14-RANTES that could have combined in any order), there is a need for a thorough reduction before attempting to refold. Though addition of 10 mM βME or 1 mM DTT at room temperature for an hour is usually enough for many chemokines[201], it has been seen that with proteins with a high number of possible disulfide combinations, a longer reduction time with a higher concentration of reducing agent are generally necessary[300]. As such, the eluted fractions were reduced in 30 mM βME at room temperature for two hours, then allowed to continue reducing overnight at 4°C as well. As both CC chemokines[201] and HIV Tat are known to behave favorably at low pH[301], there was no need to adjust pH before reduction.

Finding the proper refold conditions is often the most difficult step of refolding a protein from the inclusion body, especially one that must form proper disulfide bonds. Salt levels, pH, and temperature are all important when considering conditions for refolding proteins. Recently, with more complex proteins being made more often in foreign organisms, many additives that have been shown to aid in refolding of a protein have become available, from amino acids, to sugars, to detergents, with the volume of refold buffer also resulting in different success in folding of more complex proteins[302]. It is also important to consider the specific proteins being folded, such as if there are any cofactors involved of other, less-common additives that may help protein stabilization. For example, it's known that the HIV Tat cysteines bind zinc ions[303] (Figure 5.2A), and as such, adding zinc may aid in proper folding. Varying concentration of reducing agent or redox pair present during the refold conditions has also been shown to result in different levels of successful disulfide binding as well[304]. Many proteins will only find their correct, lowest-energy state under specific conditions, so scouting for them is an important step of the refold process.

Table 5.1 lists the conditions scouted during the His-SUMO-5P14-Linker-Tat refold process. Different salt conditions, additives, pHs, and reducing conditions were attempted throughout, as detailed in the table. While scouting for refold conditions, it is important to look for signs of precipitation during the refold process. It is also known that lower temperatures[305] and high buffer volume[306] (10-20x the volume of protein added) can aid in refolding of complex proteins by helping reduce aggregation, so all refold attempts were attempted with slow dripping into at least 15x buffer with stirring with the buffer at an ice-cold temperature. Precipitation when dripping into refold buffer is a sign of aggregation (though not all aggregation is visible to the naked eye)[307], and thus, improper folding, so precipitation during the refold process is a sign that the refold buffer is suboptimal. All refold attempts seemed clear after stirring for an entire day at 4°C to allow for a slow refold in the refold buffers, but subsequent dialysis into buffer that would allow for efficient Ulp-1 cutting (200 mM NaCl, 20 mM Tris, pH8) sometimes showed signs of precipitation. Refold conditions showing aggregation after dialysis were

discarded, as this showed that a considerable percentage of the protein had not reach its lowest-energy state during the refolding attempt.

Table 5.1: List the refold conditions scouted during the His-SUMO-5P14-Linker-Tat refold process. The numbering system, X.Y.Z corresponds to attempt number, X, meaning the number of overall attempts, during which several conditions were scouted at once to compare to one another. The buffer number, Y, corresponds to a different buffer condition, as listed under "Refold conditions". Z corresponds to whether or not the refold attempt contained Zinc, 1 meaning it did, and 2 meaning it did not.

Table 5.1: Refold attempts of 5P12-His-SUMO-Linker-Tat				
Attempt Number	**Refold conditions**	**With or Without Zinc**	**Refold Buffer volume**	**End result**
1.1.1	550 mM L-Arginine, 250 mM NaCl, 1 mM EDTA, 60 mM NaOP pH7.5, 1mM HSG, 0.1 mM GSSG	With	15x	Cut with Ulp-1 and showed protein on C4 – however, had done so with slight precipitation and unclear C4 result. Seemed like higher pH was best out of first attempt and will be continued to attempt 2.
1.1.2	550 mM L-Arginine, 250 mM NaCl, 1 mM EDTA, 60 mM NaOP pH7.5, 1mM HSG, 0.1 mM GSSG	Without	15x	Cut with Ulp-1 and showed protein on C4 – however, had done so with slight precipitation and unclear C4 result. Seemed like higher pH was best out of first attempt and will be continued to attempt 2.
1.2.1	50 mM MES, 9.6 mM NaCl, 0.4 mM KCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 550 mM L-Arginine, 0.05% PEG 3550, 1 mM GSH, 0.1 mM GSSG pH6	With	15x	Heavy precipitation after third dialysis.
1.2.2	50 mM MES, 9.6 mM NaCl, 0.4 mM KCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 550 mM L-Arginine, 0.05% PEG 3550, 1 mM GSH, 0.1 mM GSSG pH6	Without	15x	Heavy precipitation after second dialysis.
1.3.1	50 mM MES, 240 mM NaCl, 10 mM KCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$, PEG 3550, 1	With	15x	Heavy precipitation after first dialysis.

	mM GSH, 0.1 mM GSSG pH6			
1.3.2	50 mM MES, 240 mM NaCl, 10 mM KCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$, PEG 3550, 1 mM GSH, 0.1 mM GSSG pH6	Without	15x	Precipitation after first dialysis.
2.1.1	240 mM NaCl, 10 mM KCl, 1 mM EDTA, 400 mM Sucrose, 1 mM GSH, 0.1 mM GSSG, 50 mM NaOP pH 7.5	With	15x	Precipitation after third dialysis
2.1.2	240 mM NaCl, 10 mM KCl, 1 mM EDTA, 400 mM Sucrose, 1 mM GSH, 0.1 mM GSSG, 50 mM NaOP pH 7.5	Without	15x	Precipitation after third dialysis
2.2.1	9.6 mM NaCl, 0.4 mM KCl, 2 mM CaCl2, 2 mM MgCl2, 550 mM L-Arginine, 400 mM Sucrose, 1mM GSH, 0.1mM GSSG, 50 mM NaOP pH7.5	With	15x	Heavy precipitation after first dialysis
2.2.2	9.6 mM NaCl, 0.4 mM KCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 550 mM L-Arginine, 400 mM Sucrose, 1mM GSH, 0.1mM GSSG, 50 mM NaOP pH7.5	Without	15x	Heavy precipitation after second dialysis
2.3.1	240 mM NaCl, 10 mM KCl, 1 mM EDTA, 0.05% PEG 3550, 1 mM GSH, 0.1 mM GSSG, 50 mM NaOP pH7.5	With	15x	Cut with Ulp-1 and showed protein on C4 – 1D NMR showed no signs of folded protein. Seemed like higher pH and higher salt are best, but needs more additives.
2.3.2	240 mM NaCl, 10 mM KCl, 1 mM EDTA, 0.05% PEG 3550, 1 mM GSH, 0.1 mM GSSG, 50 mM NaOP pH7.5	Without	15x	Cut with Ulp-1 and showed protein on C4 – 1D NMR showed no signs of folded protein. Seemed like higher pH and higher salt are best, but needs more additives.
2.4.1	240 mM NaCl, 10 mM KCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 400 mM Sucrose, 1 mM GSH,	With	15x	Cut with Ulp-1 and showed protein on C4 – 1D NMR showed slight signs of folded protein. Seemed like higher pH and higher salt are best,

	0.1 mM GSSG, 50 mM NaOP pH 7.5			but possibly different combination of additives. Continued to attempt 3.
2.4.2	240 mM NaCl, 10 mM KCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 400 mM Sucrose, 1 mM GSH, 0.1 mM GSSG, 50 mM NaOP pH 7.5	Without	15x	Cut with Ulp-1 and showed protein on C4 – 1D NMR showed slight signs of folded protein, but less than prep with Zn. Seemed like higher pH and higher salt are best, but possibly different combination of additives. Continued to attempt 3.
3.1.1	240 mM NaCl, 10 mM KCl, 2 mM MgCl2, 2 mM CaCl2, 550 mM L-Arginine, 1 mM GSH, 0.1 mM GSSG, 50 mM Tris pH 8	With	15x	Slight precipitation before C4, compared to perfectly clear 3.2 and 3.3 attempts, so was stopped despite proper Ulp-1 cutting.
3.1.2	240 mM NaCl, 10 mM KCl, 2 mM MgCl2, 2 mM CaCl2, 550 mM L-Arginine, 1 mM GSH, 0.1 mM GSSG, 50 mM Tris pH 8	Without	15x	Slight precipitation before C4, compared to perfectly clear 3.2 and 3.3 attempts, so was stopped despite proper Ulp-1 cutting.
3.2.1	240 mM NaCl, 10 mM KCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 550 mM L-Arginine, 400 mM Sucrose, 1 mM GSH, 0.1 mM GSSG, 50 mM Tris pH 8	With	15x	1-D NMR showed slight signs of protein peaks, but weaker than 3.3.1 and 3.3.2.
3.2.2	240 mM NaCl, 10 mM KCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 550 mM L-Arginine, 400 mM Sucrose, 1 mM GSH, 0.1 mM GSSG, 50 mM Tris pH 8	Without	15x	1-D NMR showed slight signs of protein peaks, but weaker than 3.3.1 and 3.3.2.
3.3.1	240 mM NaCl, 10 mM KCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 550 mM L-Arginine, 400 mM Sucrose, 1 mM GSH, 0.05%PEG 3350, 0.1 mM GSSG, 50 mM Tris pH 8	With	15x	1-D NMR showed strongest signs of protein peaks. Additives seem good, but learned that a higher concentration of GSH and GSSG can help with proteins with many disulfides, so will attempt in prep 4. 2-D HSQC NMR showed small signs of 5P14-RANTES folding, but

				many unfolded peaks along the center.
3.3.2	240 mM NaCl, 10 mM KCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 550 mM L-Arginine, 400 mM Sucrose, 0.05%PEG 3350, 1 mM GSH, 0.1 mM GSSG, 50 mM Tris pH 8	Without	15x	1-D NMR showed second-strongest signs of protein peaks. Additives seem good, but learned that a higher concentration of GSH and GSSG can help with proteins with many disulfides, as well as higher volume of refold buffer, so will attempt in prep 4.
4.1.1	240 mM NaCl, 10 mM KCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 550 mM L-Arginine, 400 mM Sucrose, 0.05%PEG 3350, 3 mM GSH, , 0.3 mM GSSG, 50 mM Tris pH 8	With	20x	2-D HSQC NMR showed good folded peaks with a part of the spectrum that corresponds to 5P14-RANTES.
4.1.2	240 mM NaCl, 10 mM KCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 550 mM L-Arginine, 400 mM Sucrose, 3 mM GSH, 0.05%PEG 3350, 0.3 mM GSSG, 50 mM Tris pH 8	Without	20x	2-D HSQC NMR showed good folded peaks with a part of the spectrum that corresponds to 5P14-RANTES, but with slightly more unfolded peaks than with Zinc.

The first attempt seemed to show best results with the highest pH used (pH7.5), and so the second attempt continued with a higher pH. In the second attempt, it was clear that low-salt conditions were unfavorable protein-folding conditions, and so the third attempt focused on higher salt conditions and higher pH, while focusing on possible refolding additives. After the third attempt, the reducing-oxidizing pair GSH and GSSG (glutathione, reduced and oxidized) was tripled in concentration, as a higher concentration had shown success in other disulfide-rich proteins[304], and the refold volume was also increased to allow for greater dispersion of protein in the solution. This ultimately led to what appeared to be a folded variant of 5P14-Linker-Tat.

5.2.2 NMR Spectroscopy of 5P14-Linker-Tat

For a more detailed explanation on NMR spectroscopy and HSQC spectra, see section 2.2.2.

NMR samples of 60 µM 5P12-Linker-Tat or 25 µM 5P14-RANTES (expressed and purified using the same method as 5P12-Linker-C37 in section 4.2.1) were run at 25°C in 20 mM sodium phosphate buffer with 10% D_2O and 5 µM 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) with a final pH of 2.8 to check proteins for proper folding. All HSQC NMR data were acquired on a four-channel 600 MHz Bruker Avance III spectrometer equipped with a GRASP II gradient accessory and a TCI cryoprobe with

an actively shielded Z-gradient coil. The chemical shift was referenced relative to internal DSS[162]. The data were processed using NmrPipe[163] and analyzed using PIPP[164]. For HSQC spectra, sweep width was 8474.576(^{1}H) and 1766.784 Hz (^{15}N), with 1280 points in ^{1}H and 128* (256 total) points in ^{15}N.

5P14-Linker-Tat showed a folded spectrum (Figure 5.4, black) that overlaps almost perfectly with the 5P14-RANTES spectrum(Figure 5.4, red) (78/82 peaks overlapping with 5P14-RANTES including tryptophan and amide side chains), suggesting that 5P14-RANTES is fully folded within the spectrum. The spectra also show at least 65 more backbone peaks not corresponding to 5P14-RANTES, as well as some unfolded regions in the center, which likely corresponds to the flexible 10-amino acid linker and the flexible region of the 86-amino acid Tat, especially with the proper size for the full-length protein being seen on an SDS-PAGE gel.

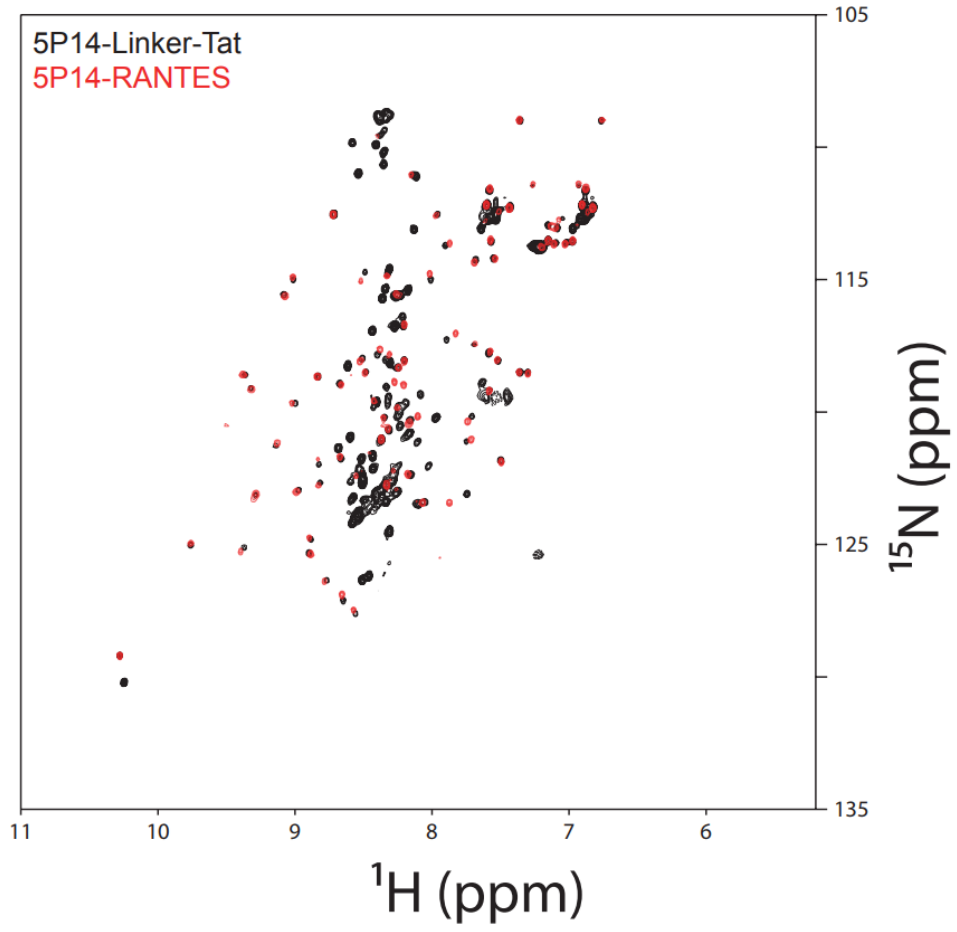

Figure 5.4: HSQC NMR spectrum of 5P14-Linker-Tat. ^{15}N-Labeled HSQC spectrum of 60 µM 5P12-Linker-Tat in black with 25 µM 5P14-RANTES spectrum overlaid in red. All spectra were run at 25°C in 20 mM NaOP Buffer, pH2.8.

5.2.3 Preliminary Results for HIV LTR Activation Assay

With the protein purified and showing folding, testing its activity was the next step. HeLa-TZM-bl (also known as JC53-bl) cells, which are used in the pseudoviral assays of chapters 3 and 4, contain an HIV viral promotor Long Terminal Repeat (LTR)

that controls transcription of the genes that follow it (β-Galactosidase or Luciferase, rather than the genes necessary for HIV virion production[231,286].) Tat released from pseudovirus recognizes and binds to the LTR and stimulates transcription of these genes by aiding in recruitment of pTEFb to RNA polymerases[308]. This results in β-Galactosidase production, which is then be detected by addition of Chlorophenol red-β-D-galactopyranoside (CPRG), as β-Galactosidase cleaves this substrate to create a color change from yellow to red[309]. As it is HIV Tat that promotes transcription at the viral promotor, it was hypothesized that Tat alone should be able to enter the cells and promote production of β-Galactosidase, which could be detected using CPRG.

An assay was set up under the same set-up as the pseudoviral assays described in sections 3.2.7 and 4.2.3. HeLa-TZM-bl cells were grown in culture media (10% FBS, penicillin/streptomycin, 25 mM HEPES in DMEM), and then 10^4 cells per well were seeded into a 96-well plate and allowed to incubate at 37°C overnight. Media was aspirated and replaced with 50µL fresh media and allowed to incubate for 2 hours. Protein samples to be tested on the cells (HIV-1 IIIB Tat Recombinant Protein alone, which was obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH[310–315], 5P14-RANTES, 5P14-Linker-Tat, and 5P14-RANTES with Tat without linker) were diluted first in PBS to resuspend lyophilized proteins, and then an appropriate volume of this mixture was added to culture media to attain double the concentrations seen in Figure 5.5. 50µL the mixtures in media were then added (with plain media to the negative control) for a final volume of 100µL/well and the plate was rotated for 1 min to mix. After 20 h, media was aspirated and replaced with 100µL/well of fresh media. After 48 more hours, the media was aspirated and the cells were lysed using 30µL/well of 0.5% NP-40 in PBS and then incubated at room temperature for 25 min. Substrate was added (30µL/well; 8mM chlorophenol red-β-d-galactopyranoside, 10 mM beta mercaptoethanol, 20 mM KCl in PBS) and the absorbance signals at wavelengths 570 and 630 nm were measured as the color change from yellow to red occurred, and the 570:630 ratio for each well was calculated and then had the negative control subtracted. 5P14-Linker-Tat and 5P14-RANTES alone were tested in triplicate, but due to very limited amount of free Tat protein available, samples with Tat alone were tested in duplicate for these preliminary results. The signals were then normalized to the Tat alone signal, in Figure 5.5.

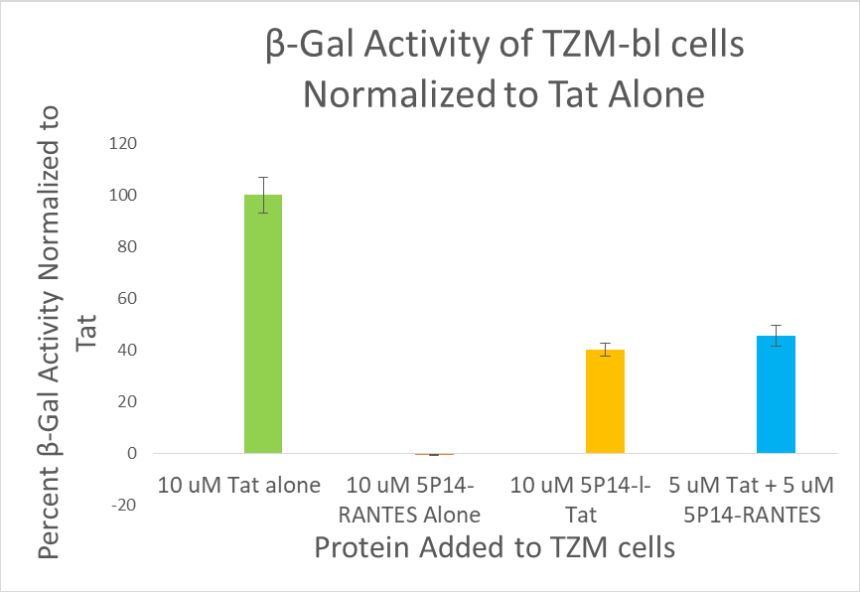

Figure 5.5: Preliminary results of HeLa TZM-bl β-Galactosidase transcription activation due to 5P14-Linker-Tat. 10 μM Tat (Green bar, error bar only from a duplicate test) was normalized to 100%, which other proteins were compared to. 5P14-RANTES (red) and 5P14-Linker-Tat were both tested in triplicate, while 5 μM Tat with 5 μM 5P14-RANTES could only be tested in duplicate due to limited volumes of free Tat.

Tat alone showed activation of β-Galactosidase production as expected, as it is known that small amounts of Tat can enter the cell and nucleus due to the PTD and NLS.[314] 5P14-RANTES alone showed no activation, as expected, as this protein would merely bind to CCR5 but have no way of activating transcription on its own. Interestingly, 5P14-RANTES in combination with Tat alone also gives a strong signal, though due to the concentration difference in the Tat as a result of limited protein availability, it is impossible to say whether the activity is higher than Tat alone or not without a test at the same concentrations. 5P14-Linker-Tat gave a similar result as 5P14-RANTES and Tat alone unlinked, albeit at double the concentration, showing that the 5P14-Linker-Tat version has more difficulty entering the cell than Tat alone, which for a selective protein, may be considered advantageous, as described in section 5.3.

5.3 Discussion and Future Directions

There is a need for new therapeutics utilizing in the shock and kill method that can be more selective than previously available therapeutics. Here, a possible novel therapeutic is introduced, one that may be able to selectively target $CCR5^+$ cells with the modified N-terminal RANTES/CCL5 mutant, 5P14-RANTES, for reactivation of HIV proviral transcription utilizing HIV Tat, a protein the viral naturally uses to activate HIV transcription. The protein 5P14-Linker-Tat was expressed and purified from *E. coli* and showed a folded spectrum (Figure 5.4) despite 5P14-RANTES containing two disulfide bonds but Tat containing seven free cysteine residues.

When the 5P14-Linker-Tat protein was tested on HeLa-TZM-bl cells, the protein showed activation of β-Galactosidase production, though a lower level of activation compared to Tat alone. These results suggest there may be promise to this method of protein delivery, as Tat alone is known to enter cells nonselectively and can escape the

endosome to then enter the nucleus[316]; however, the lower level of activation suggests there may be barrier to entry involving 5P14-Linker-Tat first binding to CCR5 before being able to enter the cell, which may be a hopeful result for their protein as a future therapeutic.

Unfortunately, 5P14-Linker-Tat also showed signs of cell toxicity at higher concentrations of use, which may be due to retention of certain cell-toxic molecules during purification, so this toxicity must also be considered when continuing this project. A different purification method utilizing molecules that are not toxic to eukaryotic cells may be necessary to attain better results, to ascertain whether the toxicity is due to Tat binding to toxic substances during the purification process or this is due to Tat's natural level of toxicity[9]. Also, as Tat is a vital control for such assays, Tat will be produced in the laboratory to be able to attain and utilize higher concentrations of the proteins, as well as allow for the possibility of utilizing certain Tat mutations[317–319] that have shown a better ability to activate transcription of HIV in latent cells. More assays must be completed in order to verify the preliminary results presented in Figure 5.5, as well as an assay showing specificity towards CCR5 cells versus other cells under an LTR promotor, which would show if the 5P14-RANTES section of the protein does, indeed, select more towards cells that contain its cognate receptor and does not allow Tat to move as nonspecifically as it does when it is free.

Overall, though this is a mere beginning to the possibility of utilizing CC chemokines as a targeting system to bring in cargo capable of reactivating latently infected HIV cells, it shows that purification of such complex proteins is possible from an *E. coli* host and the preliminary results show a very limited amount of success in attenuating Tat's ability to simply enter any available cell.

Chapter 6

Conclusions and Future Directions

The CC chemokine system is one of the most complex and diverse parts of the immune system. The system's involvement in disorders from arthritis[18] to atherosclerosis[18] to depression[33] to HIV infection[69,70] makes it of considerable interest to the medical community. Though protein-based medicines are still only a small percentage of prescribed medications, they are beginning to become a more common therapeutic method and are quickly gaining more acceptance[320]. As such, it is important to characterize as much about the proteins involved in the chemokine system as possible so that new therapeutic strategies may be discovered, as the chemokine system is so heavily involved in various diseases and disorders.

In this dissertation, there was a focus on both characterization of proteins involved in the CC chemokine and development of CC chemokine-based therapeutic methods. In Chapter 2, the structures of two proteins that subvert the inflammatory response of CC Chemokines, vCCI and vMIP-II, are studied in more detail. The study of these proteins elucidates certain residues that appear to be important for their binding; as vMIP-II evolved to specifically mimic an array of CC chemokines, while vCCI evolved to bind a vast array of CC chemokines, the study of the interaction between these two proteins will reveal key residues important to the general functions of CC chemokines. Such interactions can likely elucidate more about how CC chemokines generally interact with their receptors, as well as revealing certain regions of CC chemokines that may prove to be important locations for inhibiting these proteins in cases of over-active inflammation.

Chapter 2 reveals several important and interesting interactions between vCCI and vMIP-II, meaning that the next step is to study these interactions further, both through mutational studies and further MD simulations. The first step would be to produce a ^{13}C-laveled vCCI variant to be able to accurately assign the bound peaks of their protein when bound to various chemokines. Once assigned, there would be no need for a conservative estimate of peak shifts, and would allow one to compare large chemical shifts seen in vCCI when bound to various CC chemokines. The peaks that show largest chemical shifts could then be mutated and, if vCCI folding is unaffected (as could be seen through lack of large shifting in the unbound form on an HSQC), the bound state of this mutant can then be compared to the bound states not just with the particular chemokine that caused the large chemical shift change, but with other chemokines as well. Any differences and similarities should then be compared to the amino acid sequences of the CC chemokines and to the known bound configurations of vCCI:MIP-1β or vCCI:vMIP-II complexes to find amino acids most likely involved in these

interactions. These interactions can then be studied through molecular dynamics simulations before attempting ^{13}C labeling of chemokines or vCCI mutants to view these interactions in more detail. There is also the possibility of attempting a crystal structure of vMIP-II bound to vCCI, now that great quantities of vCCI can be produced, as the simulation suggests a great area of interaction between the usually-flexible acidic loop and vMIP-II, as well as a great area of interaction with the usually-flexible region of vMIP-II with vCCI, as well.

There is also room to simultaneously explore some of the specific amino acids of interest found in the simulations. For example, seeing how many residues are hydrophobic in the region of position 41 in several chemokines may reveal the importance of the I41(vMIP-II)/I184(vCCI) interaction (Figure 2.12A), and whether the simulation was correct about this hydrophobic interaction being the reason for loss of function when I184 of vCCI is mutated. Mutating I41 to a smaller, less hydrophobic residue such as alanine, or to a hydrophilic residue such as threonine, and studying effects on binding, either through NMR (comparing fast versus slow exchange) or ITC could reveal accuracy of the simulation. Also, mutating Y80 would be an interesting experiment; while this residue interacts with K48 of MIP-1β, no similar interaction was seen with vMIP-II (Figure 2.12B). Mutating this residue (such as to alanine) and seeing whether it maintains its folded conformation may reveal whether this residue is vital for interaction with residues in position 48 for other chemokines (which could also be tested with other chemokines with positively-charged residues at position 48), or whether this residue is necessary for propping open the acidic loop, instead, and loss of this residue results in collapse of the folded structure of the chemokine (which would result in a highly-different HSQC fingerprint compared to wild type unbound vCCI).

The effects of N-terminal cyclization of the CC chemokine based HIV entry inhibitor, 5P12-RANTES, was studied in detail in Chapter 3. As this chemokine derivative is intended to be used as a biotherapeutic, it is important to characterize this protein's structures as much as possible, not only for their specific medical use, but for development of future therapeutics, as well. Seeing that the cyclization of the 5P12-RANTES N-terminus does not appear to affect its potency or signaling properties, it allows for less focus on cyclization of the terminal Q residue when preparing this protein as a therapeutic agent. This also shows that a natural N terminus at the end of a RANTES derivative is still capable of subverting cell signaling while promoting tight binding with the receptor. This suggests that there may be more combinations of potent N-terminal residues that could result in tight binding but no downstream signaling, showing that more study of the effects of the N-terminal residues of CC chemokines with their receptors is still necessary for fully understanding of their interactions.

As such, more targeted mutations studies into the interaction of the N-terminus or a chemokine and the CC receptor can be performed. With nanodisc technology[321], where a membrane protein can be put on uniform lipid discs that can contain one receptor per disc, it would be possible to study the CCR5 receptor through NMR. Though it is difficult to visualize a protein as large as 40 kDa, there are several techniques available that could make this protein visible and capable of assignment on NMR.[322,323] However, just being able to visualize changed between the bound and unbound spectra of RANTES variants with an unlabeled CCR5 receptor would yield interesting structural information, especially for changes at the N terminus. Comparing RANTES mutants with different signaling properties, such as 5P12-RANTES which

causes no calcium signaling nor internalization of receptor[63], to 5P14-RANTES which does causes no calcium influx but does cause receptor internalization[63], and wild type RANTES could reveal information on the differences between how these N-termini interact differently with the receptors. Any large differences in shifts of the N terminus could be compared to the crystal structure of the 5P7-RANTES variant bound to CCR5[28] (Figure 1.3B) to find relevant residues in CCR5, and mutants of the receptor could be created to see if the mutants result in causing similar shifting between the RANTES variants on the NMR spectra. Also, as nanodisc technology improves, it may soon be possible to also couple a G-protein to CCR5[324], allowing for study of RANTES with the receptor properly couples with its downstream signaling partner.

Chapters 4 and 5 focused instead on creating and purifying CC chemokines-based proteins to engineer novel therapeutic methods. A previous-generation chimeric protein, 5P12-Linker-C37[93], showed some of the best inhibition of HIV of any available entry inhibitor. However, C37's poor pharmacokinetic properties make this chimeric inhibitor a sub-optimal option for clinical purposes. As a result, the next-generation inhibitors T-1144 and T-2635, were used to replace C37 to create 5P12-Linker-T1144 and 5P12-Linker-T2635. These proteins were purified and showed to be folded, and then were used for several preliminary studies. While 5P12-Linker-T1144 and 5P12-Linker-T2635 do not show an appreciable difference with common wild type viruses, the fact that they are showing an advantage over 5P12-Linker-C37 with T20 resistance mutations shows that there is a reason to continue studying these proteins. As such, confirming the results see in Chapter 4 and further testing with various viral clades, such as A and D, should be done to see if there is any variation seen between these variants with more widely-different viral clades. Also, further mutants from table 4.2 should be created and tested on the three variants, to see if the trend of making up for resistance to C37 continued. As T20 is already approved and in use as a current treatment[96], there is a great need for inhibitors that will be able to combat this protein's weaknesses, while having superior pharmacokinetic properties to C37, which would make these next generation chimeric proteins more likely to be approved for clinical use, should they continue.

Other than the inhibition of HIV, the solubility of these variants should also be compared to one another, as well, to see if the increased solubility is still true after attachment to 5P12-RANTES. Two methods used to measure protein solubility in aqueous solution are: 1), adding lyophilized protein to solvent; and 2), concentrating a protein solution by ultrafiltration. Both of these methods require that the concentration of protein in solution be increased until saturation is reached[324], allowing for simple comparison of the chimeric variants. The resistance of these proteins against degradation should also be tested; incubating the proteins in a solution with biological pH at high temperatures, and then comparing their HSQC spectra to see signs of unfolding, could compare their thermal stability. Incubating the proteins with various proteases that may be found on the skin, in the rectal cavity, or in the vaginal cavity and then comparing the results at various points of incubation on a protein gel or through mass spectroscopy can be used to see if the next-generation variants are more resistant to protease activity than 5P12-Linker-C37, as well. These proteins can also be compared to 5P12-RANTES alone, to see whether the C-peptide analogs offer any extra protection to the inhibitor. Discovering whether these proteins truly do have superior pharmacokinetic properties can reveal whether they are more qualified to be biotherapeutics in the future.

In Chapter 5, the focus is the design and study of a protein capable of targeting specific subtypes of latently infected cells that could contain HIV provirus, as most current methods would activate transcription on a very general level in all cell types, including those that do not contain HIV proviral DNA. The protein 5P14-Linker-Tat was designed and purified, with the hypothesis that the 5P14-RANTES portion of their protein will be capable of selectively targeting cells with the CCR5 receptor and cause receptor internalization, while the HIV Tat portion of the protein would allow endosomal escape, translocation into the nucleus, and activation of proviral transcription. Although purification of the protein comes with difficulties, the fact that such a protein shows folding as well as preliminarily showing that a small amount of activation of transcription from HIV LTR proves hopeful. These findings suggest that the strategy for targeting cells with a chemokine should still continue to be further explored, if not only for the case of HIV, but other possible epigenetic conditions that may require cell-type targeting.

As such, different versions of the latency-activating chimeric proteins should be attempted. While 5P14-RANTES will target CCR5, certain latent HIV reservoirs may have more specific surface proteins. In these cases, finding which surface proteins are specifically located or heavily upregulated on cell types most likely to contain latent provirus DNA is vital, and finding the ligands to these receptors can result in various different versions of possible targeting molecules that may be able to replace 5P14-RANTES for different cases of latent HIV infection. Also, although Tat is useful for activating latent HIV provirus, there is the possibility for attaching nanoparticles to the chemokine receptor-targeting proteins, allowing for delivery of non-protein based HIV-reactivating agents[325]. Also, as CRISPR-Cas9 technologies continue to improve[326], it may be possible to use this system as the cargo to be delivered by attaching it to a chemokine or chemokine analog, so that the system enters the correct cell types, hopefully decreasing the amount of such a therapeutic necessary for treatment of HIV latency.

Overall, the diverse world of CC chemokines offers the tools to be able to manipulate aspects of the immune systems vital for conditions from chronic inflammation to HIV entry and infection, making their thorough study and utilization vital when developing novel ways in order to combat these types of immune conditions. My research highlights several important structural insights into the CC chemokine system, as well as improved strategies and novel strategies to use these CC chemokines as tools to attempt and fight immunological conditions, opening the door to new possibilities for future biotherapeutics.

VII. References

(1) Abbas, A. K., Lichtman, A. H., and Pillai, S. (2012) Basic immunology: functions and disorders of the immune system, in *Basic Immunology: Functions and Disorders of the Immune System*, pp 303–305.

(2) Miedema, F., Tersmette, M., and van Lier, R. A. W. (1990) AIDS pathogenesis: a dynamic interaction between HIV and the immune system. *Immunol. Today 11*, 293–297.

(3) Cernak, I., and Noble-Haeusslein, L. J. (2010) Traumatic Brain Injury: An Overview of Pathobiology with Emphasis on Military Populations. *J. Cereb. Blood Flow Metab. 30*, 255–266.

(4) Middleton, J., Patterson, A. M., Gardner, L., Schmutz, C., and Ashton, B. A. (2002, January 1) Leukocyte extravasation: Chemokine transport and presentation by the endothelium. *Blood*. American Society of Hematology.

(5) Le, Y., Zhou, Y., Iribarren, P., and Wang, J. (2004) Chemokines and chemokine receptors: their manifold roles in homeostasis and disease. *Cell Mol Immunol 1*, 95–104.

(6) Springer, T. A. (1994) Traffic signals for lymphocyte recirculation and leukocyte emigration: The multistep paradigm. *Cell 76*, 301–314.

(7) Hamel, D. J., Sielaff, I., Proudfoot, A. E. I., and Handel, T. M. (2009, January 1) Chapter 4 Interactions of Chemokines with Glycosaminoglycans. *Methods Enzymol.* Academic Press.

(8) Onuffer, J. J., and Horuk, R. (2002) Chemokines, chemokine receptors and small-molecule antagonists: Recent developments. *Trends Pharmacol. Sci.*

(9) King, J. E., Eugenin, E. A., Buckner, C. M., and Berman, J. W. (2006, April 1) HIV tat and neurotoxicity. *Microbes Infect.* Elsevier Masson.

(10) Cathcart, M. K. (2009) Signal-activated phospholipase regulation of leukocyte chemotaxis. *J Lipid Res 50 Suppl*, S231-6.

(11) Oh-hora, M., and Rao, A. (2008, June) Calcium signaling in lymphocytes. *Curr. Opin. Immunol.* NIH Public Access.

(12) Xu, X., and Jin, T. (2015) The novel functions of the PLC/PKC/PKD signaling axis in G protein-coupled receptor-mediated chemotaxis of neutrophils. *J. Immunol. Res.* Hindawi.

(13) Tsai, F. C., Kuo, G. H., Chang, S. W., and Tsai, P. J. (2015, April 22) Ca2+ signaling in cytoskeletal reorganization, cell migration, and cancer metastasis. *Biomed Res. Int.* Hindawi.

(14) Nobes, C. D., and Hall, A. (1995) Rho, Rac, and Cdc42 GTPases regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia, and filopodia. *Cell 81*, 53–62.

(15) Yoshimura, T. (2015) Discovery of IL-8/CXCL8 (the story from Frederick). *Front. Immunol.* Frontiers Media SA.

(16) Bacon, K., Baggiolini, M., Broxmeyer, H., Horuk, R., Lindley, I., Mantovani, A., Matsushima, K., Murphy, P., Nomiyama, H., Oppenheim, J., Rot, A., Schall, T., Tsang, M., Thorpe, R., Van Damme, J., Wadhwa, M., Yoshie, O., Zlotnik, A., and Zoon, K. (2003) Chemokine/chemokine receptor nomenclature. *Cytokine 21*, 48–49.

(17) Mantovani, A. (1999, June) The chemokine system: Redundancy for robust outputs. *Immunol. Today*.

(18) White, G. E., Iqbal, A. J., and Greaves, D. R. (2013) CC chemokine receptors and chronic inflammation--therapeutic opportunities and pharmacological challenges. *Pharmacol. Rev. 65*, 47–89.

(19) Mantovani, A. (1999) Chemokines: Introduction and overview. *Chem Immunol. 72*, 1–6.

(20) Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem. 25*, 1605–1612.

(21) Wilen, C. B., Tilton, J. C., and Doms, R. W. (2012) HIV: Cell binding and entry. *Cold Spring Harb. Perspect. Med. 2*.

(22) Wang, X., Sharp, J. S., Handel, T. M., and Prestegard, J. H. (2013) Chemokine oligomerization in cell signaling and migration. *Prog. Mol. Biol. Transl. Sci. 117*, 531–578.

(23) Zheng, Y., Han, G. W., Abagyan, R., Wu, B., Stevens, R. C., Cherezov, V., Kufareva, I., and Handel, T. M. (2017) Structure of CC Chemokine Receptor 5 with a Potent Chemokine Antagonist Reveals Mechanisms of Chemokine Recognition and Molecular Mimicry by HIV. *Immunity 46*, 1005–1017.e5.

(24) Ludeman, J. P., and Stone, M. J. (2014, March) The structural role of receptor tyrosine sulfation in chemokine recognition. *Br. J. Pharmacol.* Wiley-Blackwell.

(25) Qin, L., Kufareva, I., Holden, L. G., Wang, C., Zheng, Y., Zhao, C., Fenalti, G., Wu, H., Han, G. W., Cherezov, V., Abagyan, R., Stevens, R. C., and Handel, T. M. (2015) Structural biology. Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. *Science (80-.). 347*, 1117–22.

(26) Crump, M. P. (1997) Solution structure and basis for functional activity of stromal cell-derived factor-1; dissociation of CXCR4 activation from binding and inhibition of HIV-1. *EMBO J. 16*, 6996–7007.

(27) Chevigné, A., Fievez, V., Schmit, J. C., and Deroo, S. (2011) Engineering and screening the N-terminus of chemokines for drug discovery, in *Biochemical Pharmacology*, pp 1438–1456. Elsevier.

(28) Zheng, Y., Han, G. W., Abagyan, R., Wu, B., Stevens, R. C., Cherezov, V., Kufareva, I., and Handel, T. M. (2017) Structure of CC Chemokine Receptor 5 with a Potent Chemokine Antagonist Reveals Mechanisms of Chemokine Recognition and Molecular Mimicry by HIV. *Immunity 46*, 1005–1017.

(29) Kawasaki, S., Takizawa, H., Yoneyama, H., Nakayama, T., Fujisawa, R., Izumizaki, M., Imai, T., Yoshie, O., Homma, I., Yamamoto, K., and Matsushima, K. (2001)

Intervention of thymus and activation-regulated chemokine attenuates the development of allergic airway inflammation and hyperresponsiveness in mice. *J. Immunol. 166*, 2055–62.

(30) Luster, A. D. (2001) Antichemokine immunotherapy for allergic diseases. *Curr Opin Allergy Clin Immunol 1*, 561–567.

(31) Brunner, P. M., Glitzner, E., Reininger, B., Klein, I., Stary, G., Mildner, M., Uhrin, P., Sibilia, M., and Stingl, G. (2015) CCL7 contributes to the TNF-alpha-dependent inflammation of lesional psoriatic skin. *Exp. Dermatol. 24*, 522–528.

(32) Girkin, J., Hatchwell, L., Foster, P., Johnston, S. L., Bartlett, N., Collison, A., and Mattes, J. (2015) CCL7 and IRF-7 Mediate Hallmark Inflammatory and IFN Responses following Rhinovirus 1B Infection. *J. Immunol. 194*, 4924–4930.

(33) Erickson, M. A., Morofuji, Y., Owen, J. B., and Banks, W. A. (2014) Rapid Transport of CCL11 across the Blood-Brain Barrier: Regional Variation and Importance of Blood Cells. *J. Pharmacol. Exp. Ther. 349*, 497–507.

(34) Wells, T. N., and Schwartz, T. W. (1997, December 1) Plagiarism of the host immune system: Lessons about chemokine immunology from viruses. *Curr. Opin. Biotechnol.* Elsevier Current Trends.

(35) Boomker, J. M., De Leij, L. F. M. H., The, T. H., and Harmsen, M. C. (2005) Viral chemokine-modulatory proteins: Tools and targets. *Cytokine Growth Factor Rev. 16*, 91–103.

(36) Lalani, A. S., Graham, K., Mossman, K., Rajarathnam, K., Clark-Lewis, I., Kelvin, D., and McFadden, G. (1997) The purified myxoma virus gamma interferon receptor homolog M-T7 interacts with the heparin-binding domains of chemokines. *J. Virol. 71*, 4356–63.

(37) Graham, K. A., Lalani, A. S., Macen, J. L., Ness, T. L., Barry, M., Liu, L.-Y., Lucas, A., Clark-Lewis, I., Moyer, R. W., and Mcfadden, G. (1997) The T1/35kDa Family of Poxvirus-Secreted Proteins Bind Chemokines and Modulate Leukocyte Influx into Virus-Infected Tissues. *Virology 229*, 12–24.

(38) Parry, C. M., Simas, J. P., Smith, V. P., Stewart, C. A., Minson, A. C., Efstathiou, S., and Alcami, A. (2000) A broad spectrum secreted chemokine binding protein encoded by a herpesvirus. *J. Exp. Med. 191*, 573–8.

(39) Carfí, A., Smith, C. A., Smolak, P. J., McGrew, J., and Wiley, D. C. (1999) Structure of a soluble secreted chemokine inhibitor vCCI (p35) from cowpox virus. *Proc. Natl. Acad. Sci. U. S. A. 96*, 12379–83.

(40) Burns, J. M., Dairaghi, D. J., Deitz, M., Tsang, M., and Schall, T. J. (2002) Comprehensive mapping of poxvirus vCCI chemokine-binding protein. Expanded range of ligand interactions and unusual dissociation kinetics. *J. Biol. Chem. 277*, 2785–9.

(41) Kuo, N.-W., Gao, Y.-G., Schill, M. S., Isern, N., Dupureur, C. M., and LiWang, P. J. (2014) Structural Insights into the Interaction between a Potent Anti-inflammatory Protein, Viral CC Chemokine Inhibitor (vCCI), and the Human CC Chemokine, Eotaxin-1. *J. Biol. Chem. 289*, 6592–6603.

(42) Gemma E. White, Eileen McNeill, Ivy Christou, Keith M. Channon, and D. R. G. (2011) Site-Directed Mutagenesis of the CC Chemokine Binding Protein 35K-Fc Reveals Residues Essential for Activity and Mutations That Increase the Potency of CC Chemokine Blockade. *Mol. Pharmacol. 80*, 328–36.

(43) Spinetti, G., Wang, M., Monticone, R., Zhang, J., Zhao, D., and Lakatta, E. G. (2004) Rat aortic MCP-1 and its receptor CCR2 increase with age and alter vascular smooth muscle cell function. *Arterioscler. Thromb. Vasc. Biol. 24*, 1397–1402.

(44) Beck, C. G., Studer, C., Zuber, J. F., Demange, B. J., Manning, U., and Urfer, R. (2001) The viral CC chemokine-binding protein vCCI inhibits monocyte chemoattractant protein-1 activity by masking its CCR2B-binding site. *J. Biol. Chem. 276*, 43270–6.

(45) Smith, C. a, Smith, T. D., Smolak, P. J., Friend, D., Hagen, H., Gerhart, M., Park, L., Pickup, D. J., Torrance, D., Mohler, K., Schooley, K., and Goodwin, R. G. (1997) Poxvirus genomes encode a secreted, soluble protein that preferentially inhibits beta chemokine activity yet lacks sequence homology to known chemokine receptors. *Virology 236*, 316–27.

(46) Zhang, L., DeRider, M., McCornack, M. A., Jao, S. -c., Isern, N., Ness, T., Moyer, R., and LiWang, P. J. (2006) Solution structure of the complex between poxvirus-encoded CC chemokine inhibitor vCCI and human MIP-1beta. *Proc. Natl. Acad. Sci. 103*, 13985–13990.

(47) White, G. E., Mcneill, E., Christou, I., Channon, K. M., and Greaves, D. R. (2011) Site-Directed Mutagenesis of the CC Chemokine Binding Protein 35K-Fc Reveals Residues Essential for Activity and Mutations That Increase the Potency of CC Chemokine Blockade. *Mol. Pharmacol. 80*, 328–336.

(48) Arnold, P. L., and Fremont, D. H. (2006) Structural Determinants of Chemokine Binding by an Ectromelia Virus-Encoded Decoy Receptor. *J. Virol. 80*, 7439–7449.

(49) Seet, B. T., Singh, R., Paavola, C., Lau, E. K., Handel, T. M., and McFadden, G. (2001) Molecular determinants for CC-chemokine recognition by a poxvirus CC-chemokine inhibitor. *Proc. Natl. Acad. Sci. U. S. A. 98*, 9008–13.

(50) Kledal, T. N., Rosenkilde, M. M., Coulin, F., Simmons, G., Johnsen, A. H., Alouani, S., Power, C. A., Luttichau, H. R., Gerstoft, J., Clapham, P. R., Clark-Lewis, I., Wells, T. N., and Schwartz, T. W. (1997) A broad-spectrum chemokine antagonist encoded by Kaposi's sarcoma-associated herpesvirus. *Science (80-.). 277*, 1656–1659.

(51) Sozzani, S., Luini, W., Bianchi, G., Allavena, P., Wells, T. N. C., Napolitano, M., Bernardini, G., Vecchi, A., D'Ambrosio, D., Mazzeo, D., Sinigaglia, F., Santoni, A., Maggi, E., Romagnani, S., and Mantovani, A. (1998) The Viral Chemokine Macrophage Inflammatory Protein-II Is a Selective Th2 Chemoattractant. *Blood 92*.

(52) Weber, K. S., Gröne, H. J., Röcken, M., Klier, C., Gu, S., Wank, R., Proudfoot, A. E., Nelson, P. J., and Weber, C. (2001) Selective recruitment of Th2-type cells and evasion from a cytotoxic immune response mediated by viral macrophage inhibitory protein-II. *Eur. J. Immunol. 31*, 2458–66.

(53) Weiss, R. A. (1993) How Does HIV Cause AIDS? *Science (80-.). 260*, 1273–1279.

(54) Chun, T.-W., Moir, S., and Fauci, A. S. (2015) HIV reservoirs as obstacles and

opportunities for an HIV cure. *Nat. Immunol. 16*, 584–589.

(55) Passaes, C. P., and Sáez-Cirión, A. (2014, April 1) HIV cure research: Advances and prospects. *Virology*. Academic Press.

(56) UNAIDS. (2016) Global AIDS Update 2016.

(57) Centers for Disease Control and Prevention (CDC). (2015) Diagnoses of HIV Infections in the United States and Dependent Areas, 2015. *HIV Surveill. Rep. 27*, 1–114.

(58) Larson, B. A., Brennan, A., McNamara, L., Long, L., Rosen, S., Sanne, I., and Fox, M. P. (2010) Early loss to follow up after enrolment in pre-ART care at a large public clinic in Johannesburg, South Africa. *Trop. Med. Int. Health 15 Suppl 1*, 43–7.

(59) Spinner, C. D., Boesecke, C., Zink, A., Jessen, H., Stellbrink, H. J., Rockstroh, J. K., and Esser, S. (2016, April 15) HIV pre-exposure prophylaxis (PrEP): a review of current knowledge of oral systemic HIV PrEP in humans. *Infection*. Springer Berlin Heidelberg.

(60) Kuritzkes, D. R. (2009) HIV-1 Entry Inhbitors: An Overview. *Curr Opin HIV AIDS 4*, 82–87.

(61) Klasse, P. J., Shattock, R. J., and Moore, J. P. (2006) Which topical microbicides for blocking HIV-1 transmission will work in the real world? *PLoS Med. 3*, 1501–1507.

(62) McGowan, I. (2006, December) Microbicides: A new frontier in HIV prevention. *Biologicals*.

(63) Gaertner, H., Cerini, F., Escola, J. M., Kuenzi, G., Melotti, A., Offord, R., Rossitto-Borlat, I., Nedellec, R., Salkowitz, J., Gorochov, G., Mosier, D., and Hartley, O. (2008) Highly potent, fully recombinant anti-HIV chemokines: reengineering a low-cost microbicide. *Proc Natl Acad Sci U S A 105*, 17706–17711.

(64) Zhu, P., Chertova, E., Bess, J., Lifson, J. D., Arthur, L. O., Liu, J., Taylor, K. A., and Roux, K. H. (2003) Electron tomography analysis of envelope glycoprotein trimers on HIV and simian immunodeficiency virus virions. *Proc. Natl. Acad. Sci. 100*, 15812–15817.

(65) Pancera, M., Zhou, T., Druz, A., Georgiev, I. S., Soto, C., Gorman, J., Huang, J., Acharya, P., Chuang, G.-Y., Ofek, G., Stewart-Jones, G. B. E., Stuckey, J., Bailer, R. T., Joyce, M. G., Louder, M. K., Tumba, N., Yang, Y., Zhang, B., Cohen, M. S., Haynes, B. F., Mascola, J. R., Morris, L., Munro, J. B., Blanchard, S. C., Mothes, W., Connors, M., and Kwong, P. D. (2014) Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature 514*, 455–461.

(66) Wilen, C. B., Tilton, J. C., and Doms, R. W. (2012) Molecular mechanisms of HIV entry. *Adv. Exp. Med. Biol. 726*, 223–242.

(67) Sattentau, Q., and Moore, J. (1993) The Role of CD4 in HIV Binding and Entry. *Philos. Trans. Biol. Sci. 342*, 59–66.

(68) Abbas, A. K., Lichtman, A. H., and Pillai, S. (2012) Basic immunology: functions and disorders of the immune system, in *Basic Immunology: Functions and Disorders of the Immune System*, pp 303–305.

(69) Dragic, T., Litwin, V., Allaway, G. P., Martin, S. R., Huang, Y., Nagashima, K. A., Cayanan, C., Maddon, P. J., Koup, R. A., Moore, J. P., and Paxton, W. A. (1996) HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature 381*, 667–673.

(70) Brumme, Z. L., Goodrich, J., Mayer, H. B., Brumme, C. J., Henrick, B. M., Wynhoven, B., Asselin, J. J., Cheung, P. K., Hogg, R. S., Montaner, J. S. G., and Harrigan, P. R. (2005) Molecular and clinical epidemiology of CXCR4-using HIV-1 in a large population of antiretroviral-naive individuals. *J. Infect. Dis. 192*, 466–474.

(71) Jones, P. L. S. J., Korte, T., and Blumenthal, R. (1998) Conformational changes in cell surface HIV-1 envelope glycoproteins are triggered by cooperation between cell surface CD4 and co-receptors. *J. Biol. Chem. 273*, 404–409.

(72) Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997) Core Structure of gp41 from the HIV Envelope Glycoprotein. *Cell 89*, 263–273.

(73) Clapham, P. R., Weber, J. N., Whitby, D., McIntosh, K., Dalgleish, A. G., Maddon, P. J., Deen, K. C., Sweet, R. W., and Weiss, R. A. (1989) Soluble CD4 blocks the infectivity of diverse strains of HIV and SIV for T cells and monocytes but not for brain and muscle cells. *Nature 337*, 368–70.

(74) Li, Z., Zhou, N., Sun, Y., Ray, N., Lataillade, M., Hanna, G. J., and Krystal, M. (2013) Activity of the HIV-1 attachment inhibitor BMS-626529, the active component of the prodrug BMS-663068, against CD4-independent viruses and HIV-1 envelopes resistant to other entry inhibitors. *Antimicrob. Agents Chemother. 57*, 4172–4180.

(75) Jacobson, J. M., Kuritzkes, D. R., Godofsky, E., DeJesus, E., Larson, J. A., Weinheimer, S. P., and Lewis, S. T. (2009) Safety, pharmacokinetics, and antiretroviral activity of multiple doses of ibalizumab (formerly TNX-355), an anti-CD4 monoclonal antibody, in human immunodeficiency virus type 1-infected adults. *Antimicrob. Agents Chemother. 53*, 450–457.

(76) Veazey, R. S., Ling, B., Green, L. C., Ribka, E. P., Lifson, J. D., Piatak, M., Lederman, M. M., Mosier, D., Offord, R., and Hartley, O. (2009) Topically Applied Recombinant Chemokine Analogues Fully Protect Macaques from Vaginal Simian-Human Immunodeficiency Virus Challenge.

(77) Palani A1, Shapiro S, Clader JW, Greenlee WJ, Cox K, Strizki J, Endres M, B. B. (2001) Discovery of 4-[(Z)-(4-bromophenyl)- (ethoxyimino)methyl]-1'-[(2,4-dimethyl-3-pyridinyl)carbonyl]-4'-methyl-1,4'- bipiperidine N-oxide (SCH 351125): an orally bioavailable human CCR5 antagonist for the treatment of HIV infection. *J Med Chem 44*, 3339–42.

(78) Ding, X., Zhang, X., Chong, H., Zhu, Y., Wei, H., Wu, X., He, J., Wang, X., and He, Y. (2017) Enfuvirtide (T20)-Based Lipopeptide Is a Potent HIV-1 Cell Fusion Inhibitor: Implications for Viral Entry and Inhibition. *J. Virol. 91*, JVI.00831-17.

(79) Forssmann, W. G., The, Y. H., Stoll, M., Adermann, K., Albrecht, U., Tillmann, H. C., Barlos, K., Busmann, A., Canales-Mayordomo, A., Gimenez-Gallego, G., Hirsch, J., Jimenez-Barbero, J., Meyer-Olson, D., Munch, J., Perez-Castells, J., Standker, L., Kirchhoff, F., and Schmidt, R. E. (2010) Short-term monotherapy in HIV-infected patients with a virus entry inhibitor against the gp41 fusion peptide. *Sci Transl Med 2*, 63re3.

(80) Alexandre, K. B., Gray, E. S., Mufhandu, H., McMahon, J. B., Chakauya, E., O'Keefe, B. R., Chikwamba, R., and Morris, L. (2012) The lectins griffithsin, cyanovirin-N and scytovirin inhibit HIV-1 binding to the DC-SIGN receptor and transfer to CD4(+) cells. *Virology 423*, 175–86.

(81) Cocchi, F., DeVico, A. L., Garzino-Demo, A., Arya, S. K., Gallo, R. C., and Lusso, P. (1995) Identification of RANTES, MIP-1α, and MIP-1β as the Major HIV-Suppressive Factors Produced by CD8+ T Cells. *Science (80-.). 270*, 1811–1815.

(82) Arenzana-Seisdedos, F., Virelizier, J.-L., Rousset, D., Clark-Lewis, I., Loetscher, P., Moser, B., and Baggiolini, M. (1996) HIV blocked by chemokine antagonist. *Nature 383*, 400–400.

(83) Laurence, J. S., LiWang, A. C., and LiWang, P. J. (1998) Effect of N-terminal truncation and solution conditions on chemokine dimer stability: Nuclear magnetic resonance structural analysis of macrophage inflammatory protein 1β mutants. *Biochemistry 37*, 9346–9354.

(84) Walli, R., Reinhart, B., Luckow, B., Lederer, E., Loch, O., Malo, A., Wank, R., Schlöndorff, D., and Goebel, F. D. (1998) HIV-1-infected long-term slow progressors heterozygous for delta32-CCR5 show significantly lower plasma viral load than wild-type slow progressors. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol. 18*, 229–233.

(85) Proudfoot, A. E. I., Power, C. A., Hoogewerf, A. J., Montjovent, M. O., Borlat, F., Offord, R. E., and Wells, T. N. C. (1996) Extension of recombinant human RANTES by the retention of the initiating methionine produces a potent antagonist. *J. Biol. Chem. 271*, 2599–2603.

(86) Suresh, P., and Wanchu, A. (2006) Chemokines and chemokine receptors in HIV infection: role in pathogenesis and therapeutics. *J Postgr. Med 52*, 210–217.

(87) Hartley, O., Gaertner, H., Wilken, J., Thompson, D., Fish, R., Ramos, A., Pastore, C., Dufour, B., Cerini, F., Melotti, A., Heveker, N., Picard, L., Alizon, M., Mosier, D., Kent, S., and Offord, R. (2004) Medicinal chemistry applied to a synthetic protein: development of highly potent HIV entry inhibitors. *Proc Natl Acad Sci U S A 101*, 16460–16465.

(88) Hartley, O., Dorgham, K., Perez-Bercoff, D., Cerini, F., Heimann, A., Gaertner, H., Offord, R. E., Pancino, G., Debré, P., and Gorochov, G. (2003) Human Immunodeficiency Virus Type 1 Entry Inhibitors Selected on Living Cells from a Library of Phage Chemokines. *J. Virol. 77*, 6637–6644.

(89) Nedellec, R., Coetzer, M., Lederman, M. M., Offord, R. E., Hartley, O., and Mosier, D. E. (2011) Resistance to the CCR5 inhibitor 5P12-RANTES requires a difficult evolution from CCR5 to CXCR4 Coreceptor use. *PLoS One 6*, e22020.

(90) Cuevas, J. M., Geller, R., Garijo, R., López-Aldeguer, J., and Sanjuán, R. (2015) Extremely High Mutation Rate of HIV-1 In Vivo. *PLoS Biol.* (Rowland-Jones, S. L., Ed.) *13*, e1002251.

(91) Wagner, T. A., and Frenkel, L. M. (2008) Potential limitation of CCR5 antagonists: drug resistance more often linked to CXCR4-utilizing than to CCR5-utilizing HIV-1. *AIDS 22*, 2393–5.

(92) Vicenzi, E., Liò, P., and Poli, G. (2013) The puzzling role of CXCR4 in human immunodeficiency virus infection. *Theranostics*.

(93) Zhao, B., Mankowski, M. K., Snyder, B. A., Ptak, R. G., and Liwang, P. J. (2011) Highly potent chimeric inhibitors targeting two steps of HIV cell entry. *J. Biol. Chem. 286*, 28370–81.

(94) Root, M. J. (2001) Protein Design of an HIV-1 Entry Inhibitor. *Science (80-.). 291*, 884–888.

(95) Eckert, D. M., and Kim, P. S. (2001) Mechanisms of Viral Membrane Fusion and Its Inhibition. *Annu. Rev. Biochem. 70*, 777–810.

(96) FDA. (2003) Highlights of Prescribing Information: Fuzeon (enfuvirtide) for Injection.

(97) Baldwin, C., and Berkhout, B. (2007) HIV-1 drug-resistance and drug-dependence. *Retrovirology 4*, 78.

(98) Otaka, A., and Nakamura, M. (2002) Remodeling of gp41-C34 Peptide Leads to Highly Effective Inhibitors of the Fusion of HIV-1 with Target Cells. *Angew. … 41*, 2937–2940.

(99) He, Y., Cheng, J., Lu, H., Li, J., Hu, J., Qi, Z., Liu, Z., Jiang, S., and Dai, Q. (2008) Potent HIV fusion inhibitors against Enfuvirtide-resistant HIV-1 strains. *Proc. Natl. Acad. Sci. U. S. A. 105*, 16332–7.

(100) Dwyer, J. J., Wilson, K. L., Davison, D. K., Freel, S. A., Seedorff, J. E., Wring, S. A., Tvermoes, N. A., Matthews, T. J., Greenberg, M. L., and Delmedico, M. K. (2007) Design of helical, oligomeric HIV-1 fusion inhibitor peptides with potent activity against enfuvirtide-resistant virus. *Proc. Natl. Acad. Sci. U. S. A. 104*, 12772–7.

(101) Qian, K., Morris-Natschke, S. L., and Lee, K.-H. (2009) HIV entry inhibitors and their potential in HIV therapy. *Med. Res. Rev. 29*, 369–93.

(102) Ryom, L., Lundgren, J. D., De Wit, S., Kovari, H., Reiss, P., Law, M., El-Sadr, W., Monforte, A. D., Mocroft, A., Smith, C., Fontas, E., Dabis, F., Phillips, A., and Sabin, C. (2016) Use of antiretroviral therapy and risk of end-stage liver disease and hepatocellular carcinoma in HIV-positive persons. *AIDS 30*, 1731–1743.

(103) Hamlyn, E., Ewings, F. M., Porter, K., Cooper, D. A., Tambussi, G., Schechter, M., Pedersen, C., Okulicz, J. F., McClure, M., Babiker, A., Weber, J., and Fidler, S. (2012) Plasma HIV Viral Rebound following Protocol-Indicated Cessation of ART Commenced in Primary and Chronic HIV Infection. *PLoS One* (Verhasselt, B., Ed.) *7*, e43754.

(104) Tarrago-Litvak, L., Andreola, M. L., Fournier, M., Nevinsky, G. A., Parissi, V., de Soultrait, V. R., and Litvak, S. (2002) Inhibitors of HIV-1 reverse transcriptase and integrase: classical and emerging therapeutical approaches. *Curr. Pharm. Des. 8*, 595–614.

(105) Hu, W.-S., and Hughes, S. H. (2012) HIV-1 reverse transcription. *Cold Spring Harb. Perspect. Med. 2*, a006882.

(106) Lusic, M., and Siliciano, R. F. (2016) Nuclear landscape of HIV-1 infection and integration. *Nat. Rev. Microbiol. 15*, 69–82.

(107) Craigie, R., and Bushman, F. D. (2012) HIV DNA integration. *Cold Spring Harb. Perspect. Med. 2*, a006890.

(108) Archin, N. M., Sung, J. M., Garrido, C., Soriano-Sarabia, N., and Margolis, D. M. (2014) Eradicating HIV-1 infection: seeking to clear a persistent pathogen. *Nat. Rev. Microbiol. 12*, 750–764.

(109) Siliciano, J. D., Kajdas, J., Finzi, D., Quinn, T. C., Chadwick, K., Margolick, J. B., Kovacs, C., Gange, S. J., and Siliciano, R. F. (2003) Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nat. Med. 9*, 727–728.

(110) Kandathil, A. J., Sugawara, S., and Balagopal, A. (2016) Are T cells the only HIV-1 reservoir? *Retrovirology 13*, 86.

(111) Deeks, S. G. (2012) HIV: Shock and kill. *Nature 487*, 439–440.

(112) Limsirichai, P., Gaj, T., and Schaffer, D. V. (2016) CRISPR-mediated Activation of Latent HIV-1 Expression. *Mol. Ther. 24*, 499–507.

(113) Pegu, A., Asokan, M., Wu, L., Wang, K., Hataye, J., Casazza, J. P., Guo, X., Shi, W., Georgiev, I., Zhou, T., Chen, X., O'Dell, S., Todd, J.-P., Kwong, P. D., Rao, S. S., Yang, Z., Koup, R. A., Mascola, J. R., and Nabel, G. J. (2015) Activation and lysis of human CD4 cells latently infected with HIV-1. *Nat. Commun. 6*, 8447.

(114) Van Praag, R. M. E., Prins, J. M., Roos, M. T. L., Schellekens, P. T. A., Ten Berge, I. J. M., Yong, S. L., Schuitemaker, H., Eerenberg, A. J. M., Jurriaans, S., De Wolf, F., Fox, C. H., Goudsmit, J., Miedema, F., and Lange, J. M. A. (2001) OKT3 and IL-2 treatment for purging of the latent HIV-1 reservoir in vivo results in selective long-lasting CD4+ T cell depletion. *J. Clin. Immunol. 21*, 218–226.

(115) Schwartz, C., Bouchat, S., Marban, C., Gautier, V., Van Lint, C., Rohr, O., and Le Douce, V. (2017, July 4) On the way to find a cure: Purging latent HIV-1 reservoirs. *Biochem. Pharmacol.* Elsevier.

(116) Griffith, J. W., Sokol, C. L., and Luster, A. D. (2014) Chemokines and Chemokine Receptors: Positioning Cells for Host Defense and Immunity. *Annu. Rev. Immunol. 32*, 659–702.

(117) Nguyen, A. F., Kuo, N. W., Showalter, L. J., Ramos, R., Dupureur, C. M., Colvin, M. E., and LiWang, P. J. (2017) Biophysical and computational studies of the vCCI:VMIP-II complex. *Int. J. Mol. Sci. 18*, 1778.

(118) Heidarieh, H., Hernáez, B., and Alcamí, A. (2015) Immune modulation by virus-encoded secreted chemokine binding proteins. *Virus Res. 209*, 67–75.

(119) Alexander, J. M., Nelson, C. A., Van Berkel, V., Lau, E. K., Studts, J. M., Brett, T. J., Speck, S. H., Handel, T. M., and Virgin, H. W. (2002) Structural Basis of Chemokine Sequestration by a Herpesvirus Decoy Receptor. *Cell 111*, 343–356.

(120) Webb, L. M. C., Smith, V. P., and Alcami, A. (2004) The gammaherpesvirus chemokine binding protein can inhibit the interaction of chemokines with glycosaminoglycans. *FASEB J. 18*, 571–3.

(121) Alexander-Brett, J. M., and Fremont, D. H. (2007) Dual GPCR and GAG mimicry by the M3 chemokine decoy receptor. *J. Exp. Med. 204*, 3157–72.

(122) Alejo, A., Ruiz-Argüello, M. B., Ho, Y., Smith, V. P., Saraiva, M., and Alcami, A. (2006) A chemokine-binding domain in the tumor necrosis factor receptor from variola (smallpox) virus. *Proc. Natl. Acad. Sci. U. S. A. 103*, 5995–6000.

(123) Carfí, A., Smith, C. A., Smolak, P. J., McGrew, J., and Wiley, D. C. (1999) Structure of a soluble secreted chemokine inhibitor vCCI (p35) from cowpox virus. *Proc. Natl. Acad. Sci. U. S. A. 96*, 12379–83.

(124) Liwang, A. C., Wang, Z. X., Sun, Y., Peiper, S. C., and Liwang, P. J. (1999) The solution structure of the anti-HIV chemokine vMIP-II. *Protein Sci. 8*, 2270–80.

(125) Zhao, B., and Liwang, P. J. (2010) Characterization of the interactions of vMIP-II, and a dimeric variant of vMIP-II, with glycosaminoglycans. *Biochemistry 49*, 7012–7022.

(126) Takami, S., Minami, M., Nagata, I., Namura, S., and Satoh, M. (2001) Chemokine Receptor Antagonist Peptide, Viral MIP-II, Protects the Brain Against Focal Cerebral Ischemia in Mice. *J. Cereb. Blood Flow Metab.* 1430–1435.

(127) Ghirnikar, R. S., Lee, Y. L., and Eng, L. F. (2001) Chemokine antagonist infusion promotes axonal sparing after spinal cord contusion injury in rat. *J. Neurosci. Res. 64*, 582–589.

(128) Bedke, J., Stojanovic, T., Kiss, E., Behnes, C.-L., Proudfoot, A. E., and Gröne, H.-J. (2010) Viral macrophage inflammatory protein-II improves acute rejection in allogeneic rat kidney transplants. *World J. Urol. 28*, 537–42.

(129) Zhang, L. (2008) Structural Study of the Interaction Between Poxvirus-Encoded CC Chemokine Inhibitor VCCI and Hhuman MIP-1β: A Dissertation.

(130) Oswald, N. (2008) What you need to know about OD600. BitesizeBio.

(131) Studier, F. W., and Moffatt, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol. 189*, 113–130.

(132) Rosano, G. L., and Ceccarelli, E. A. (2014) Recombinant protein expression in Escherichia coli: Advances and challenges. *Front. Microbiol.* Frontiers Media SA.

(133) A. Qasim, M., and Taha, M. (2012) Investigation of the Mechanism of Protein Denaturation by Guanidine Hydrochloride-Induced Dissociation of Inhibitor-Protease Complexes. *Protein Pept. Lett. 20*, 187–191.

(134) Wingfield, P. T. (2015) Overview of the purification of recombinant proteins. *Curr. Protoc. Protein Sci. 80*, 6.1.1-6.1.35.

(135) Holmgren, A. (1985) Thioredoxin. *Annu. Rev. Biochem. 54*, 237–271.

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

(137) Sela, M., and Lifson, S. (1959) On the reformation of disulfide bridges in proteins. *Biochim. Biophys. Acta 36*, 471–478.

(138) Life technologies. (2015) Ni-NTA Purification System.

(139) Edgcomb, S. P., and Murphy, K. P. (2002) Variability in the pKa of histidine side-chains correlates with burial within proteins. *Proteins Struct. Funct. Genet. 49*, 1–6.

(140) Simonian, M. H., Simonian, and H., M. (2002) Spectrophotometric Determination of Protein Concentration. *Curr. Protoc. Food Anal. Chem.* B.1.3.1-B.1.3.7.

(141) Skala, W., Goettig, P., and Brandstetter, H. (2013) Do-it-yourself histidine-tagged bovine enterokinase: A handy member of the protein engineer's toolbox. *J. Biotechnol. 168*, 421–425.

(142) Duong-Ly, K. C., and Gabelli, S. B. (2014) Using ion exchange chromatography to purify a recombinantly expressed protein. *Methods Enzymol. 541*, 95–103.

(143) Molnar, I., and Horvath, C. (1976) Reverse phase chromatography of polar biological substances: separation of catechol compounds by high performance liquid chromatography. *Clin. Chem. 22*, 1497–1502.

(144) Chang, Y.-G., Kuo, N.-W., Tseng, R., and LiWang, A. (2011) Flexibility of the C-terminal, or CII, ring of KaiC governs the rhythm of the circadian clock of cyanobacteria. *Proc. Natl. Acad. Sci. U. S. A. 108*, 14431–6.

(145) Sanders, J. K. M., and Hunter, B. K. (1987) Modern NMR spectroscopy : a guide for chemists. Oxford University Press.

(146) Rith, K., and Schäfer, A. (1999) The Mystery of Nucleon Spin. *Sci. Am. 281*, 58–63.

(147) Cavanagh, J. (2007) Protein NMR spectroscopy : principles and practice. Academic Press.

(148) Keeler, J. (2002) NMR and energy levels, in *Understanding NMR Spectroscopy*.

(149) Bloch, F. (1946) Nuclear induction. *Phys. Rev. 70*, 460–474.

(150) Levitt, M. H. (1997) The signs of frequencies and phases in NMR. *J. Magn. Reson. 126*, 164–182.

(151) Hoult, D. I., and Ginsberg, N. S. (2001) The Quantum Origins of the Free Induction Decay Signal and Spin Noise. *J. Magn. Reson. 148*, 182–199.

(152) Wüthrich, K. (2001) The way to NMR structures of proteins. *Nat. Struct. Biol. 8*, 923–925.

(153) Keeler, J. (2007) Two-dimensional NMR, in *Understanding NMR Spectroscopy*.

(154) Bodenhausen, G., and Ruben, D. J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett. 69*, 185–189.

(155) Morris, G. A., and Freeman, R. (1979) Enhancement of nuclear magnetic resonance signals by polarization transfer. *J. Am. Chem. Soc. 101*, 760–762.

(156) Hahn, E. L., and Maxwell, D. E. (1952) Spin echo measurements of nuclear spin coupling in molecules. *Phys. Rev. 88*, 1070–1084.

(157) Konrat, R. (2014) NMR contributions to structural dynamics studies of intrinsically disordered proteins. *J. Magn. Reson. 241*, 74–85.

(158) Goodchild, S. C., Curmi, P. M. G., and Brown, L. J. (2011, September) Structural gymnastics of multifunctional metamorphic proteins. *Biophys. Rev.* Springer.

(159) Higman, V. (2012) 1H-15N HSQC, in *Protein NMR, A Practical Guide*.

(160) Zhao, B. (2011) Structural and Functional Studies of the Receptor-Binding and Glycosoaminoglycan-Binding Mechanism of a Viral Chemokine Analog vMIP-II and Ration Design of Chemokine-Based Highly Potent HIV-1 Entry Inhibitors: A Dissertation.

(161) Foster, M. P., McElroy, C. A., and Amero, C. D. (2007) Solution NMR of large molecules and assemblies. *Biochemistry 46*, 331–340.

(162) Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) 1 H, 13 C and 15 N chemical shift referencing in biomolecular NMR. *J.Biomolec.NMR 6*, 135–140.

(163) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6*, 277–293.

(164) Garrett, D. S., Powers, R., Gronenborn, A. M., and Clore, G. M. (1991) A common sense approach to peak picking in two-, three-, and four-dimensional spectra using automatic computer analysis of contour diagrams. *J. Magn. Reson. 95*, 214–220.

(165) Csermely, P., Palotai, R., and Nussinov, R. (2010) Induced fit, conformational selection and independent dynamic segments: An extended view of binding events. *Trends Biochem. Sci. 35*, 539–546.

(166) Williamson, M. P. (2013) Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.*

(167) Farmer, B. T. I., Constantine, K. L., Goldfarb, V., Friedrichs, M. S., Wittekind, M., Yanchunas J., J., Robertson, J. G., and Mueller, L. (1996) Localizing the NADP+ binding site on the murb enzyme by NMR. *Nat. Struct. Biol. 3*, 995–997.

(168) Garrett, D. S., Seok, Y. J., Peterkofsky, A., Clore, G. M., and Gronenborn, A. M. (1997) Identification by NMR of the binding surface for the histidine- containing phosphocarrier protein HPr on the N-terminal domain of enzyme I of the Escherichia coli phosphotransferase system. *Biochemistry 36*, 4393–4398.

(169) Damian, L. (2013) Isothermal Titration Calorimetry for Studying Protein–Ligand Interactions, in *Protein-Ligand Interactions, Methods and Applications*, pp 103–118.

(170) Pierce, M. M., Raman, C. S., and Nall, B. T. (1999) Isothermal Titration Calorimetry of Protein–Protein Interactions. *Methods 19*, 213–221.

(171) Keserü, G., Swinney, D. C., Mannhold, R., Kubinyi, H., and Folkers, G. (2015) Thermodynamics and Kinetics of Drug Binding.

(172) TA Istruments. (2014) Nano ITC: Getting Started, Revision F.

(173) Johnson, M. L., Holt, J. M., and Ackers, G. K. (2009) Thermodynamics of 2-Cys

Peroxiredoxin Assembly Determined by ITC, in *Biothermodynamics, Part 2*, pp 418–422. Academic Press/Elsevier.

(174) Wiseman, T., Williston, S., Brandts, J. F., and Lin, L. N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem. 179*, 131–7.

(175) Doyle, M. L., and Hensley, P. (1998) Tight ligand binding affinities determined from thermodynamic linkage to temperature by titration calorimetry. *Methods Enzymol. 295*, 88–99.

(176) Albrecht, A. (1961) Polarizations and assignments of transitions: The method of photoselection. *J. Mol. Spectrosc. 6*, 84–108.

(177) Pollard, T. D. (2010) A Guide to Simple and Informative Binding Assays. *Mol. Biol. Cell 21*, 4061–4067.

(178) Marchand, J. B., Kaiser, D. A., Pollard, T. D., and Higgs, H. N. (2001) Interaction of WASP/Scar proteins with actin and vertebrate Arp2/3 complex. *Nat. Cell Biol. 3*, 76–82.

(179) Conlan, L. H., and Dupureur, C. M. (2002) Dissecting the metal ion dependence of DNA binding by PvuII endonuclease. *Biochemistry 41*, 1335–1342.

(180) Reid, S. L., Parry, D., Liu, H. H., and Connolly, B. A. (2001) Binding and recognition of GATATC target sequences by the EcoRV restriction endonuclease: A study using fluorescent oligonucleotides and fluorescence polarization. *Biochemistry 40*, 2484–2494.

(181) Abraham, M. J., Murtola, T., Schulz, R., P??ll, S., Smith, J. C., Hess, B., and Lindah, E. (2015) Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX 1–2*, 19–25.

(182) Pronk, S., Pall, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., Van Der Spoel, D., Hess, B., and Lindahl, E. (2013) GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics 29*, 845–854.

(183) Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: algorithms for highly efficient, load balanced, and scalable molecular simulations. *J. Chem. Theory Comput. 4*, 435–447.

(184) Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997) LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem. 18*, 1463–1472.

(185) Bussi, G., Donadio, D., and Parrinello, M. (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys. 126*, 1–7.

(186) Parrinello, M., and Rahman, A. (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys. 52*, 7182–7190.

(187) Lindorff-Larsen, K., Piana, S., Palmo, K., Maragakis, P., Klepeis, J. L., Dror, R. O., and Shaw, D. E. (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct. Funct. Bioinforma. 78*, 1950–1958.

(188) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79*, 926.

(189) Zhang, L., and Liwang, P. J. (2002) Letter to the Editor Resonance assignments of the 34 kD rabbitpox vCCI:human MIP-1b complex. *J. Leukoc. Biol. J. Biomol. NMR 72*, 24–3449.

(190) Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*, 2577–2637.

(191) PDBePISA (Proteins, Interfaces, Structures and Assemblies).

(192) Krissinel, E., and Henrick, K. (2007) Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol. 372*, 774–797.

(193) Krissinel, E. (2009) Valence Bond Theory for Chemical Dynamics. *J. Comput. Chem. 28*, 73–86.

(194) Taylor, P. C., and Feldmann, M. (2009) Anti-TNF biologic agents: still the therapy of choice for rheumatoid arthritis. *Nat. Rev. Rheumatol. 5*, 578–582.

(195) Yi, X., Manickam, D. S., Brynskikh, A., and Kabanov, A. V. (2014) Agile delivery of protein therapeutics to CNS. *J. Control. Release 190*, 637–663.

(196) Boomker, J. M., De Leij, L. F. M. H., The, T. H., and Harmsen, M. C. (2005) Viral chemokine-modulatory proteins: Tools and targets. *Cytokine Growth Factor Rev. 16*, 91–103.

(197) Alcami, A., and Viejo-Borbolla, A. (2009) Chapter 8 Identification and Characterization of Virus-Encoded Chemokine Binding Proteins, pp 173–191.

(198) Adcock, S. a, and Mccammon, J. A. (2006) Molecular Dynamics : Survey of Methods for Simulating the Activity of Proteins Molecular Dynamics : Survey of Methods for Simulating the Activity of Proteins. *Chem. Rev. 106*, 1589–1615.

(199) Shaw, D. E., Maragakis, P., Lindorff-larsen, K., Piana, S., Shan, Y., and Wriggers, W. (2010) Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science (80-.). 330*, 341–347.

(200) Dill, K. A., and MacCallum, J. L. (2012) The Protein-Folding Problem, 50 Years On. *Science (80-.). 338*, 1042–1046.

(201) Nguyen, A. F., Schill, M. S., Jian, M., and LiWang, P. J. (2017) The Effect of N-Terminal Cyclization on the Function of the HIV Entry Inhibitor 5P12-RANTES. *Int. J. Mol. Sci. 18*, 1575.

(202) Unaids. (2013) UNAIDS report on the global AIDS epidemic 2013.

(203) Boskey, E. R., Telsch, K. M., Whaley, K. J., Moench, T. R., and Cone, R. A. (1999) Acid production by vaginal flora in vitro is consistent with the rate and extent of vaginal acidification. *Infect. Immun. 67*, 5170–5175.

(204) Evans, D. F., Pye, G., Bramley, R., Clark, A. G., Dyson, T. J., and Hardcastle, J.

D. (1988) Measurement of gastrointestinal pH profiles in normal ambulant human subjects. *Gut 29*, 1035–1041.

(205) Wilton, J. (2012) Putting a number on it: The risk from an exposure to HIV. *CATIE*.

(206) Suresh, P., and Wanchu, A. (2006) Chemokines and chemokine receptors in HIV infection: role in pathogenesis and therapeutics. *J. Postgrad. Med. 52*, 210–7.

(207) Dragic, T., Litwin, V., Allaway, G. P., Martin, S. R., Huang, Y., Nagashima, K. A., Cayanan, C., Maddon, P. J., Koup, R. A., Moore, J. P., and Paxton, W. A. (1996) HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. *Nature 381*, 667–73.

(208) Cerini, F., Landay, A., Gichinga, C., Lederman, M. M., Flyckt, R., Starks, D., Offord, R. E., Xois, F., Gal, L., and Hartley, O. (2008) Chemokine Analogues Show Suitable Stability for Development as Microbicides. *J Acquir Immune Defic Syndr 49*, 472–476.

(209) Wiktor, M., Hartley, O., and Grzesiek, S. (2013) Characterization of structure, dynamics, and detergent interactions of the anti-HIV chemokine variant 5P12-RANTES. *Biophys. J. 105*, 2586–2597.

(210) Laurence, J. S., Blanpain, C., De Leener, A., Parmentier, M., and LiWang, P. J. (2001) Importance of basic residues and quaternary structure in the function of MIP-1β: CCR5 binding and cell surface sugar interactions. *Biochemistry 40*, 4990–4999.

(211) Arenzana-Seisdedos, F., Virelizier, J.-L., Rousset, D., Clark-Lewis, I., Loetscher, P., Moser, B., and Baggiolini, M. (1996) HIV blocked by chemokine antagonist. *Nature 383*, 400–400.

(212) Simmons, G., Clapham, P. R., Picard, L., Offord, R. E., Rosenkilde, M. M., Schwartz, T. W., Buser, R., Wells, T. N., Proudfoot, A. E., Cocchi, F., Paxton, W. A., Dragic, T., Alkhatib, G., Deng, H., Arenzana-Seisdedos, F., Moriuchi, H., Schmidtmayerova, H., Sherry, B., Bukrinsky, M., Lusti-Narasimhan, M., Simmons, G., Feng, Y., Broder, C. C., Kennedy, P. E., Berger, E. A., McKnight, Á., Simmons, G., Clapham, P. R., Blanc, D., Weiss, R. A., Clapham, P. R., McKnight, Á., Weiss, R. A., Dragic, T., Charneau, P., Clavel, F., Alizon, M., Endres, M. J., and McKnight, Á. (1997) Potent inhibition of HIV-1 infectivity in macrophages and lymphocytes by a novel CCR5 antagonist. *Science 276*, 276–9.

(213) White, G. E., Iqbal, A. J., and Greaves, D. R. (2013) CC chemokine receptors and chronic inflammation--therapeutic opportunities and pharmacological challenges. *Pharmacol. Rev. 65*, 47–89.

(214) Pastore, C., Picchio, G. R., Galimi, F., Fish, R., Hartley, O., Offord, R. E., and Mosier, D. E. (2003) Two mechanisms for human immunodeficiency virus type 1 inhibition by N-terminal modifications of RANTES. *Antimicrob. Agents Chemother. 47*, 509–17.

(215) Yu, L., Vizel, A., Huff, M. B., Young, M., Remmele, R. L., and He, B. (2006) Investigation of N-terminal glutamate cyclization of recombinant monoclonal antibody in formulation development. *J. Pharm. Biomed. Anal. 42*, 455–463.

(216) Dick, L. W., Kim, C., Qiu, D., and Cheng, K.-C. (2007) Determination of the origin

of the N-terminal pyro-glutamate variation in monoclonal antibodies using model peptides. *Biotechnol. Bioeng. 97*, 544–553.

(217) Kumar, M., Chatterjee, A., Khedkar, A. P., Kusumanchi, M., and Adhikary, L. (2013) Mass spectrometric distinction of in-source and in-solution pyroglutamate and succinimide in proteins: A case study on rhG-CSF. *J. Am. Soc. Mass Spectrom. 24*, 202–212.

(218) Liu, Y. D., Goetze, A. M., Bass, R. B., and Flynn, G. C. (2011) N-terminal glutamate to pyroglutamate conversion in vivo for human IgG2 antibodies. *J. Biol. Chem. 286*, 11211–11217.

(219) Tritsch, G. L., and Moore, G. E. (1962) Spontaneous decomposition of glutamine in cell culture media. *Exp. Cell Res. 28*, 360–364.

(220) Chelius, D., Jing, K., Lueras, A., Rehder, D. S., Dillon, T. M., Vizel, A., Rajan, R. S., Li, T., Treuheit, M. J., and Bondarenko, P. V. (2006) Formation of pyroglutamic acid from N-terminal glutamic acid in immunoglobulin gamma antibodies. *Anal. Chem. 78*, 2370–2376.

(221) Yan, B., Valliere-Douglass, J., Brady, L., Steen, S., Han, M., Pace, D., Elliott, S., Yates, Z., Han, Y., Balland, A., Wang, W., and Pettit, D. (2007) Analysis of post-translational modifications in recombinant monoclonal antibody IgG1 by reversed-phase liquid chromatography/mass spectrometry. *J. Chromatogr. A 1164*, 153–161.

(222) Kumar, A., and Bachhawat, A. K. (2012) Pyroglutamic acid: Throwing light on a lightly studied metabolite. *Curr. Sci. 102*, 288–297.

(223) Cerini, F., Gaertner, H., Madden, K., Tolstorukov, I., Brown, S., Laukens, B., Callewaert, N., Harner, J. C., Oommen, A. M., Harms, J. T., Sump, A. R., Sealock, R. C., Peterson, D. J., Johnson, S. K., Abramson, S. B., Meagher, M., Offord, R., and Hartley, O. (2016) A scalable low-cost cGMP process for clinical grade production of the HIV inhibitor 5P12-RANTES in Pichia pastoris. *Protein Expr. Purif. 119*, 1–10.

(224) Wiktor, M. (2013) Towards NMR analysis of the HIV-1 coreceptor CCR5 and its interaction with RANTES. *Diss. Univ. Basel* 1–171.

(225) Platt, E. J., Bilska, M., Kozak, S. L., Kabat, D., and Montefiori, D. C. (2009) Evidence that Ecotropic Murine Leukemia Virus Contamination in TZM-bl Cells Does Not Affect the Outcome of Neutralizing Antibody Assays with Human Immunodeficiency Virus Type 1. *J. Virol. 83*, 8289–8292.

(226) Takeuchi, Y., McClure, M. O., and Pizzato, M. (2008) Identification of Gammaretroviruses Constitutively Released from Cell Lines Used for Human Immunodeficiency Virus Research. *J. Virol. 82*, 12585–12588.

(227) Wei, X., Decker, J. M., Liu, H., Zhang, Z., Arani, R. B., Kilby, J. M., Saag, M. S., Wu, X., Shaw, G. M., and Kappes, J. C. (2002) Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. *Antimicrob. Agents Chemother. 46*, 1896–1905.

(228) Derdeyn, C. A., Decker, J. M., Sfakianos, J. N., Wu, X., O'Brien, W. A., Ratner, L., Kappes, J. C., Shaw, G. M., and Hunter, E. (2000) Sensitivity of human immunodeficiency virus type 1 to the fusion inhibitor T-20 is modulated by coreceptor

specificity defined by the V3 loop of gp120. *J. Virol. 74*, 8358–67.

(229) Platt, E. J., Wehrly, K., Kuhmann, S. E., Chesebro, B., and Kabat, D. (1998) Effects of CCR5 and CD4 Cell Surface Concentrations on Infections by Macrophagetropic Isolates of Human Immunodeficiency Virus Type 1. *J. Virol. 72*, 2855–2864.

(230) Laurence, J. S., Blanpain, C., Burgner, J. W., Parmentier, M., and LiWang, P. J. (2000) CC chemokine MIP-1β can function as a monomer and depends on Phe13 for receptor binding. *Biochemistry 39*, 3401–3409.

(231) Li, M., Gao, F., Mascola, J. R., Stamatatos, L., Polonis, V. R., Koutsoukos, M., Voss, G., Goepfert, P., Gilbert, P., Greene, K. M., Bilska, M., Kothe, D. L., Salazar-gonzalez, J. F., Wei, X., Decker, J. M., Hahn, B. H., and Montefiori, D. C. (2005) Human Immunodeficiency Virus Type 1 env Clones from Acute and Early Subtype B Infections for Standardized Assessments of Vaccine-Elicited Neutralizing Antibodies. *J. Virol. 79*, 10108–10125.

(232) Derdeyn, C. a, Decker, J. M., Bibollet-Ruche, F., Mokili, J. L., Muldoon, M., Denham, S. a, Heil, M. L., Kasolo, F., Musonda, R., Hahn, B. H., Shaw, G. M., Korber, B. T., Allen, S., and Hunter, E. (2004) Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science 303*, 2019–22.

(233) Wei, X., Decker, J. M., Liu, H., Zhang, Z., Arani, R. B., Kilby, J. M., Saag, M. S., Shaw, G. M., Kappes, J. C., and Wu, X. (2002) Emergence of Resistant Human Immunodeficiency Virus Type 1 in Patients Receiving Fusion Inhibitor (T-20) Monotherapy Emergence of Resistant Human Immunodeficiency Virus Type 1 in Patients Receiving Fusion Inhibitor (T-20) Monotherapy. *Antimicrob. Agents Chemother. 46*, 1896–1905.

(234) Wei, X., Decker, J. M., Wang, S., Hui, H., Kappes, J. C., Wu, X., Salazar-Gonzalez, J. F., Salazar, M. G., Kilby, J. M., Saag, M. S., Komarova, N. L., Nowak, M. A., Hahn, B. H., Kwong, P. D., and Shaw, G. M. (2003) Antibody neutralization and escape by HIV-1. *Nature 422*, 307–312.

(235) Jin, H., Shen, X., Baggett, B. R., Kong, X., and LiWang, P. J. (2007) The human CC chemokine MIP-1beta dimer is not competent to bind to the CCR5 receptor. *J. Biol. Chem. 282*, 27976–83.

(236) Shih, W., and Yamada, S. (2012) N-cadherin-mediated cell-cell adhesion promotes cell migration in a three-dimensional matrix. *J. Cell Sci. 125*, 3661–3670.

(237) Grewe, C., Beck, A., and Gelderblom, H. R. (1990) HIV: early virus-cell interactions. *J. Acquir. Immune Defic. Syndr. 3*, 965–974.

(238) Allen, S. J., Crown, S. E., and Handel, T. M. (2007) Chemokine:Receptor Structure, Interactions, and Antagonism. *Annu. Rev. Immunol. 25*, 787–820.

(239) Appay, V., Brown, A., Cribbes, S., Randle, E., and Czaplewski, L. G. (1999) Aggregation of RANTES Is Responsible for Its Inflammatory Properties: Characterization of Nonaggregating, Noninflammatory RANTES Mutants. *J. Biol. Chem. 274*, 27505–27512.

(240) Burton, D. R., and Hangartner, L. (2016) Broadly Neutralizing Antibodies to HIV

and Their Role in Vaccine Design. *Annu. Rev. Immunol. 34*, 635–659.

(241) Klein, F., Halper-Stromberg, A., Horwitz, J. A., Gruell, H., Scheid, J. F., Bournazos, S., Mouquet, H., Spatz, L. A., Diskin, R., Abadir, A., Zang, T., Dorner, M., Billerbeck, E., Labitt, R. N., Gaebler, C., Marcovecchio, P. M., Incesu, R.-B., Eisenreich, T. R., Bieniasz, P. D., Seaman, M. S., Bjorkman, P. J., Ravetch, J. V., Ploss, A., and Nussenzweig, M. C. (2012) HIV therapy by a combination of broadly neutralizing antibodies in humanized mice. *Nature 492*, 118–122.

(242) Clark, A. J., Gindin, T., Zhang, B., Wang, L., Abel, R., Murret, C. S., Xu, F., Bao, A., Lu, N. J., Zhou, T., Kwong, P. D., Shapiro, L., Honig, B., and Friesner, R. A. (2017) Free Energy Perturbation Calculation of Relative Binding Free Energy between Broadly Neutralizing Antibodies and the gp120 Glycoprotein of HIV-1. *J. Mol. Biol. 429*, 930–947.

(243) Pancera, M., Shahzad-ul-Hussan, S., Doria-Rose, N. A., McLellan, J. S., Bailer, R. T., Dai, K., Loesgen, S., Louder, M. K., Staupe, R. P., Yang, Y., Zhang, B., Parks, R., Eudailey, J., Lloyd, K. E., Blinn, J., Munir Alam, S., Haynes, B. F., Amin, M. N., Wang, L.-X., Burton, D. R., Koff, W. C., Nabel, G. J., Mascola, J. R., Bewley, C. A., and Kwong, P. D. (2013) Structural basis for diverse N-glycan recognition by HIV-1–neutralizing V1–V2–directed antibody PG16 *8*.

(244) Alexandre, K. B., Gray, E. S., Lambson, B. E., Moore, P. L., Choge, I. A., Mlisana, K., Karim, S. S. A., McMahon, J., O'Keefe, B., Chikwamba, R., and Morris, L. (2010) Mannose-rich glycosylation patterns on HIV-1 subtype C gp120 and sensitivity to the lectins, Griffithsin, Cyanovirin-N and Scytovirin. *Virology 402*, 187–196.

(245) Akkouh, O., Ng, T. B., Singh, S. S., Yin, C., Dan, X., Chan, Y. S., Pan, W., and Cheung, R. C. F. (2015) Lectins with anti-HIV activity: A review. *Molecules 20*, 648–668.

(246) Pang, W., Tam, S.-C., and Zheng, Y.-T. (2009) Current peptide HIV type-1 fusion inhibitors. *Antivir. Chem. Chemother. 20*, 1–18.

(247) Kagiampakis, I., Gharibi, A., Mankowski, M. K., Snyder, B. A., Ptak, R. G., Alatas, K., and LiWang, P. J. (2011) Potent Strategy To Inhibit HIV-1 by Binding both gp120 and gp41. *Antimicrob. Agents Chemother. 55*, 264–275.

(248) Qin, L., Kufareva, I., Holden, L. G., Wang, C., Zheng, Y., Zhao, C., Fenalti, G., Wu, H., Han, G. W., Cherezov, V., Abagyan, R., Stevens, R. C., and Handel, T. M. (2015) Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. *Science 347*, 1117–22.

(249) Zhang, L., Herrera, C., Coburn, J., Olejniczak, N., Ziprin, P., Kaplan, D. L., and LiWang, P. J. (2017) Stabilization and sustained release of HIV inhibitors by encapsulation in silk fibroin disks. *ACS Biomater. Sci. Eng.* acsbiomaterials.7b00167.

(250) Paul D. Boyer, Edwin G. Krebs, D. S. S. (1970) The enzymes, Third Edition, pp 127–130. Academic Press, New York.

(251) Leal, L., Lucero, C., Gatell, J. M., Gallart, T., Plana, M., and García, F. (2017) New challenges in therapeutic vaccines against HIV infection. *Expert Rev. Vaccines 16*, 587–600.

(252) Ashkenazi, A., Faingold, O., Kaushansky, N., Ben-Nun, A., and Shai, Y. (2013) A highly conserved sequence associated with the HIV gp41 loop region is an

immunomodulator of antigen-specific T cells in mice. *Blood 121*, 2244–52.

(253) Gendelman, H. E., Orenstein, J. M., Martin, M. A., Ferrua, C., Mitra, R., Phipps, T., Wahl, L. A., Lane, H. C., Fauci, A. S., and Burke, D. S. (1988) Efficient isolation and propagation of human immunodeficiency virus on recombinant colony-stimulating factor 1-treated monocytes. *J. Exp. Med. 167*, 1428–41.

(254) Gendelman, H. E., Orenstein, J. M., Baca, L. M., Weiser, B., Burger, H., Kalter, D. C., and Meltzer, M. S. (1989) The macrophage in the persistence and pathogenesis of HIV infection. *AIDS 3*, 475–495.

(255) Westervelt, P., Gendelmant, H. E., and Ratner, L. (1991) Identification of a determinant within the human immunodeficiency virus 1 surface envelope glycoprotein critical for productive infection of primary monocytes (macrophage/tropism/third variable domain of gpl20/CD4). *Med. Sci. 88*, 3097–3101.

(256) Gendelman, H. E., Baca, L. M., Kubrak, C. A., Genis, P., Burrous, S., Friedman, R. M., Jacobs, D., and Meltzer, M. S. (1992) Induction of IFN-alpha in peripheral blood mononuclear cells by HIV-infected monocytes. Restricted antiviral activity of the HIV-induced IFN. *J. Immunol. 148*, 422–9.

(257) Li, M., Salazar-Gonzalez, J. F., Derdeyn, C. A., Morris, L., Williamson, C., Robinson, J. E., Decker, M., Li, Y., Salazar, M. G., Victoria, R., Mlisana, K., Karim, S. A., Greene, K. M., Bilska, M., Zhou, J., Allen, S., Chomba, E., Mulenga, J., Gao, F., Zhang, M., Korber, B. T. M., Hahn, B. H., Montefiori, D. C., Li, M., Salazar-gonzalez, J. F., Derdeyn, C. A., Morris, L., Williamson, C., Robinson, J. E., Decker, J. M., Li, Y., Salazar, M. G., Polonis, V. R., Mlisana, K., Karim, S. A., Hong, K., Greene, K. M., Bilska, M., Zhou, J., Allen, S., Chomba, E., Mulenga, J., and Vwalika, C. (2006) Genetic and Neutralization Properties of Subtype C Human Immunodeficiency Virus Type 1 Molecular env Clones from Acute and Early Heterosexually Acquired Infections in Southern Africa. *J. Virol. 80*, 11776–11790.

(258) Page, K. A., Landau, N. R., and Littman, D. R. (1990) Construction and use of a human immunodeficiency virus vector for analysis of virus infectivity. *J. Virol. 64*, 5270–6.

(259) Nameki, D., Kodama, E., Ikeuchi, M., Mabuchi, N., Otaka, A., Tamamura, H., Ohno, M., Fujii, N., and Matsuoka, M. (2005) Mutations conferring resistance to human immunodeficiency virus type 1 fusion inhibitors are restricted by gp41 and Rev-responsive element functions. *J. Virol. 79*, 764–70.

(260) He, Y., Cheng, J., Lu, H., Li, J., Hu, J., Qi, Z., Liu, Z., Jiang, S., and Dai, Q. (2008) Potent HIV fusion inhibitors against Enfuvirtide-resistant HIV-1 strains. *Proc. Natl. Acad. Sci. U. S. A. 105*, 16332–7.

(261) Pang, W., Tom, S. C., and Zheng, Y. T. (2009) Current peptide HIV type-1 fusion inhibitors. *Antivir. Chem. Chemother. 20*, 1–18.

(262) Jordan, A., Defechereux, P., and Verdin, E. (2001) The site of HIV-1 integration in the human genome determines basal transcriptional activity and response to Tat transactivation. *EMBO J. 20*, 1726–1738.

(263) Bönsch, C., Munteanu, M., Rossitto-Borlat, I., Fürstenberg, A., and Hartley, O.

(2015) Potent anti-HIV chemokine analogs direct post-endocytic sorting of CCR5. *PLoS One* (Spencer, J., Ed.) *10*, e0125396.

(264) Bouchat, S., Gatot, J.-S., Kabeya, K., Cardona, C., Colin, L., Herbein, G., De Wit, S., Clumeck, N., Lambotte, O., Rouzioux, C., Rohr, O., and Van Lint, C. (2012) Histone methyltransferase inhibitors induce HIV-1 recovery in resting CD4+ T cells from HIV-1-infected HAART-treated patients. *AIDS 26*, 1473–1482.

(265) Shirakawa, K., Chavez, L., Hakre, S., Calvanese, V., and Verdin, E. (2013) Reactivation of latent HIV by histone deacetylase inhibitors. *Trends Microbiol. 21*, 277–285.

(266) Bosque, A., and Planelles, V. (2009) Induction of HIV-1 latency and reactivation in primary memory CD4+ T cells. *Blood 113*, 58–65.

(267) Spina, C. A., Anderson, J., Archin, N. M., Bosque, A., Chan, J., Famiglietti, M., Greene, W. C., Kashuba, A., Lewin, S. R., Margolis, D. M., Mau, M., Ruelas, D., Saleh, S., Shirakawa, K., Siliciano, R. F., Singhania, A., Soto, P. C., Terry, V. H., Verdin, E., Woelk, C., Wooden, S., Xing, S., and Planelles, V. (2013) An In-Depth Comparison of Latent HIV-1 Reactivation in Multiple Cell Model Systems and Resting CD4+ T Cells from Aviremic Patients. *PLoS Pathog. 9*, 1–15.

(268) An, W. (2007) Histone acetylation and methylation: combinatorial players for transcriptional regulation. *Subcell. Biochem. 41*, 351–69.

(269) Gilmore, T. D. (2006) Introduction to NF-κB: players, pathways, perspectives. *Oncogene 25*, 6680–6684.

(270) Crabtree, G. R., and Olson, E. N. (2002) NFAT signaling: Choreographing the social lives of cells. *Cell 109*, S67–S79.

(271) Chen, W.-H., Xu, X.-D., Luo, G.-F., Jia, H.-Z., Lei, Q., Cheng, S.-X., Zhuo, R.-X., and Zhang, X.-Z. (2013) Dual-Targeting Pro-apoptotic Peptide for Programmed Cancer Cell Death via Specific Mitochondria Damage. *Sci. Rep. 3*, 3468.

(272) Yin, C., Zhang, T., Qu, X., Zhang, Y., Putatunda, R., Xiao, X., Li, F., Xiao, W., Zhao, H., Dai, S., Qin, X., Mo, X., Young, W. Bin, Khalili, K., and Hu, W. (2017) In Vivo Excision of HIV-1 Provirus by saCas9 and Multiplex Single-Guide RNAs in Animal Models. *Mol. Ther. 25*, 1168–1186.

(273) Das, A. T., Harwig, A., and Berkhout, B. (2011) The HIV-1 Tat Protein Has a Versatile Role in Activating Viral Transcription. *J. Virol. 85*, 9506–9516.

(274) Geng, G., Liu, B., Chen, C., Wu, K., Liu, J., Zhang, Y., Pan, T., Li, J., Yin, Y., Zhang, J., Huang, F., Yu, F., Chen, J., Ma, X., Zhou, J., Kuang, E., Liu, C., Cai, W., and Zhang, H. (2016) Development of an Attenuated Tat Protein as a Highly-effective Agent to Specifically Activate HIV-1 Latency. *Mol. Ther. 24*, 1–10.

(275) Kovalevich, J., and Langford, D. (2012) Neuronal toxicity in HIV CNS disease. *Future Virol. 7*, 687–698.

(276) Bukrinsky, M. I., Haggerty, S., Dempsey, M. P., Sharova, N., Adzhubei, A., Spitz, L., Lewis, P., Goldfarb, D., Emerman, M., and Stevenson, M. (1993) A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells.

Nature 365, 666–669.

(277) Wadia, J. S., Stan, R. V, and Dowdy, S. F. (2004) Transducible TAT-HA fusogenic peptide enhances escape of TAT-fusion proteins after lipid raft macropinocytosis. *Nat. Med. 10*, 310–315.

(278) Tahirov, T. H., Babayeva, N. D., Varzavand, K., Cooper, J. J., Sedore, S. C., and Price, D. H. (2010) Crystal structure of HIV-1 Tat complexed with human P-TEFb.pdf. *Nature 465*, 747–51.

(279) Péloponèse, J. M., Grégoire, C., Opi, S., Esquieu, D., Sturgis, J., Lebrun, E., Meurs, E., Collette, Y., Olive, D., Aubertin, A. M., Witvrow, M., Pannecouque, C., De Clercq, E., Bailly, C., Lebreton, J., and Loret, E. P. (2000) 1H-13C nuclear magnetic resonance assignment and structural characterization of HIV-1 Tat protein. *C. R. Acad. Sci. III. 323*, 883–94.

(280) Tahirov, T. H., Babayeva, N. D., Varzavand, K., Cooper, J. J., Sedore, S. C., and Price, D. H. (2010) Crystal structure of HIV-1 Tat complexed with human P-TEFb. *Nature 465*, 747–751.

(281) Gu, J., Babayeva, N. D., Suwa, Y., Baranovskiy, A. G., Price, D. H., and Tahirov, T. H. (2014) Crystal structure of HIV-1 Tat complexed with human P-TEFb and AFF4. *Cell Cycle 13*, 1788–1797.

(282) Bayer, P., Kraft, M., Ejchart, A., Westendorp, M., Frank, R., and Rösch, P. (1995) Structural studies of HIV-1 Tat protein. *J. Mol. Biol. 247*, 529–35.

(283) Opi, S., Péloponèse, J. M., Grégoire, C., Esquieu, D., Sturgis, J., Lebrun, E., Meurs, E., Collette, Y., Olive, D., Aubertin, a M., Witvrow, M., Pannecouque, C., De Clercq, E., Bailly, C., Lebreton, J., and Loret, E. P. (2000) 1H-13C nuclear magnetic resonance assignment and structural characterization of HIV-1 Tat protein. *Comptes Rendus l' 323*, 883–94.

(284) Kirsch, T., Boehm, M., Schuckert, O., Metzger, A. U., Willbold, D., Frank, R. W., and Rösch, P. (1996) Cloning, high-yield expression in Escherichia coli, and purification of biologically active HIV-1 Tat protein. *Protein Expr. Purif. 8*, 75–84.

(285) Patki, A. H., and Lederman, M. M. (1996) HIV-1 Tat protein and its inhibitor Ro 24-7429 inhibit lymphocyte proliferation and induce apoptosis in peripheral blood mononuclear cells from healthy donors. *Cell. Immunol. 169*, 40–46.

(286) Wei, X., Decker, J. M., Liu, H., Zhang, Z., Arani, R. B., Kilby, J. M., Saag, M. S., Wu, X., Shaw, G. M., and Kappes, J. C. (2002) Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. *Antimicrob. Agents Chemother. 46*, 1896–1905.

(287) Berkmen, M. (2012) Production of disulfide-bonded proteins in Escherichia coli. *Protein Expr. Purif. 82*, 240–51.

(288) Fischer, B., Sumner, I., and Goodenough, P. (1993) Isolation, renaturation, and formation of disulfide bonds of eukaryotic proteins expressed in Escherichia coli as inclusion bodies. *Biotechnol. Bioeng. 41*, 3–13.

(289) Vallejo, L. F., and Rinas, U. (2004) Strategies for the recovery of active proteins

through refolding of bacterial inclusion body proteins. *Microb. Cell Fact. 3*, 11.

(290) Costa, S., Almeida, A., Castro, A., and Domingues, L. (2014) Fusion tags for protein solubility, purification, and immunogenicity in Escherichia coli: The novel Fh8 system. *Front. Microbiol. 5*, 63.

(291) Malhotra, A. (2009, January 1) Chapter 16 Tagging for Protein Expression. *Methods Enzymol.* Academic Press.

(292) Hopp, T. P., Prickett, K. S., Price, V. L., Libby, R. T., March, C. J., Pat Cerretti, D., Urdal, D. L., and Conlon, P. J. (1988) A Short Polypeptide Marker Sequence Useful for Recombinant Protein Identification and Purification. *Bio/Technology 6*, 1204–1210.

(293) LaVallie, E. R., DiBlasio, E. A., Kovacic, S., Grant, K. L., Schendel, P. F., and McCoy, J. M. (1993) A Thioredoxin Gene Fusion Expression System That Circumvents Inclusion Body Formation in the E. coli Cytoplasm. *Nat. Biotechnol. 11*, 187–193.

(294) Smith, D. B., and Johnson, K. S. (1988) Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene 67*, 31–40.

(295) Kapust, R. B., and Waugh, D. S. (1999) Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci. 8*, 1668–1674.

(296) Marblestone, J. G. (2006) Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Sci. 15*, 182–189.

(297) San-Miguel, T., Pérez-Bermúdez, P., and Gavidia, I. (2013) Production of soluble eukaryotic recombinant proteins in E. coli is favoured in early log-phase cultures induced at low temperature. *Springerplus 2*, 89.

(298) Pace, C. N. (1986) Determination and Analysis of Urea and Guanidine Hydrochloride Denaturation Curves. *Methods Enzymol. 131*, 266–280.

(299) Porath, J. (1992) Immobilized metal ion affinity chromatography. *Protein Expr. Purif. 3*, 263–281.

(300) Kiedzierska, A., Czepczynska, H., Smietana, K., and Otlewski, J. (2008) Expression, purification and crystallization of cysteine-rich human protein muskelin in Escherichia coli. *Protein Expr. Purif. 60*, 82–88.

(301) Brand, S. R., Kobayashi, R., and Mathews, M. B. (1997) The Tat protein of human immunodeficiency virus type 1 is a substrate and inhibitor of the interferon-induced, virally activated protein kinase, PKR. *J. Biol. Chem. 272*, 8388–8395.

(302) Yamaguchi, S., Yamamoto, E., Mannen, T., and Nagamune, T. (2013, January 1) Protein refolding using chemical refolding additives. *Biotechnol. J.* WILEY-VCH Verlag.

(303) Frankel, A. D., Chen, L., Cotter, R. J., and Pabo, C. O. (1988) Dimerization of the tat protein from human immunodeficiency virus: a cysteine-rich peptide mimics the normal metal-linked dimer interface. *Proc. Natl. Acad. Sci. U. S. A. 85*, 6297–300.

(304) Moghadam, M., Ganji, A., Varasteh, A., Falak, R., and Sankian, M. (2015) Refolding process of cysteine-rich proteins:Chitinase as a model. *Reports Biochem. Mol.*

Biol. 4, 19–24.

(305) Xie, Y., and Wetlaufer, D. B. (1996) Control of aggregation in protein refolding: the temperature-leap tactic. *Protein Sci. 5*, 517–523.

(306) Yamaguchi, H., and Miyazaki, M. (2014) Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. *Biomolecules 4*, 235–251.

(307) Fink, A. L. (1998, February 1) Protein aggregation: Folding aggregates, inclusion bodies and amyloid. *Fold. Des.* Cell Press.

(308) Karn, J., and Stoltzfus, C. M. (2012) Transcriptional and posttranscriptional regulation of HIV-1 gene expression. *Cold Spring Harb. Perspect. Med. 2*, a006916.

(309) Vanderperren, E., Demaré, W., Blust, R., Cooreman, K., and Bossier, P. (2001) Oestrogenic activity of CPRG (chlorophenol red-β-D-galactopyranoside), a β-galactosidase substrate commonly used in recombinant yeast oestrogenic assays. *Biomarkers 6*, 375–380.

(310) Kashanchi, F., Duvall, J. F., and Brady, J. N. (1992) Electroporation of viral transactivator proteins into lymphocyte suspension cells. *Nucleic Acids Res. 20*, 4673–4674.

(311) Kashanchi, F., Piras, G., Radonovich, M. F., Duvall, J. F., Fattaey, A., Chiang, C.-M., Roeder, R. G., and Brady, J. N. (1994) Direct interaction of human TFIID with the HIV-1 transactivator Tat. *Nature 367*, 295–299.

(312) Kashanchi, F., Shibata, R., Ross, E. K., Brady, J. N., and Martin, M. A. (1994) Second-site long terminal repeat (LTR) revertants of replication- defective human immunodeficiency virus: effects of revertant TATA box motifs on virus infectivity, LTR-directed expression, in vitro RNA synthesis, and binding of basal transcription factor. *J Virol 68*, 3298–3307.

(313) Gutekunst, K. A., Kashanchi, F., Brady, J. N., and Bednarik, D. P. (1993) Transcription of the HIV-1 LTR is regulated by the density of DNA CpG methylation. *JAIDS J. … 6*, 541–549.

(314) Bohan, C. A., Kashanchi, F., Ensoli, B., Buonaguro, L., Boris-Lawrie, K. A., and Brady, J. N. (1992) Analysis of Tat transactivation of human immunodeficiency virus transcription in vitro. *Gene Expr. 2*, 391–407.

(315) Chang, H. C., Samaniego, F., Nair, B. C., Buonaguro, L., and Ensoli, B. (1997) HIV-1 Tat protein exits from cells via a leaderless secretory pathway and binds to extracellular matrix-associated heparan sulfate proteoglycans through its basic region. *AIDS 11*, 1421–1431.

(316) Ensoli, B., Buonaguro, L., Barillari, G., Fiorelli, V., Gendelman, R., Morgan, R. A., Wingfield, P., and Gallo, R. C. (1993) Release, uptake, and effects of extracellular human immunodeficiency virus type 1 Tat protein on cell growth and viral transactivation. *J. Virol. 67*, 277–87.

(317) Jeang, K. T., Xiao, H., and Rich, E. A. (1999) Multifaceted activities of the HIV-1 transactivator of transcription, Tat. *J. Biol. Chem. 274*, 28837–28840.

(318) Geng, G., Liu, B., Chen, C., Wu, K., Liu, J., Zhang, Y., Pan, T., Li, J., Yin, Y., Zhang, J., Huang, F., Yu, F., Chen, J., Ma, X., Zhou, J., Kuang, E., Liu, C., Cai, W., and Zhang, H. (2016) Development of an Attenuated Tat Protein as a Highly-effective Agent to Specifically Activate HIV-1 Latency. *Mol. Ther. 24*, 1–10.

(319) Donahue, D. A., Kuhl, B. D., Sloan, R. D., and Wainberg, M. A. (2012) The Viral Protein Tat Can Inhibit the Establishment of HIV-1 Latency. *J. Virol. 86*, 3253–3263.

(320) Kinch, M. S. (2015) An overview of FDA-approved biologics medicines. *Drug Discov. Today 20*, 393–398.

(321) Denisov, I. G., and Sligar, S. G. (2016) Nanodiscs for structural and functional studies of membrane proteins. *Nat. Struct. Mol. Biol. 23*, 481–486.

(322) Yu, H. (1999) Extending the size limit of protein nuclear magnetic resonance. *Proc. Natl. Acad. Sci. USA 96*, 332–334.

(323) Frueh, D. P., Goodrich, A. C., Mishra, S. H., and Nichols, S. R. (2013, October) NMR methods for structural studies of large monomeric and multimeric proteins. *Curr. Opin. Struct. Biol.* NIH Public Access.

(324) Bayburt, T. H., and Sligar, S. G. (2010, May 3) Membrane protein assembly into Nanodiscs. *FEBS Lett.* NIH Public Access.

(325) Singh, R., and Lillard, J. W. (2009, June) Nanoparticle-based targeted drug delivery. *Exp. Mol. Pathol.* NIH Public Access.

(326) Wang, H., La Russa, M., and Qi, L. S. (2016) CRISPR/Cas9 in Genome Editing and Beyond. *Annu. Rev. Biochem. 85*, 227–264.

Appendix A

Supporting Information for Chapter 2

Supporting information available: Overview of design of vCCI, overview of purification of vCCI, overview of purification of vMIP-II, general overview of one pulse NMR experiment, side chains containing N-H groups in HSQC, chemical shift perturbations, slow versus fast exchange, Isothermal Titration Calorimetry results for titration of vMIP-II into vCCI

Figure A-1: Overview of protein design of vCCI used in Chapter 2. A detailed overview of the vCCI construct used in chapter 2, as discussion in section 2.2.1. Blue sections represent fusion tags used to aid in protein purification and which will be cleaved off by the cleavage protein, enterokinase. Green section represents the protein sequence for rabitpox vCCI, though with the original M1P2 N-terminal sequence of vCCI mutated to A1M2A3 to allow for easy cleavage by enterokinase.

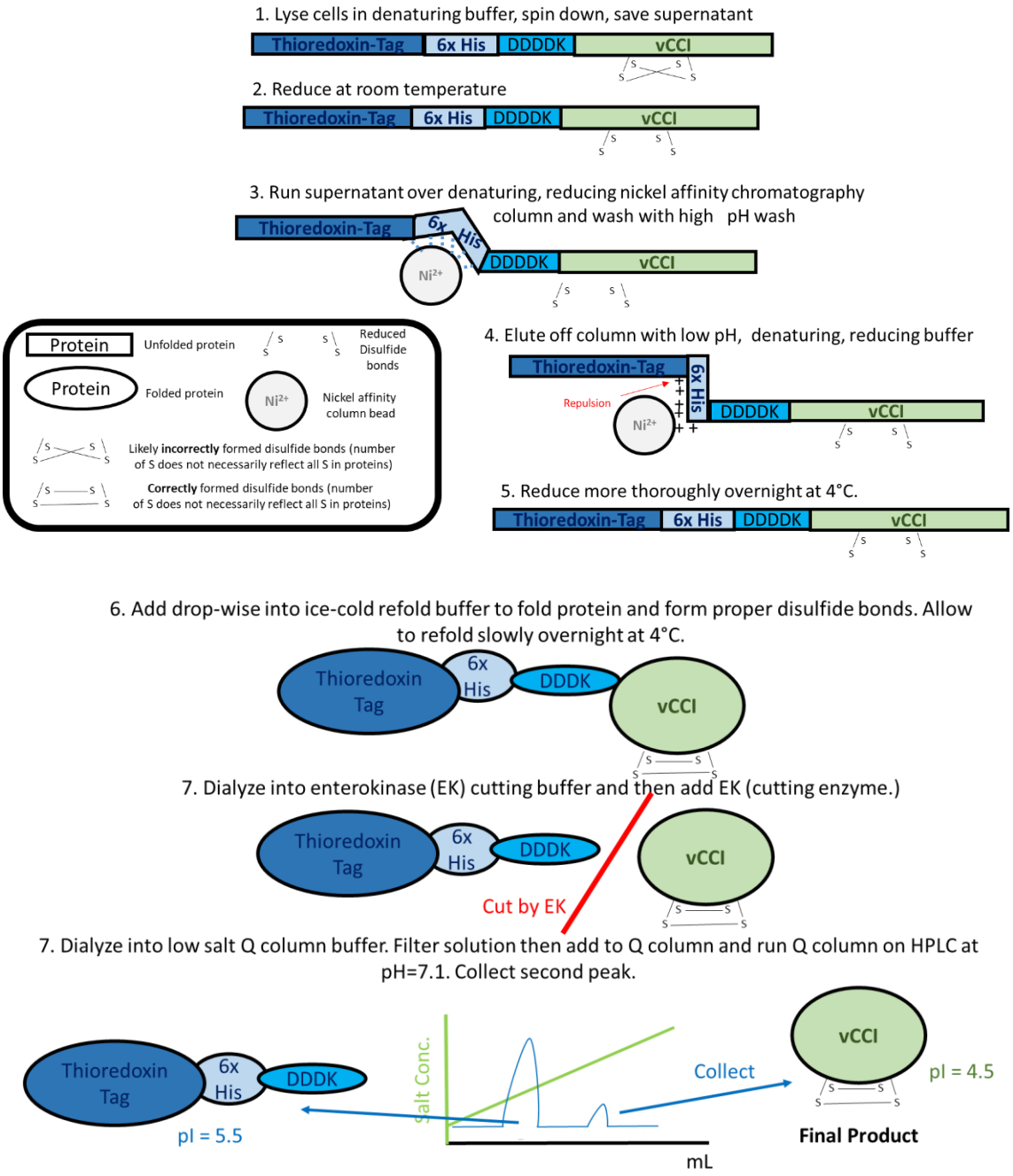

Figure A-2: Overview of the Purification of vCCI. A general overview of the purification of the protein seen in more detail in figure A-1, as described in section 2.2.1. Blue sections represent purification tags, green section represents vCCI protein. Square sections represent steps where the protein is not yet folded properly, while rounded sections represent properly-folded protein. White boxed section shows legend.

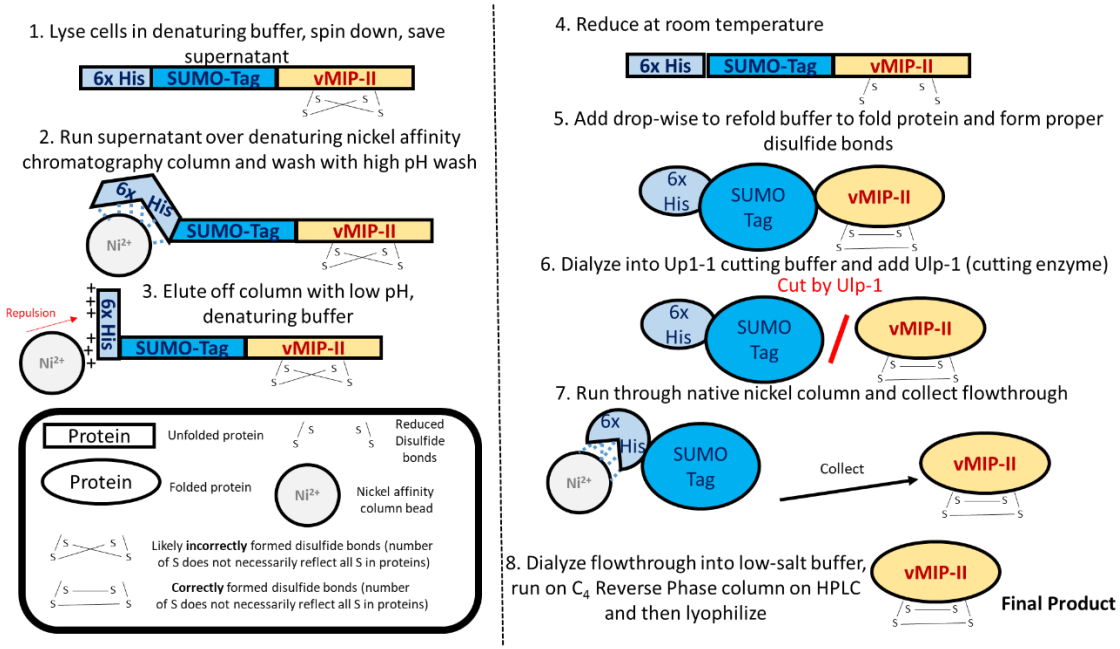

Figure A-3: Overview of the purification of vMIP-II. General overview of the purification of the vMIP-II protein from chapter 2, as described in section 2.2.1. Blue sections represent purification tags, pale orange section represents vMIP-II protein. Square sections represent steps where the protein is not yet folded properly, while rounded sections represent properly-folded protein. White boxed section shows legend.

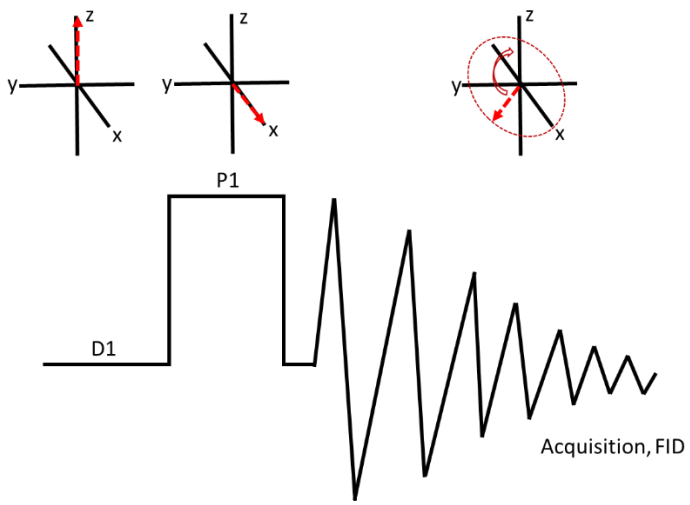

Figure A-4: General overview of a one-pulse NMR experiment. A one-pulse NMR experiment where the static magnetic field is in the z direction and the corresponding orientations of the net magnetization (red arrow) throughout the experiment. At the delay (D1), the orientation the net magnetization is aligned with the magnetic field (z) at equilibrium. At the radiofrequency pulse (P1), the orientation of the net magnetization moves to align with the x axis. After the pulse ends, the magnetization begins to precess along the x–y plane (dotted red circle) as it slowly returns back to equilibrium along the z axis, resulting in FID recorded during the acquisition period.

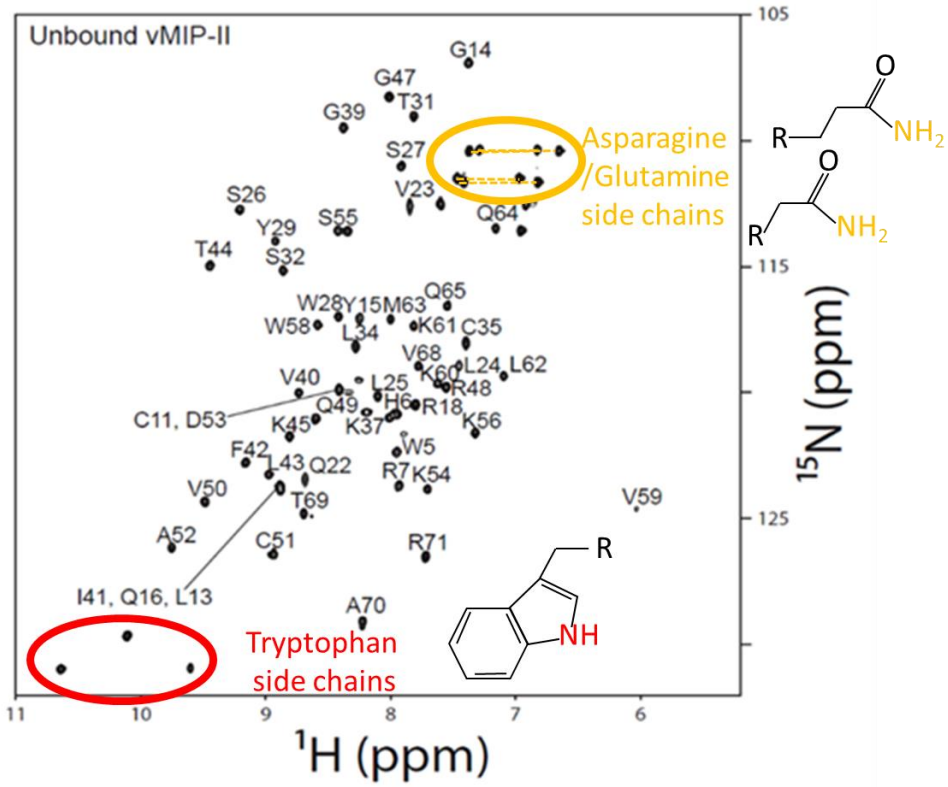

Figure A-5: Side chains containing N-H groups in HSQC. Unbound vMIP-II spectrum from Figure 2-3C with side chain residues circled. Asparagine and glutamine side-chain peaks are circled in orange (side chain chemical structure drawn on the right, outside spectrum). They show doublet peaks due to two hydrogens (x axis) binding to the same nitrogen (y axis), allowing these side chains to be easily recognized due to this doubling pattern. Doublets are connected by orange dashed line. Red circle contains peaks for tryptophan side chain (side chain chemical structure drawn on right, inside spectrum) and are recognizable due to their position generally being above 9.5 ppm in the proton dimension. All other labeled peaks represent backbone residues, corresponding to the N-H bonds on the backbone for each amino acid labeled.

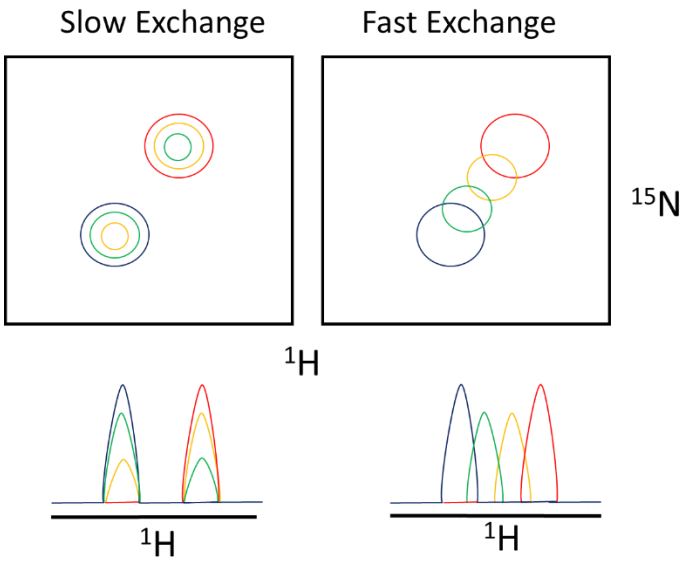

Figure A-6: Chemical shift perturbations, slow versus fast exchange. A very simple schematic representation of chemical shift perturbations in a ^{15}N-HSQC (top) and corresponding ^{1}H 1-D spectra (bottom). In both situations, peaks go from unbound (blue) to roughly 30% bound (green) to roughly 70% bound (orange) to fully bound (red). In slow exchange (left), the unbound peak decreases as the bound peak increases, both in 2-D and 1-D spectra, while in fast exchange, there is a weighted average of the peaks as they shift from the unbound position to the bound position.

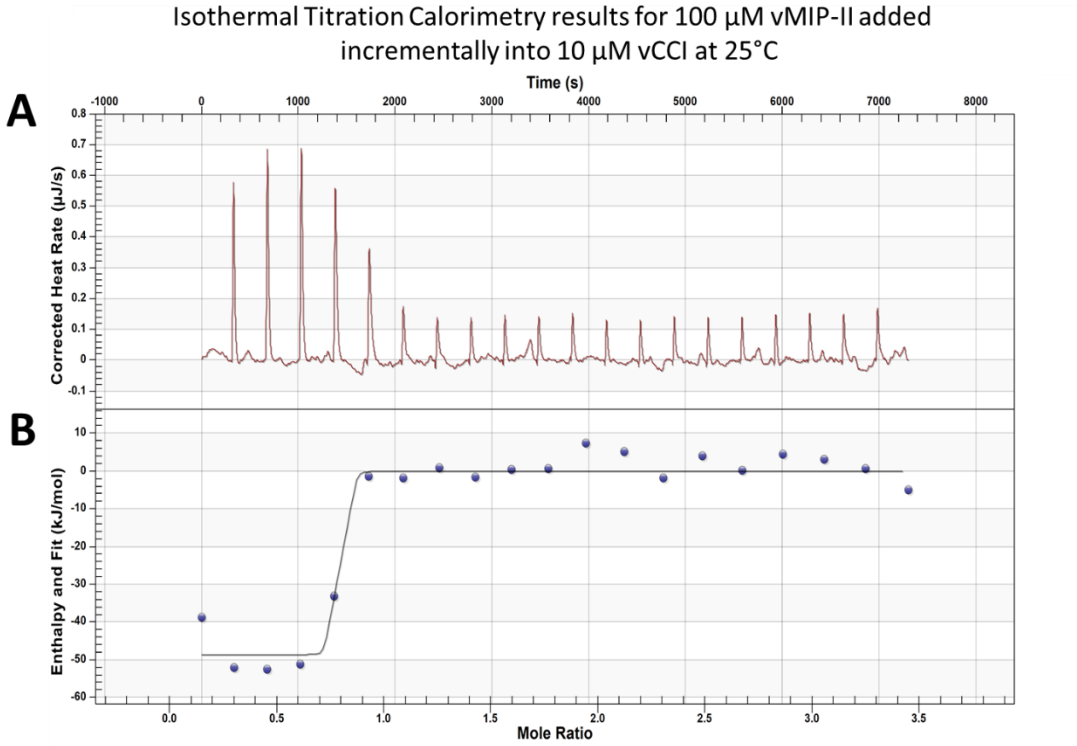

Figure A-7: Isothermal Titration Calorimetry Results for titration of vMIP-II into vCCI. ITC results for a single experiment for 100 µM vMIP-II titrated in 2.5 µL increments into 10 µM vCCI at 25°C. A corresponding buffer-into-buffer experiment as described in section 2.2.3 was subtracted from data above. Figure shows (**A**) raw data corresponding to the heat change observed in each of the 20 injections and (**B**) the data under each peak in (A) integrated, with data fitted to a one-site independent binding site model.

Appendix B

Supporting Information for Chapter 3

Supporting information available: Overview of the purification of 5P12-RANTES variants, example of plate after pseudoviral assay.

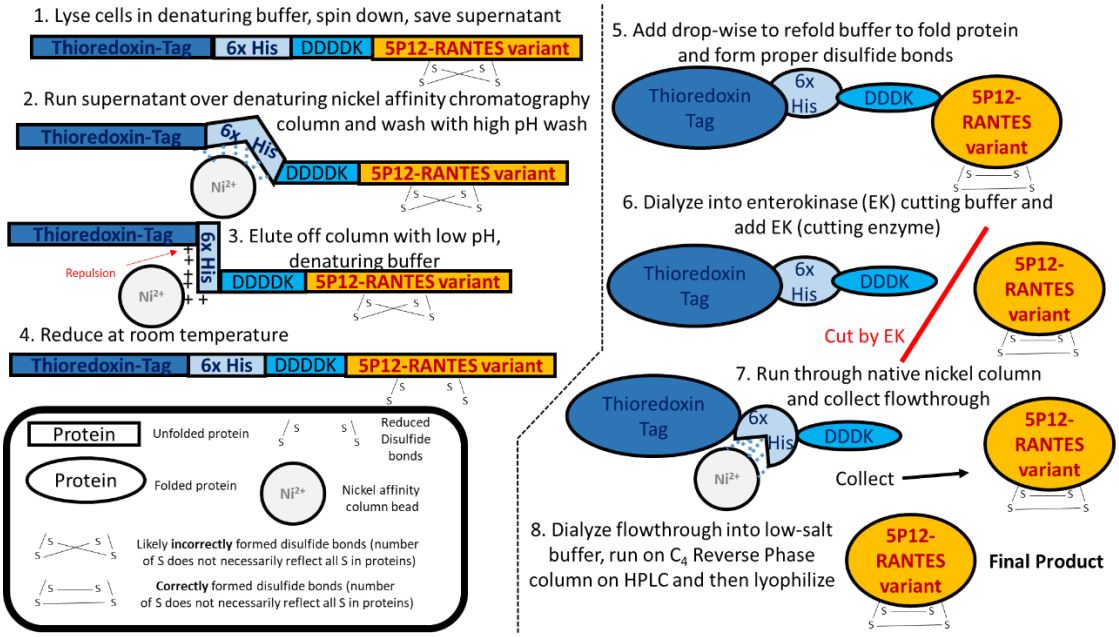

Figure B-1: Overview of Purification of 5P12-RANTES variants. A general overview of the purification of the 5P12-RANTES variants (5P12-RANTES, 5P12-RANTES-Q0E, etc.), as described in section 3.2.1. Blue sections represent purification tags, orange section represents 5P12-RANTES variant. Square sections represent steps where the protein is not yet folded properly, while rounded sections represent properly-folded protein. White boxed section shows legend.

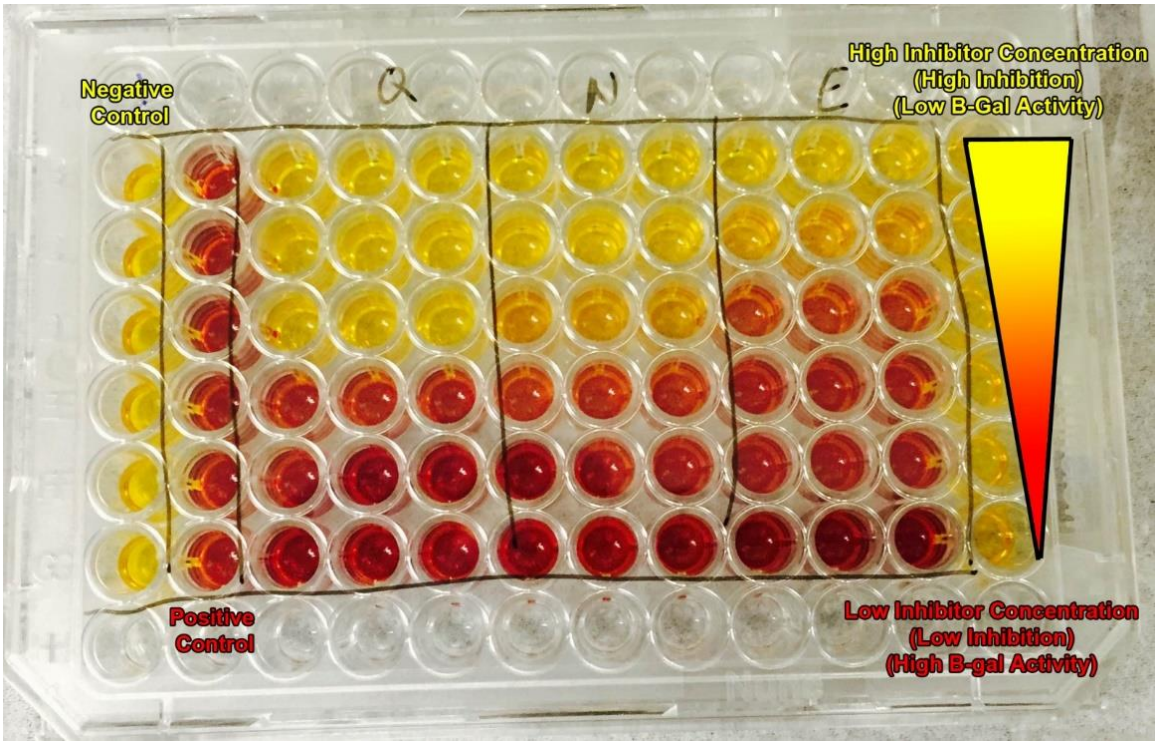

Figure B-2: Example of plate after pseudoviral assay. A 96-well plate showing an example of a single triplicate experiment comparing three different inhibitors (5P12-RANTES, 5P12-RANTES-Q0E, 5P12-RANTES-Q0N). First column on the left shows the negative control (no inhibitor added, no pseudovirus added), second column shows positive control (no inhibitor added, pseudovirus added), while next columns show triplicates of wells containing both inhibitor (higher concentrations of inhibitor in the top wells into a gradient leading to lower concentrations of inhibitor in the bottom wells). Yellow color represents CPRG that was not cleaved because no β-Galactosidase was produced due to full inhibitor of the virus (or no addition of the virus in the case of the negative control), while redder wells represent more β-Galactosidase activity resulting from increased pseudoviral infection due to decreasing concentration of inhibitors.

Appendix C

Supporting Information for Chapter 4

Supporting information available: Purification overview of the 5P12-Linker Variants.

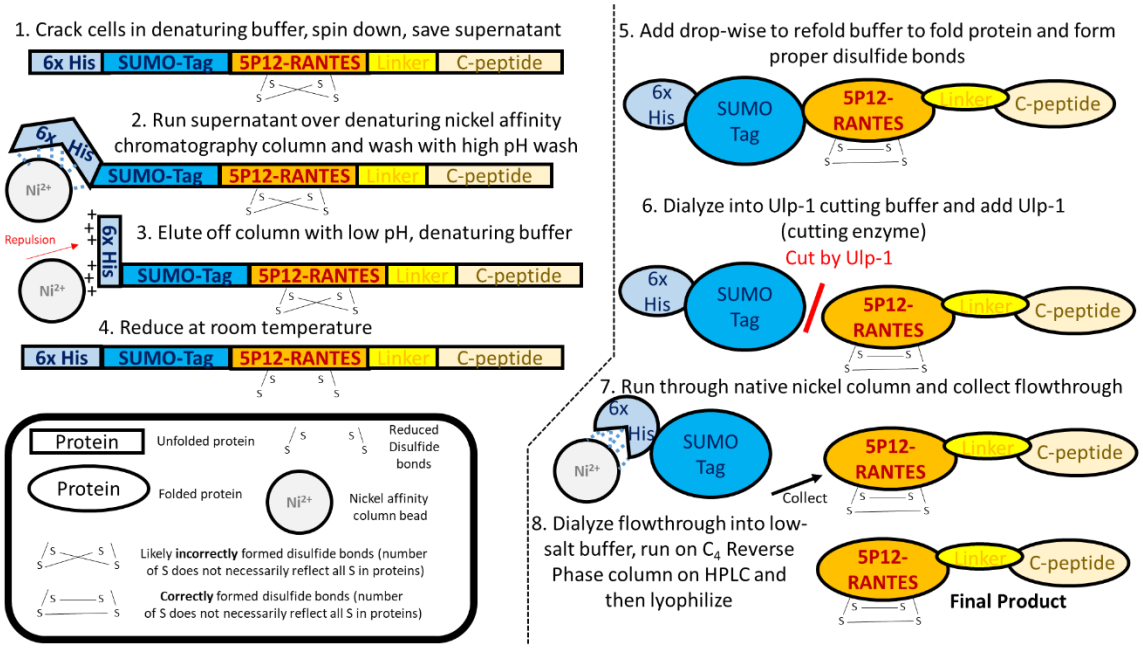

Figure C-1: Overview of the purification of 5P12-Linker variant proteins. A general overview of the purification of the 5P12-Linker variants (5P12-Linker-C37, 5P12-Linker-T1144, 5P12-Linker-T2635), as described in section 4.2.1. Blue sections represent purification tags, orange section represents 5P12-RANTES, yellow section represents the flexible linker, and the tan section represents the C-peptide (C37, T-1144, or T-2635). Square sections represent steps where the protein is not yet folded properly, while rounded sections represent properly-folded protein. White boxed section shows legend.